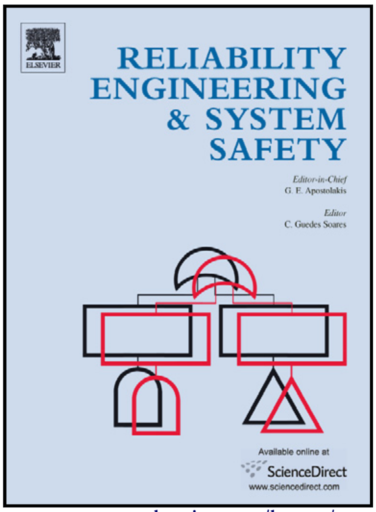
Author's Accepted Manuscript



www.elsevier.com/locate/ress

Two efficient heuristics to solve the integrated load distribution and production planning problem

Yuvraj Gajpal, Mustapha Nourelfath

PII: S0951-8320(15)00222-7
DOI: http://dx.doi.org/10.1016/j.ress.2015.06.019
Reference: RESS5371

To appear in: *Reliability Engineering and System Safety*

Received date: 1 July 2014
Revised date: 10 June 2015
Accepted date: 12 June 2015

Cite this article as: Yuvraj Gajpal, Mustapha Nourelfath, Two efficient heuristics to solve the integrated load distribution and production planning problem, *Reliability Engineering and System Safety,* http://dx.doi.org/10.1016/j.ress.2015.06.019

Revision for Resubmission to Reliability Engineering and System Safety, June 2015

Two efficient heuristics to solve the integrated load distribution and production planning problem

Yuvraj Gajpal[a,*], Mustapha Nourelfath[b]

[a] Asper School of Business, University of Manitoba

Winnipeg, Manitoba, R3T 5V4, Canada

[b] Université Laval

Interuniversity Research Center on Enterprise Networks, Logistics and Transportation (CIRRELT)

Faculty of Science and Engineering

Mechanical Engineering Department

Quebec (QC.) G1K 7P4, Canada

* Corresponding author. E-mail address: Yuvraj.Gajpal@umanitoba.ca (Yuvraj Gajpal).

Abstract

This paper considers a multi-period production system where a set of machines are arranged in parallel. The machines are unreliable and the failure rate of machine depends on the load assigned to the machine. The expected production rate of the system is considered to be a non-monotonic function of its load. Because of the machine failure rate, the total production output depends on the combination of loads assigned to different machines. We consider the integration of load distribution decisions with production planning decision. The product demands are considered to be known in advance. The objective is to minimize the sum of holding costs, backorder costs, production costs, setup costs, capacity change costs and unused capacity costs while satisfying the demand over specified time horizon. The constraint is not to exceed available repair resources required to repair the machine breakdown. The paper develops two heuristics to solve the integrated load distribution and production planning problem. The first heuristic consists of a three-phase approach, while the second one is based on tabu search metaheuristic. The efficiency of the proposed heuristics is tested through the randomly generated problem instances.

Keywords: Heuristics; Production planning; Load distribution; Reliability; Multi-state systems, Lot-sizing

1. Introduction

This paper develops two heuristics to solve the problem of integrated load distribution and production planning problem. The problem was recently proposed in [1]. Many empirical studies of mechanical systems and computer systems have proven that the workload strongly affects the failure rate [2]. In production industries, machines are often run at different production rates to meet the production requirements. Some engineering systems are even designed to support varying amounts of loads, such as conveyers, computer processors, load-carrying systems, cutting tools, etc. When machines are over-utilized or run at higher speed rate, more failures and interruptions are observed. In industry, machines are often run at overloaded conditions. For example, the machines are overloaded when the demand is very high so that backorder can be avoided. However, the overloading of machines increases their failure rate which ultimately reduces the production rate. This situation is frequently observed in sawmills and in other manufacturing lines. In this situation managers often face the problem of optimal load distribution while planning their production systems. It is therefore important to consider load versus failure rate relationship while performing production planning optimization.

In [1], authors considered a series-parallel multi-state production system containing a set of non-identical machines. In most of the production systems, machines are arranged in parallel. Thus, we consider a multi-state production system in this paper in which machines are arranged in parallel. Each machine can be assigned to a discrete load which corresponds to a possible production rate of the machine. In general, the failure rate of the machine increases with the increase of assigned load. When machines are assigned to a higher load, their production rates increase for a short period of time. However, the higher machine load also increases the number of failures, which ultimately reduces the average production rate over a long planning horizon. Therefore, the expected machine performance behaves as a non-monotonic function of its assigned load. In [3], the authors developed a model that determines the optimal load, on each component of a series-parallel multi-state system, to provide the maximal expected performance. The objective of this paper is to develop the solution technique for an integrated system in which load distribution decisions and tactical production planning decisions are combined. The integrated problem can be defined as follows. The system produces a set of products during a given planning horizon. A demand is to be satisfied at the end of period for each product. The integrated plan should determine the quantities of items to be produced (lot-sizes) for each period and the optimal load on each machine. The available repair resources required to fix the machine breakdowns are limited. Thus the loads are assigned in the machines in such a way that the machines can be fixed using the available repair resources (i.e. the required repair resources are fewer than the available repair resources). The objective is to minimize the sum of setup costs, holding costs, backorder costs, production costs, capacity change costs and unused capacity costs while satisfying the demand for all products over the entire planning horizon.

Recently, a number of authors considered optimal load distribution system with component failures ([17], [18] [19], [21]). The authors of [17] considered the optimal loading in series-parallel system with aim to maximize system availability. The problem of optimal reliability in over-actuated systems is considered in [18]. In [19] the authors solve the optimal load distribution problem in which mission cost is minimized while satisfying the system reliability constraints. In [21] authors analyze the parameter dependency in stochastic modeling. The analysis provided in [21] can be applied in load optimization problem with system failure.

The integrated model described in [1] includes the costs of changing machine capacities and the costs of unused capacities. Usually the varying production rates from one period to another period increases the costs. A load distribution plan that has less frequent changes in machine capacities will have lower costs while the load distribution plan that has higher changes in machine capacity will have higher costs. Furthermore, the consequences of idle equipment may be undesirable. The leftover capacity may result in a positive cost or penalty. An optimal control of manufacturing systems under production-dependent failure rates is considered in [22]. The authors of [22] considered two parallel machines with production-dependent failure rate, single type of product and minimization of inventory and storage cost in infinite time horizon.

The integrated model described in [1] has shown that the combining load distribution decisions with production planning decisions may reduce the total expected cost. In fact, production planning and load distribution planning may be in conflict. The objective of production planning is to minimize the total production cost while the objective of load distribution planning is to maximize the production rate. The optimal load distribution plan tends to maximize the production rate independently, which may lead to the higher costs of unused capacity. Consequently, if load distribution planning and production planning activities are performed sequentially, the integrated production and load distribution plans could be non-optimal with respect to the objective of minimizing the combined cost.

There are many papers in reliability engineering that deal with optimal load distributions. One can distinguish between static and dynamic problems to consider the effects of load on machine failures. In the static load distribution models, the machine load is constant for a given time period. In the dynamic load distribution models, the load on a given machine may vary with time [12]. Although the dynamic load distribution approach has a wide range of applications, they require more information to capture the information on the system state. On the other hand, the static load distribution problem requires less information and does not require tracking of the system states over a time. Hence, the static load distribution problems are considered to be more practical [3] because of the fewer record keeping activities. This paper considers static load distribution problem where machine loads are constant during a given time period. Although, the machine loads may vary among different time periods.

A number of research papers can be found in the literature dealing with the tactical production planning decision (See [8] and [9] for the advancement in this area). A comparison of lot sizing methods considering capacity change costs can be found in [13]. In [14], the authors study the impact of the unused capacity cost on production planning of manufacturing systems. In [15], the authors develop a mathematical model for production lot-sizing with variable production rate and explicit idle capacity cost. In general, solution methodologies for multi-product capacitated lot-sizing problems vary from traditional linear mixed-integer programming and associated branch and bound based exact methods to the heuristic methods: see [10] for a survey on solution methods for lot sizing problems. A review on metaheuristics for dynamic lot sizing problem can be found at [23]. In recent times, numerous research papers have considered solving the variant of capacitated lot sizing problems (see [24]-[27]).

To the best of our knowledge, the only papers which address the issue of combining load distribution and production planning are [1] and [20]. The problem addressed in [1] is important since it considers the load versus failure rate relationship while optimizing planning of production systems. The benefit of integrated model was illustrated using an exhaustive evaluation of all the load distribution methods in [1]. The exhaustive enumeration approach in not suitable for large problem instances because the computer CPU time increases exponentially with the problem size. Therefore, we present two heuristics, a three-phase based heuristic and a tabu search based heuristic to solve the integrated production and load distribution planning model for multi-state production system in which machines are arranged in parallel. Our heuristics solves the multi-state system with only parallel machine arrangement. However, the heuristics can be easily extended to the multi-state system with series-parallel arrangement. The ideas of proposed three-phase based heuristic and initial results were presented in [20]. In the present paper, we provide the complete description of the three-phase based heuristic and propose a tabu search based heuristic to solve the problem. We perform extensive numerical experiment to evaluate the performance of proposed heuristics.

The remainder of the paper is structured as follows. Section 2 presents some definitions, preliminaries and formulation of integrated load distribution and production planning model. Section 3 presents our three-phase based heuristic. The proposed tabu search based heuristic is presented in Section 4. Numerical examples are presented in Section 5 followed by the conclusions in Section 6.

2. Model description

In this section, first we provide some concepts needed to describe the model and then we present the mathematical model for the integrated system considered in this paper. We use the following notations to describe the mathematical model.

Notation

H planning horizon

T number of periods

t period index ($t = 1, 2, \ldots, T$)

P set of products

p product index ($p \in P$)

d_{pt} demand of product p by the end of period t

n number of machines

j machine index ($j = 1, 2, \ldots, n$)

L_j^t load on machine j during period t

$\overline{L_j}$ maximum allowed load on machine j

$\underline{L_j}$ minimum allowed load on machine j

L_{j0} baseline load of machine j

λ_j^t failure rate of machine j during period t

α_j parameter of component j power law

A_j^t steady-state availability of machine j during period t

μ_j repair rate of machine j

g_j^t production rate of machine j during period t

G_{MSS} average production rate of the multi-state system (MSS)

h_{pt} inventory holding cost per unit of product p by the end of period t

b_{pt} backorder cost per unit of product p by the end of period t

s_{pt} fixed set-up cost of producing product p in period t

π_t variable cost of producing one unit of product p in period t

u unitary cost of unused capacity

c_{jt} cost of changing load distribution for machine j from period t-1 to period t

δ_{ij} Kronecker delta function; $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise

RT required repair resource

RT_0 available repair resource

Decision variables

x_{pt} quantity of product p to be produced in period t

I_{pt} inventory level of product p at the end of period t

B_{pt} backorder level of product p at the end of period t

y_{pt} binary variable, which is equal to 1 if the setup of product p occurs at the end of period t, and 0 otherwise

L^t vector of loads of all the machines for a given period t, $\mathrm{L}^t = \left\{L_1^t, L_2^t, \ldots, L_n^t\right\}$

2.1. The machine failure rate

In this paper, we assume that the failure rate of machine j depends on the load assigned to the machine. Let L_j^t be the load on machine j during period t and let $\lambda_j^t\left(L_j^t\right)$ be the failure rate of machine j when load L_j^t is assigned to the machine j. There is an upper and lower limit on the amount of load that can be assigned to machine j such that $L_j^t \in \left[\underline{L}_j, \overline{L}_j\right]$. Here $\underline{L}_j \in N$ and $\overline{L}_j \in N$ are, respectively, the minimum and the maximum allowed loads on machine j. It is assumed that L_j^t takes only integer values. The function that describes the relationship between the load and failure rate $\lambda_j^t\left(L_j^t\right)$ is known. We use the accelerated failure-time model (AFTM) described in [4] to express the relationship between the load and the failure rate of machine. This relationship is defined as follows:

$$\lambda_j^t\left(L_j^t\right) = \lambda_j\left(L_{j0}\right)\left(L_j^t / L_{j0}\right)^{\alpha_j},$$

(1)

Here L_{j0} is the initial load of machine j, $\lambda_j\left(L_{j0}\right)$ is the failure rate at the initial load, and α_j is the parameter of component j power law. This relationship can be simplified as follows:

$$\lambda_j^t\left(L_j^t\right) = k_j\, {L_j^t}^{\alpha_j}, \text{ where } k_j = \frac{\lambda_j\left(L_{j0}\right)}{{L_{j0}}^{\alpha_j}}.$$

(2)

2.2. The machines availability

We need the estimate for average availability of the machine to calculate the average production rate of the system. The availability of machine depends on the failure and repair rate of the machine. Let $A_j^t\left(L_j^t\right)$ denote the steady-state availability for machine j at time period t when machine j is assigned to load L_j^t. Let μ_j denote the repair rate of machine j then the availability can be expressed as:

$$A_j^t\left(L_j^t\right)=\frac{\mu_j}{\lambda_j^t\left(L_j^t\right)+\mu_j}. \tag{3}$$

2.3. The average production rate of the system

The average production rate of the system depends on the loads assigned to the machines. Let $\mathbf{L}^t=\left\{L_1^t, L_2^t, \ldots, L_n^t\right\}$ be the vector of loads assigned to all the machines for a given time period t. The average production rate of the entire system is denoted by $G_{MSS}(\mathbf{L}^t)$ which is a function of the load vector $\mathbf{L}^t$. It depends on the performance of its machines and their availability. Once the average availability is calculated for each period, and for each machine, an appropriate evaluation method can be used to calculate $G_{MSS}(\mathbf{L}^t)$. A detailed study of these evaluation methods can be found in [6] and [7]. In case of parallel arrangement of the machine, the average production rate of the system is expressed as a sum of the production rate of an individual machine.

2.4. The integrated model

The problem under study considers an integrated load distribution, and production planning for a multi-period production system. The problem involves determining the load assigned to the machines at each time period and the quantities of items (lot-sizes) to be produced at each time period. The objective is to minimize the sum of holding costs, backorder costs, production costs, setup costs, capacity change costs and unused capacity costs while satisfying the demand over specified time horizon. The integrated model is mathematically formulated as follows:

Minimize

$$\sum_{p\in P}\sum_{t=1}^{T}\left(h_{pt}I_{pt}+b_{pt}B_{pt}+\pi_{pt}x_{pt}+s_{pt}y_{pt}\right)+\sum_{j=1}^{n}\sum_{t=1}^{T}c_{jt}\left(1-\delta_{L_j^t L_j^{t-1}}\right)+\sum_{t=1}^{T}u\left(G_{MSS}\left(\mathbf{L}^t\right)-\sum_{p\in P}x_{pt}\right) \tag{4}$$

subject to $$I_{pt} - B_{pt} = I_{pt-1} - B_{pt-1} + x_{pt} - d_{pt},\ \ p \in P,\ \ t = 1,\ 2,\ ...,\ T, \tag{5}$$

$$x_{pt} \le \left(\sum_{q \ge t} d_{pq} \right) y_{pt},\ \ p \in P,\ \ t = 1,\ 2,\ ...,\ T, \tag{6}$$

$$\sum_{p \in P} x_{pt} \le G_{MSS}(\mathrm{L}^t),\ t = 1,\ 2,\ ...,\ T, \tag{7}$$

$$RT(\mathbf{L}^t) \le RT_0,\ t = 1,\ 2,\ ...,\ T, \tag{8}$$

$$RT(\mathbf{L}^t) = \sum_{j=1}^{n} \frac{\lambda_j^t\left(L_j^t\right)}{\mu_j},\quad t = 1,\ 2,\ ...,\ T. \tag{9}$$

$$\mathbf{L}^t = \left\{ L_1^t, L_2^t, ..., L_n^t \right\},\ t = 1,\ 2,\ ...,\ T, \tag{10}$$

$$L_j^t \in \left\{ \underline{L_j},\ \underline{L_j}+1,\ \underline{L_j}+2,\ ...,\ \overline{L_j}-1,\ \overline{L_j} \right\},\quad j = 1,\ 2,\ ...,\ n;\quad t = 1,\ 2,\ ...,\ T, \tag{11}$$

$$x_{pt}, I_{pt}, B_{pt}, \underline{L_j}, \overline{L_j} \in \mathbb{N};\ y_{pt} \in \{0,\ 1\}. \tag{12}$$

The objective function (4) consists of three types of costs. The first term represents the sum of inventory cost, back order cost, production cost and setup cost. The second term represents the cost of changing machine capacity between period t-1 to period t. Here, δ_{ij} is the Kronecker delta function, which takes value 1 if $i = j$ and zero otherwise. The third term represents the cost of unused capacity. Equation (5) ensures the balance of inventory flow between two consecutive periods. Equation (6) ensures production quantity x_{pt} is 0 if setup variable y_{pt} takes value zero and the production quantity is positive if setup variable y_{pt} takes value 1. Equation (7) ensures that the total production quantity is less than the average production quantity of the system. Equation (8) specifies the constraint on the total repair resources that should not exceed RT_0. Equation (9) provides formula for calculating the total expected repair time.

Equation (10) and (11) restrict the decision variables to take values between the minimum and the maximum allowed loads. Finally equation (12) enforces integer and binary decision variables. This integrated model determines the optimal values of production plan and the load assigned to the machines. For each product p, and for each period t, the decision variables are x_{pt}, I_{pt}, B_{pt}, y_{pt} and $\mathbf{L}^t$. For a period (t = 1, …, T), each possible load distribution is represented by a vector $\mathbf{L}^t = \left\{ L_j^t \right\}$ with j = 1, …, n.

3. The proposed three-phase based heuristic

This section describes the proposed three-phase based heuristic and lower bound calculation for the problem under investigation. The ideas of proposed three-phase based heuristic were initially presented in [20]. In this paper, we provide the complete descriptions and details of the proposed three-phase based heuristic.

3.1. Solution procedure

The integrated load distribution and production planning problem is solved using a three-phase based heuristic. The problem under investigation combines two important aspects of production system, load distribution and production planning. Our heuristic solves the problem in three phases. In the first phase, an optimal production rate is determined irrespective of the time period. The optimal production rate imposes an upper limit restriction for the quantity of product which can be produced in a given time period. The second phase solves a production planning problem by imposing the optimal production rate limit obtained in phase 1 as an available production capacity. In the third phase, a load distribution problem is solved to determine the machine load which satisfied the production quantity obtained in phase 2. The second phase minimizes the production planning cost alone while the third phase minimizes the unused capacity cost and load changing cost. The detailed description of the three-phase approach is described in the following subsections.

3.1.1. Phase 1: Determine optimal production rate

This phase determines the feasible optimal production rate that can be achieved by *n* machines, while respecting the maximum and minimum allowed load on the machines. The load distribution problem aims to maximize the production rate in such a way that the total repair resources used are less than the available repair resources and the load assigned to the machine is within their minimum and maximum limit. This phase seeks to solve the following load distribution problem in the first phase. Let's call this problem a problem P1.

Problem P1:

Maximize $G_{MSS}(L)$

(13)

Subject to $RT(L) = \sum_{j=1}^{n} \frac{\lambda_j(L_j)}{\mu_j}$

(14)

$$RT(L) \le RT_0$$

(15)

$$L = \{L_1, L_2, ..., L_n\}$$

(16)

$$L_j \in \{\underline{L_j}, \underline{L_j}+1, ..., \overline{L_j}-1, \overline{L_j}\},\ j = 1, 2, ..., n$$

(17)

This problem cannot be solved exactly because of non-linear objective functions and constraints. Therefore, we propose a heuristic which solves problem P1 in two steps. The first step finds the optimal production rate by relaxing the available resource capacity constraint ($RT(L) \le RT_0$) and the second step finds the feasible solution by decreasing the machine's load until the required repair resources are within the available repair resources.

Step 1: Finding Initial Solution

In this paper, we consider a production system where machines are arranged in a parallel. In this case, the total production rate $G_{MSS}(L)$ is expressed as a sum of individual production rates. Thus the problem is decomposed into an n sub problem where the j^{th} sub-problems can be formulated as:

Problem P2:

Maximize $\quad G_{MSS}(L_j)$

(18)

Subject to $\quad L_j \in \{\underline{L_j}, \underline{L_j}+1, ..., \overline{L_j}-1, \overline{L_j}\}$

(19)

We consider only two states, working state and failure state for the machines. The probability associated with the working state is $A_j(L_j)$ and failed state is $(1-A_j(L_j))$. The expected average production rate $G_{MSS}(L_j)$, thus, can be represented as an $A_j(L_j)\times L_j+(1-A_j(L_j))\times 0$ which in turn can be simplified as an $A_j(L_j)\times L_j$. The stationary availability of machine j for assigned load L_j is expressed as:

$$A_j(L_j)=\frac{\mu_j}{\mu_j+\lambda_j(L_j)} \tag{20}$$

After substituting the expression for $A_j(L_j)$, problem P2 is expressed as:

Maximize $\quad \dfrac{\mu_j \times L_j}{\mu_j+\lambda_j(L_j)}$

(21)

Subject to $\quad L_j \in \{\underline{L_j}, \underline{L_j}+1, ..., \overline{L_j}-1, \overline{L_j}\}$

(22)

The objective function is a simple continuous non-linear function of machine load which maximizes at load $L_j^{free} = \left(\frac{\mu_j}{k_j(\alpha_j - 1)} \right)^{1/\alpha_j}$. With maximum and minimum load constraint, the optimal load is expressed as an $L_j^{continious} = \min(\max(L_j^{free}, \underline{L_j}), \overline{L_j})$. If the value of $L_j^{continious}$ is integer then this solution is considered as an optimal load, otherwise two adjacent integer points are checked for the optimal load. We will use notation $L_j^{Glob-opt}$ and $G_{MSS}^{Glob-opt}$ to represent the load on machine *j* and optimal production rate obtained by solving problem P2. We use notation $RT^{Glob-opt}$ to represent the total resources used when machine loads are set to $L_j^{Glob-opt}$.

Step 2: Finding feasible solution

The solution obtained in step 1 could be infeasible with respect to resource capacity constraint. If $\sum_{j=1}^{n} \frac{\lambda_j^t \left(L_j^{Glob-opt} \right)}{\mu_j} \le RT_0$ then the solution obtained in step 1 is considered to be feasible otherwise it is infeasible. If the problem is infeasible then an attempt is made to find the feasible solution by reducing the machines load to bring the required repair resources below the available repair resources. When a machine load is reduced from current load, the average production rate and the required repair resource also decreases. In order to find the feasible solution, we start with the optimal load $L_j^{Glob-opt}$. We identify the machine which reduces the average production rate least and repair resources most when machine load is reduced by one unit. Then the load of the identified machine is reduced by 1. We keep on reducing the machine load till the required repair resources becomes less than the available repair resources. The complete description of step 2 is provided below.

Step 0: Initialize the current load $L_j^{Current}$, and current required repair resources $RT^{Current}$ as follows:

$$L_j^{Current} = L_j^{Glob-opt}, j = 1, 2, ..., n$$

$$RT^{Current} = RT^{Glob-opt}$$

Step 1: Calculate reduction in required repair resources and average production rate if load of machine *j* is reduced by one unit at $L_j^{Trial} = L_j^{Current} - 1$. Let ΔRT_j and ΔG_{MSS}^j denote the reduction in required repair resources and reduction in production rate respectively. These values are calculated for only those machines whose current load is more than their minimum allowed load. Let these machines be called eligible machines, then the ΔRT_j and ΔG_{MSS}^j for each eligible machine *j* is calculated using the following expression:

$$\Delta RT_j = \frac{\lambda_j(L_j^{Current}) - \lambda_j(L_j^{Trial})}{\mu_j}$$

$$\Delta G_{MSS}^j = \frac{\mu_j \times L_j^{Current}}{\mu_j + \lambda_j(L_j^{Current})} - \frac{\mu_j \times L_j^{Trial}}{\mu_j + \lambda_j(L_j^{Trial})}$$

Step 2: Calculate the ratio of ΔG_{MSS}^j and ΔRT_j for each eligible machine *j* and select the machine that minimizes this ratio. Let's assume that machine *k* has the minimum ratio of $\Delta G_{MSS}^j / \Delta RT_j$.

Step 3: Update the current load $L_k^{Current}$ for machine *k*, and current required repair resources as:

$$L_k^{Current} = L_k^{Trial}$$

$$RT^{Current} = RT^{Current} - \Delta RT_j$$

Step 4: If $RT^{Current} \leq RT_0$ then go to step 5, otherwise go back to step 1.

Step 5: The current machine load $L_j^{Current}$ is the heuristic optimal solution for problem P2. Let's denote the optimal load by L_j^{Opt} and the corresponding production rate by G_{MSS}^{Opt}.

3.1.2. Second Phase: Solving production planning problem

The second phase of the proposed three-phase based heuristic solves the production planning problem by imposing an optimal production rate as a production capacity constraint on the master problem. Note that all n machines can feasibly produce only G_{MSS}^{Opt} units in a given time period due to failure rate of machine, maximum allowable load on machine and due to availability of repair resources. The objective of the second phase is to minimize the production planning cost and determine the production quantity, backorder quantity and inventory level for product at each period. We solve the following production planning problem which is referred as a problem P3.

Problem P3:

Minimize $$\sum_{p\in P}\sum_{t=1}^{T}\left(h_{pt}I_{pt}+b_{pt}B_{pt}+\pi_{pt}x_{pt}+s_{pt}y_{pt}\right)$$

(23)

Subject to $$I_{pt}-B_{pt}=I_{pt-1}-B_{pt-1}+x_{pt}-d_{pt},\ \ p\in P,\ \ t=1,\ 2,\ ...,\ T,$$

(24)

$$x_{pt}\le\left(\sum_{q\ge t}d_{pq}\right)y_{pt},\ \ p\in P,\ \ t=1,\ 2,\ ...,\ T,$$

(25)

$$\sum_{p\in P}x_{pt}\le G_{MSS}^{Opt},\ \ t=1,\ 2,\ ...,\ T,$$

(26)

$$x_{pt},\ I_{pt},\ B_{pt}\in\mathbb{N};\ y_{pt}\in\{0,\ 1\}. \tag{27}$$

We use CPLEX to solve the problem P3 exactly. CPLEX is an optimization software package of IBM which can be used for solving large and difficult mixed-integer programming problems. This phase determines the production quantity x_{pt}, the back order quantity B_{pt}, and the inventory quantity I_{pt}. Let $X_t=\sum_{p\in P}x_{pt}$ denote the total quantity produced in period t which is determined after solving problem P3. Quantity X_t is now a known quantity. The third phase tries to minimize the unused capacity cost by finding the machine load close to the value X_t.

3.1.3. Third Phase: Solve load distribution problem

The third phase solves the modified load distribution problem to minimize the unused capacity cost by setting the machine load in such a way that the total production quantity is close to the required production quantity X_t obtained in second phase. Our aim in third phase is to solve the following problem called problem P4.

Problem P4:

Minimize $\sum_{j=1}^{n}\sum_{t=1}^{T} c_{jt}\left(1-\delta_{L_j^t L_j^{t-1}}\right)+\sum_{t=1}^{T} u\left(G_{MSS}\left(\mathbf{L}^t\right)-X_t\right)$

(28)

Subject to $G_{MSS}(L^t)=\sum_{j=1}^{n}\frac{\mu_j \times L_j}{\mu_j+\lambda_j(L_j)}, t=1,2,...,T,$

(29)

$$G_{MSS}(L^t)\geq X_t, t=1,2,...,T,$$

(30)

$$L_j^t \in \{\underline{L_j}, \underline{L_j}+1,...,\overline{L_j^{Opt}}-1, \overline{L_j^{Opt}}\}, j=1,2,...,n; t=1,2,...,T.$$

(31)

This problem cannot be solved exactly due to non-linearity in objective function and constraints, therefore, we use heuristic to solve the problem. It can be observed that when machines are assigned to optimal load L_j^{Opt}, the average production rate is higher than the required production rate X_t because of the constraint (26). Thus if machines are assigned to load L_j^{Opt} then it will satisfy constraints (29)-(31) but the unused capacity cost will be high which can be lowered by reducing the machine loads. Thus, our heuristic starts with the optimal load L_j^{Opt}, for all the machines and then it tries to improve the solution by decreasing the load to bring the average production rate (i.e., $G_{MSS}(L^t)$) close to the required production quantity (i.e., X_t). The complete description of the heuristic for time period *t* is described below:

Step 0: Initialize the current load $L_j^{Current}$, and current average production rate as follows:

$$L_j^{Current} = L_j^{Opt},\ j = 1, 2, ..., n,$$

$$G_{MSS}(L^{Current}) = \sum_{j=1}^{n} \frac{\mu_j \times L_j^{Opt}}{\mu_j + \lambda_j(L_j^{Opt})}$$

Step 1: Calculate reduction in average production rate when load of machine *j* is reduced by one unit to set $L_j^{Trial} = L_j^{Current} - 1$. Let ΔG_{MSS}^j denote the reduction in average production rate when load on machine *j* is reduced by one unit. The value ΔG_{MSS}^j is calculated for only those machines whose current load is more than their minimum allowed load. Let's call these machines eligible machines, then ΔG_{MSS}^j for each eligible machine *j* is calculated as:

$$\Delta G_{MSS}^j = \frac{\mu_j \times L_j^{Current}}{\mu_j + \lambda_j(L_j^{Current})} - \frac{\mu_j \times L_j^{Trial}}{\mu_j + \lambda_j(L_j^{Trial})}$$

Step 2: Select the machine that has maximum ΔG_{MSS}^j. Let's assume that machine *k* has the maximum value of ΔG_{MSS}^j.

Step 3: If $G_{MSS}(L^{Current}) - \Delta G_{MSS}^k \geq X_t$,

Then

Update the current load $L_j^{Current}$ for machine *k*, and current average production rate as follows and go back to step 1:

$$L_k^{Current} = L_k^{Trial}$$

$$G_{MSS}(L_k^{Current}) = G_{MSS}(L_k^{Current}) - \Delta G_{MSS}^k$$

Else

Go to step 4:

Step 4: The current machine load L_j^{Curent}, j = 1,2..n is the heuristic optimal solution for problem P4.

At the end of phase 3, we try to minimize the load changing cost by using a simple local search process. In this local search process we look for the improvement of the solution by keeping the equal machine load in two different periods.

The three-phase approach described above determines the complete solution for integrated load distribution production planning problem. The second phase determines the production quantity x_{pt}, the back order quantity B_{pt}, and inventory quantity I_{pt} while the third phase determines the load of machine j for period t.

3.2. Lower-bound Calculation

The optimal solution of the integrated problem can be obtained by solving the integrated problem described by equations (4)-(12). However, the non-linear functions and constraints do not allow the problem to be solved exactly using a known method or software. In order to compare the evaluation of heuristic solution, usually lower-bound value is calculated. The lower-bound value indicates that the optimal value of the problem cannot be less than the lower-bound value. The optimal value of the problem will be always higher than the lower bound value. Usually lower-bound value is obtained by removing some constraints from the problem in such a way that the resultant problem can be solved exactly. The mathematical formulation of our integrated problem has two non-linear functions, $G_{MSS}(\mathrm{L}^t)$ and $\mathrm{RT}(\mathrm{L}^t)$. Therefore, we relax the problem by removing the non-linear functions from the formulation. Since the production planning problem can be solved exactly, we consider the production planning based formulation to find the lower bound while removing the load distribution related variables from the constraints and objective function. The lower-bound formulation can be stated as follows:

Problem LB:

Minimize $$\sum_{p\in P}\sum_{t=1}^{T}\left(h_{pt}I_{pt}+b_{pt}B_{pt}+\pi_{pt}x_{pt}+s_{pt}y_{pt}\right)$$

(32)

Subject to $$I_{pt}-B_{pt}=I_{pt-1}-B_{pt-1}+x_{pt}-d_{pt},\ \ p\in P,\ \ t=1,\ 2,\ ...,\ T,$$

(33)

$$x_{pt}\le\left(\sum_{q\ge t}d_{pq}\right)y_{pt},\ \ p\in P,\ \ t=1,\ 2,\ ...,\ T,$$

(34)

$$\sum_{p\in P}x_{pt}\le G_{MSS}^{Glob-Opt},\ t=1,\ 2,\ ...,\ T,$$

(35)

$$x_{pt},\ I_{pt},\ B_{pt}\in\mathbb{N};\ y_{pt}\in\{0,\ 1\}. \qquad (36)$$

Here $G_{MSS}^{Glob-opt}$ is the average production quantity when available resources are set to infinite. The value $G_{MSS}^{Glob-opt}$ is obtained in step 1 of phase 1 of the proposed three-phase based heuristic. The above problem will be the lower bound for the original formulation because the capacity on production quantity constraint (7) has been relaxed by introducing $G_{MSS}^{Glob-Opt}$ as an upper limit on the production quantity. The introduction of $G_{MSS}^{Glob-Opt}$ in turns makes constraints (8)-(11) irrelevant.

The LB formulation is similar to problem P3 with the only difference being that the production quantity is restricted by $G_{MSS}^{Glob-opt}$ while in problem P3 the production quantity is restricted by G_{MSS}^{Opt}. The proposed lower-bound solution can be improved by formulating the problem just like problem P3. But it is difficult to find the optimal value of G_{MSS}^{Opt} that satisfies the available repair resource constraint. We use heuristic procedure to find the value of G_{MSS}^{Opt} in the first phase of the proposed three-phase based heuristic. Therefore, we cannot use formulations like P3 to calculate the tight lower bound for the integrated problem.

Another way that can be used to tighten the lower bound is by introducing the load change and unused capacity components. However, as mentioned in the beginning, these components are nonlinear and hence it cannot be incorporated in the formulation which can be solved exactly using the known software or method.

4. The proposed tabu search based heuristic

Tabu search (TS) is a meta-heuristic method originally proposed in [28-30]. It consists in an iterative procedure which explores the solution space by moving at each iteration from the current solution s to the best solution in its neighborhood $N(s)$, until some stopping criterion has been satisfied. The neighborhood structure is defined as follows. At each iteration of TS, the local transformations (or moves), that can be applied to the current solution s, define a set of neighboring solutions as: $N(s)$={Solutions obtained by applying a single move to s}. Different types of neighborhood moves have been tested and are explained in subsection 4.2. The TS algorithm starts from an initial feasible solution: subsection 4.1 presents how this initial solution is determined.

At each iteration, the best solution s' in a subset $V(s) \subseteq N(s)$ is selected and considered as a tabu solution for some next iterations. $V(s)$ (referred to as the effective neighborhood) is generated by eliminating the tabu solutions from $N(s)$. Tabus are stored in a short-term memory of the search (tabu list). A previously visited solution is added to the tabu list in order to forbid the repetition of solutions. That is, tabus are used to prevent cycling when moving away from local optima through non-improving moves.

An important feature of our proposed TS is the utilization of a penalty function while allowing infeasible solutions. Solution infeasibility is expressed in terms of exceeding the required repair resources defined by equation (8) in the model formulation (i.e., $RT(\mathbf{L}^t) \leq RT_0$). Consider a solution s and let $C(s)$ denote the objective function value expressed by equation (4). Let $q(s)$ denote the total violation of the required resources. That is, the total violation $q(s)$ is computed as

$\sum_{t=1}^{T} RT(L^t) - RT_0$, where $RT(L^t)$ is total resources used at time period *t* and RT_0 is the available resource. The penalized objective function to be minimized is $C_{penalized}(s) = C(s) + \gamma q(s)$, where γ is a positive weight parameter. In this equation, a weighted penalty for constraint violation is added to the objective function $C(s)$. This penalty function discourages, but allows, the TS algorithm to search into the infeasible boundary region. An interesting way to find correct weights for constraint violations is to use self-adjusting penalties [29]. The value of the weight γ is then dynamically adjusted in our TS implementation. The weight γ is increased if an infeasible solution is encountered, and decreased when a feasible solution is encountered. This technique, known as strategic oscillation, was first introduced in [28] and has been used since in several successful TS procedures. After each iteration the value of γ is modified by a factor $1+\delta$ where $\delta > 0$. If the neighbor solution is found to be feasible with respect to the available resource then the value of γ is divided by $1+\delta$; otherwise it is multiplied by $1+\delta$. This iteration is repeated for η times and the best feasible solution s^* encountered during the search process is reported.

The overall algorithm can be described as follows:

1. Generate an initial feasible solution *s*
2. Set $\gamma = u$, where u is the unitary cost of unused capacity. Set $s^* := s$ and $c(s^*) := c(s)$.
3. While $i \le \eta$ do
   - $i = 1$
   - Choose a neighbor solution $s' \in N(s)$ that minimizes $f(s')$ and store it in the tabu list
   - If solution s' is feasible and $c(s') < c(s^*)$, then set $s^* := s$ and $c(s^*) := c(s)$
   - If $q(s') > 0$ then set $\gamma := \gamma(1+\delta)$; otherwise set $\gamma := \gamma / (1+\delta)$.
   - Increment *i*

End While

4. Report s^* as the final solution.

We set η = 10,000 iterations, penalty factor $\gamma = u$ and penalty multiplication factor δ = 0.5. Thus the proposed algorithm explores the 10,000 neighbor solution. Initially penalty γ for infeasible solution is set to the unitary cost of unused capacity. The value of γ is multiplied by 1.5 if the neighbor solution is feasible, otherwise, it is divided by 1.5. The construction of initial solution and generation of neighborhood solution is described in the next subsections.

4.1. Initial solution

The procedure to find the initial solution uses phase 1 of the three-phase based heuristic proposed in Section 3 to set the machine load. The phase 1 of the algorithm provides, the load of machines L^t for all time periods and G_{MSS}, the average production rate of the multi-state system. The product quantity is set in such a way that total production quantity is less than $G_{MSS}(L^t)$ for time period *t*. We try to allocate production quantity x_{pt} equal to the product demand d_{pt} at time period *t*. First preference is given to the product with the highest inventory holding cost while setting production quantity. If setting x_{pt} equivalent to the product demand d_{pt} violates the total production quantity limit $G_{MSS}(L^t)$ then x_{pt} is set in such a way that the total production quantity is equal to the $G_{MSS}(L^t)$. In this case, production quantity of previous time period is increased in such a way that the inventory of previous periods can be used to satisfy the product demand d_{pt}. There is a possibility of backorder for product demand d_{pt} if the production quantity of previous time period had already reached its maximum allowed production limit. After setting the production quantity x_{pt}, the inventory level I_{pt} and backorder level B_{pt} is determined using the inventory balance formula expressed in equation (5).

4.2. Neighbor solution

As explained before, tabu search explores the solution space by moving at each iteration from the current solution *s* to the best solution in its neighborhood *N*(*s*). We use five different types of neighborhood moves to find the best neighborhood solution.

Neighbor Solution 1: Increase the machine load

In this neighbor solution, the machine load is increased by one unit. We check the increase in load for each machine at each time period *t*. The machine load increase can decrease the average production quantity $G_{MSS}(L^t)$ for time period *t*, which can make the solution infeasible. We consider load increase for only those machines which keep the average production quantity $G_{MSS}(L^t)$ more than the total production quantity $\sum_{p\in P} x_{pt}$. Note that this neighbor solution can produce infeasible solution in terms of required resources exceeding available resources.

Neighbor Solution 2: Increase the machine load and the production quantity

The machine load and the production quantity are increased by one unit each. We check the load increase for each machine at each time period and the production quantity increase for each product. We consider load increase for only those machines that keep the average production quantity $G_{MSS}(L^t)$ more than the total production quantity $\sum_{p\in P} x_{pt}$. This neighbor solution can also produce infeasible solution in terms of required resources exceeding available resources.

Neighbor Solution 3: Reduce the machine load

In this neighbor solution, the machine load is reduced by one unit. We check the increase in load for each machine at each time period *t*. The machine load reduction can decrease the average production quantity $G_{MSS}(L^t)$ for time period *t,* which can make the solution infeasible. Again, we consider a load reduction for only those machines that keep the average production quantity $G_{MSS}(L^t)$ more than the total production quantity $\sum_{p\in P} x_{pt}$.

Neighbor Solution 4: Increase the production quantity

The production quantity is increased by one unit. We check the production quantity increase for each product at each time period. We consider increasing the production quantity only for those products that keep total production quantity $\sum_{p\in P} x_{pt}$ less than the average production quantity $G_{MSS}(L^t)$.

Neighbor Solution 5: Reduce the production quantity

The production quantity is reduced by one unit. We check the production quantity increase for each product at each time period. In this neighbor solution, total production quantity will be less than the average production quantity.

Neighbor Solution 6: Equal load of machine in all time periods

In this neighbor solution, the machine load is kept the same for all time periods. We check each machine at each time period. This neighbor solution will change the average production quantity $G_{MSS}(L^t)$ for all time periods. If the total production quantity $\sum_{p\in P} x_{pt}$ exceeds the average production quantity $G_{MSS}(L^t)$, then the production quantity is reduced to keep the total production quantity below the average production quantity. This is achieved by reducing the production quantity of each product by one unit iteratively.

We define the following five notations for tabu lists (to avoid recycling of neighbor solutions):

1) *TL1*[*i*] makes the load increase for machine *i* tabu,

2) *TL2*[*i*] makes the load reduction for machine *i* tabu,

3) *TL3*[*p*] makes the production quantity increase for product *p* tabu,

4) *TL4*[*p*] makes the production quantity reduction for product *p* tabu, and

5) *TL5*[*i*] makes the consideration of machine *i* for neighbor solution 6 tabu.

Furthermore, we use two terms, θ_1 and θ_2, to make the acceptance of the solution forbidden for the next few iterations. We use tabu list *TL1*[*i*] to check the solution while exploring neighbor solution 1 and neighbor solution 2, *TL2*[*i*] for neighbor solution 3, *TL3*[*p*] for neighbor solution 4, *TL4*[*p*] for neighbor solution 5, and *TL5*[*i*] for neighbor solution 6. When the tabu lists *TL1*[*i*], *TL2*[*i*] and *TL3*[*i*] are updated, they are incremented for the next θ_1 iterations. When tabu lists *TL4*[*p*] and *TL5*[*p*] are updated, they are incremented for the next θ_2 iterations. The value of θ_1 is set to $n \times T \times 0.5$ and the value of θ_2 is set to $n \times p \times 0.5$. The tabu list updating depends on the best neighbor solution obtained by the type of neighbor solution in a given iteration. The tabu list *TL1*[*i*] is updated when neighbor solution 3 obtains the best neighbor solution, *TL1*[*2*] is updated when neighbor solution 1 or neighbor solution 2 obtains the best neighbor solution, *TL3*[*p*] is updated when neighbor solution 5 obtains the best neighbor solution, *TL4*[*p*] is updated when neighbor solution 4 obtains the best neighbor solution and *TL5*[*i*] is updated when neighbor solution 6 obtains the best neighbor solution.

5. Numerical experiments

This section presents the data set generation procedure and numerical analysis to evaluate the performance of the proposed three-phase based and tabu search based heuristics solution.

5.1. Development of problem sets

We generated three types of problem sets on the basis of effectiveness of available resources (i.e., effectiveness of constraint $RT(L^t) \leq RT_0$). The three types of data sets are denoted by setT, setC and setL. The letters T, C and L stand for tight, comfortable and loose. The available resources are very tight and effective in setT while they are not too tight but effective in setC. The available resource constraint is loose and does not impose restriction in setL. Each data set consists of 5 problem sets, A, B, C, D and E where the number of machines varies from 5 to 25.

For each problem set we generate 10 instances. The number of machines and time periods for five problem sets are given in Table 1:

Table 1: Number of machines and time periods for five problem sets.

Problem set	n	T
A1	5	5
A2	5	10
B1	10	5
B2	10	10
C1	15	5
C2	15	10
D1	20	5
D2	20	10
E1	25	5
E2	25	10

Set of product P is set to 2 for all problem sets. Other parameter values related to machines for these problem sets are generated as follows:

1. α_j : Parameter of machine j power law, is generated randomly between U[1.3, 1.8]
2. L_j : Baseline load of machine is set to 30.
3. $\lambda_j(L_{j0})$: Failure rate at base line load is set to vary between 1 % to 4 %, thus, it is generated randomly between U[0.001, 0.004].

4. k_j : Multiplication factor for determining failure rate at load L_j for machine *j* is determined by the equation $k_j = \dfrac{\lambda_j(L_{j0})}{L_{j0}^{\ \alpha_j}}$.

5. μ_j repair rate of machine *j* is set in such a way that the optimal load (i.e., the load at which production rate is maximum at machine *j*) varies between U[10, 150]. As we know that the optimal load for machine *j* is $L_j^{Opt} = \left(\dfrac{\mu_j}{k_j(\alpha_j - 1)}\right)^{1/\alpha_j}$. Thus, we use expression $\mu_j = (U[10,150]^{\alpha_j})(k_j)(\alpha_j - 1)$ to generate the repair rate of machine *j*.

6. $\underline{L_j}$: Minimum allowed load on machine *j* is generated between U[10, 50].

7. $\overline{L_j}$: Maximum allowed load on machine *j* is set above minimum allowed load $\underline{L_j}$ by U[50, 100] i.e., $\overline{L_j}$ = $\underline{L_j}$ +U[50,100].

8. c_{jt} : Cost of changing load distribution for machine *j* from period *t*-1 to period *t* is generated between U[50, 100]

Cost component of product *p* at period *t* are generated as follows:

1. u : Unitary cost of unused capacity is generated between [50, 100].
2. h_{pt} : Inventory holding cost per unit of product *p* is generated between U[25, 75].
3. b_{pt} : Backorder cost per unit of product *p* is set to be 2 to 8 times higher than the inventory holding cost i.e., b_{pt} = U[2, 8]* h_{pt} .
4. s_{pt} : Fixed setup cost of producing one unit of product *p* is generated between U[500, 700].
5. π_{pt} : Variable cost of producing one unit of product *p* is set to be 1.5 to 2 times higher than the inventory holding cost i.e., π_{pt} = U[1.5,2]* h_{pt} .

We use the information of optimal average production rate and corresponding required resources RT to generate available repair resources RT_0 and demand d_{pt}. Let G_{MSS}^{Opt} denote the optimal average production rate and let RT^{Opt} denote the corresponding required repair resources. Then demands of products are set in such a way that the average demand of all the products are G_{MSS}^{Opt} and the maximum and minimum demands are 50 units away from the G_{MSS}^{Opt}. Thus the demand of product p at period t is generated between [G_{MSS}^{Opt} -50, G_{MSS}^{Opt} + 50].

The available resources RT_0 are set in three different ways for three types of data set, setT, setC and setL. Let $RT^{\min}$ represent the minimum resources required which is obtained when the loads of all machines are set at minimum load (i.e, $\overline{L_j}$). Then RT_0 for setT is set to be $RT^{\min}$ + 0.5*(RT^{Opt} - $RT^{\min}$). In setC, RT_0 is set to be $RT^{\min}$ + 0.75*(RT^{Opt} - $RT^{\min}$) while in setL, RT_0 is set to be $RT^{\min}$ + 1.25*(RT^{Opt} - $RT^{\min}$).

For a particular problem instance, all the values for setT, setC and setL are the same except the available resources. In setC, available resources are set exactly between minimum resources and optimal resources, thus, the constraint $RT(L^t) \le RT_0$ becomes too tight. In setC, available resources are set more towards the optimal resources, thus, the constraint $RT(L^t) \le RT_0$ becomes less tight compared to setT. In setL, available resources are set more than optimal resources, thus, the constraint $RT(L^t) \le RT_0$ becomes redundant because in this case available production capacity constraint (i.e., constraint $\sum_{p \in P} x_{pt} \le G_{MSS}$) dominates the available resources constraint (i.e. constraint $RT(L^t) \le RT_0$)

5.2. Numerical results for proposed three-phase based heuristic

The proposed three-phase based heuristic is coded in C and implemented on AMD Opteron 2.3 GHz with 16 GB of RAM. We used CPLEX 12.4.0.0 version to solve exact problem in phase 2 and to calculate lower-bound formulation. CPLEX is an IBM optimization software package which can be used for solving large and difficult mixed-integer programming problems.

Initially, we tested our three-phase based heuristic using the set of problems used in [1]. Our three-phase based heuristic found the optimal solution reported in [1]. Note however that the optimal solution obtained in [1] used small examples, with just two loads for each machine. In order to get the optimal solution using method used in [1], we need to solve a lot sizing problem $\sum_{j=1}^{n}(\overline{L_j}-\underline{L_j})^n$ times. This has restricted obtaining optimal solutions for only small instances. This comparison of optimal solution with our heuristic solution for such small problems does not provide conclusive remarks on performance of the proposed heuristic. Therefore, we compare the solution of our heuristic to lower bound estimate.

Table 2: Average cost for lower bound and three-phase based heuristic solution for setT

Problem set	n	T	Lower Bound Cost	Heuristic Cost	% Gap
A1	5	5	76281.70	79835.99	4.43
A2	5	10	193749.50	210389.20	7.54
B1	10	5	116841.00	121785.22	4.22
B2	10	10	241425.00	260434.90	7.11
C1	15	5	180174.40	191048.70	5.41
C2	15	10	316494.60	341121.60	8.02
D1	20	5	258671.50	267757.40	3.40
D2	20	10	491834.40	519274.60	5.40

E1	25	5	286653.90	301970.90	4.96
E2	25	10	593923.50	615295.00	3.43
Average			275604.95	290891.35	5.39

The lower bound value and the proposed three-phase based heuristic solution values are reported in Table 2. Table 3 reports the average absolute value for three types of data sets, setT, setC and setL and their percentage gap from lower-bound solution. For each problem set the average value of 10 instances is reported in Table 2 and Table 3.

The results reported in Table 2 show that the proposed heuristic is on average 5.39 % away from the lower-bound solution. The average difference of 5.39 % is the average of all 100 problem instances reported in Table 2. The percentage gap for problem setS A1, B1, C1, D1 and E1 is 4.43, 4.22, 5.41, 3.40 and 4.96. This result gives an indication that the proposed heuristic is able to produce consistently good solutions irrespective of the problem size. This consistency is supported by the results obtained for setC and setL as well.

Comparing the results between problem sets A1 &A2, B1 & B2, C1 & C2 and D1 & D2 for setT indicate that when the number of time periods increases the optimal percentage gap increases. These results give an impression that heuristic performance decreases with number of time period increases, however, this is not always true. Comparison of lower-bound percentage gap between problem sets E1 & E2 from setT, B1&B2, E1 & E2 from setC and A1 & A2, C1 & C2 from setL shows improvement of heuristic performance with increase in number of time periods.

Table 3: Comparison of three-phase based heuristic solution for data type setT, setC and setL.

	Average absolute Value			% Gap		
S. N.	setT	setC	setL	setT	setC	setL

A1	79835.99	77605.29	77150.42	4.43	1.76	1.30
A2	210389.20	198990.90	195622.80	7.54	2.68	1.23
B1	121785.22	118449.11	118298.19	4.22	1.50	1.42
B2	260434.90	245158.70	244560.80	7.11	1.46	1.49
C1	191048.70	182287.80	182009.30	5.41	1.14	1.12
C2	341121.60	321211.50	319349.90	8.02	1.57	0.86
D1	267757.40	260074.80	260898.00	3.40	0.56	0.93
D2	519274.60	496489.40	496507.00	5.40	1.00	1.01
E1	301970.90	288407.90	289070.20	4.96	0.63	0.87
E2	615295.00	597358.50	599407.80	3.43	0.59	0.97
Average	290891.35	278603.39	278287.44	5.39	1.29	1.12

The results reported in Table 3 show that the three-phase based heuristic solution is on average 5.39 %, 1.29 % and 1.12 % away from the lower-bound solution for setT, setC and setL respectively. The percentage gap from lower bound for setT is maximum because the available repair resource constraint is too tight. The percentage gap from lower bound for setC is less than setT but greater than setL because the available repair resources are more than setT but less than setL. The percentage gap from lower bound for setL is lowest because the available repair resources constraint becomes redundant for setL. Note that the tightness of our current lower bound depends on gap between production rate for infinite resources and the production rate for given resources (i.e., the gap between $G_{MSS}^{Glob-opt}$ and G_{MSS}^{opt}). When gap between $G_{MSS}^{Glob-opt}$ and G_{MSS}^{opt} increases, calculation of our lower-bound value deteriorates. The gap between $G_{MSS}^{Glob-opt}$ and G_{MSS}^{opt} is zero for setL and therefore the lower-bound calculation for setL is considered to be close to the optimal solution for setL. The percentage gap from lower bound for setL is just 1.12 %. This result shows that our proposed three-phase based heuristic is able to produce the results close to the optimal solution for problem instances varying from 5 to 25 machines.

Other parameters that can affect the lower-bound gap is the value of unused capacity cost u and the cost of changing capacity c_{jt} because these costs are not considered in the lower-bound calculation. The percentage gap presented in Table 3 might be high if more weight is assigned to the unused capacity cost and load distribution changing cost compared to all the other costs. However in our numerical experiment these costs are already high compared to the other costs. Hence, these results will change only when the cost of u and c_{jt} is exceptionally high which might not be the case for most of the real world production planning problem.

5.3. Numerical results for the proposed tabu search based heuristic

The proposed tabu search based heuristic is coded in C and implemented on AMD Opteron 2.3 GHz with 16 GB of RAM. The numerical results are reported in Table 4 and Table 5. We report the average absolute value over 10 problem instances for each problem set. The following abbreviations are used:

1. Initial: Initial solution used in tabu search based heuristic as described in sub-section 4.1.
2. TS: Solution obtained by tabu search based heuristic described in Section 4.
3. HS: Solution obtained by three-phase based heuristic described in Section 3.
4. % IP: Percentage improvement of TS solution over the initial solution.
5. % Dev: Percentage deviation of TS over the solution of the three-phase based heuristic.

The results reported in Table 4 show that the tabu search based heuristic improves the initial solution by 1.9 %, 1.74 % and 1.24 % for problem instances setT, setC and setL, respectively. The percentage improvement is highest for setT while it is lowest for setL.

The results reported in Table 5 show that the percentage deviations of TS over the three-phase based heuristic are -4.10 %, -2.49 % and -2.25 % for setT, setC and setL, respectively. While these negative deviations illustrate the efficiency of the three-phase based heuristic proposed in Section 3, their low values show the efficiency of the tabu search based heuristic as well. The CPU time comparison of three-phase based heuristic and tabu search based heuristic is shown in

Table 6. Note that for both heuristics, and considering all the solved instances, the worst computation time did not exceed 40 seconds.

It is worth mentioning that in the current problem, the machine load and the production quantity are correlated. Such correlation between the machine load and production quantity may create notorious local minima for the solution space. Hence, it becomes difficult to jump out of the local minima once it is trapped inside it. Therefore, the proposed tabu search used 6 different types of neighbor solutions to generate the best neighbor solution (while 10,000 such neighbor solutions are generated).

Table 4: Improvement of TS over the initial solution for problem instances SetT, setC and setL

	SetT			SetC			SetL		
Problem set	Initial	TS	% IP	Initial	TS	% IP	Initial	TS	% IP
A1	88988.94	84230.46	5.35	83844.25	81564.36	2.72	82781.62	81418.87	1.65
A2	228169.30	222921.80	2.30	215389.80	211373.80	1.86	209869.60	206538.20	1.59
B1	130936.68	126372.31	3.49	125889.10	120275.09	4.46	125176.79	123124.49	1.64
B2	276358.10	274385.90	0.71	256056.80	253039.60	1.18	253904.30	250967.10	1.16
C1	207109.50	199688.90	3.58	189536.90	186254.80	1.73	188153.30	185480.40	1.42
C2	373783.80	369099.90	1.25	338448.90	331661.50	2.01	334008.30	329839.10	1.25
D1	271149.80	270618.60	0.20	263067.20	261199.90	0.71	263111.70	260969.70	0.81

D2	547783.90	538483.80	1.70	509032.30	503844.60	1.02	506747.50	501472.20	1.04
E1	307464.70	307005.90	0.15	292251.30	290660.10	0.54	292712.40	290068.60	0.90
E2	624669.40	623203.70	0.23	608671.10	601631.20	1.16	609982.60	604497.30	0.90
Average	305641.41	301601.13	1.90	288218.77	284150.50	1.74	286644.81	283437.60	1.24

Table 5: Comparison of TS with HS solution for problem instances SetT, setC and setL

	SetT			SetC			SetL		
Problem set	HS	TS	% Dev	HS	TS	% Dev	HS	TS	% Dev
A1	79835.99	84230.46	-5.50	77605.29	81564.36	-5.10	77150.42	81418.87	-5.53
A2	210389.20	222921.80	-5.96	198990.90	211373.80	-6.22	195622.80	206538.20	-5.58
B1	121785.22	126372.31	-3.77	118449.11	120275.09	-1.54	118298.19	123124.49	-4.08
B2	260434.90	274385.90	-5.36	245158.70	253039.60	-3.21	244560.80	250967.10	-2.62
C1	191048.70	199688.90	-4.52	182287.80	186254.80	-2.18	182009.30	185480.40	-1.91
C2	341121.60	369099.90	-8.20	321211.50	331661.50	-3.25	319349.90	329839.10	-3.28
D1	267757.40	270618.60	-1.07	260074.80	261199.90	-0.43	260898.00	260969.70	-0.03
D2	519274.60	538483.80	-3.70	496489.40	503844.60	-1.48	496507.00	501472.20	-1.00

E1	301970.90	307005.90	-1.67	288407.90	290660.10	-0.78	289070.20	290068.60	-0.35
E2	615295.00	623203.70	-1.29	597358.50	601631.20	-0.72	599407.80	604497.30	-0.85
Average	290891.35	301601.13	-4.10	278603.39	284150.50	-2.49	278287.44	283437.60	-2.52

Table 6: CPU time (in seconds) for TS and HS

	Set T		Set C		Set L	
Problem set	HS	TS	HS	TS	HS	TS
A1	< 1	1	< 1	1	< 1	1.3
A2	< 1	6.2	< 1	6.2	< 1	5.5
B1	< 1	3	< 1	3.1	< 1	2.6
B2	< 1	10.8	< 1	12.7	< 1	12
C1	< 1	5.5	< 1	5.2	< 1	4.6
C2	< 1	18.4	< 1	18.2	< 1	19.1
D1	< 1	7.6	< 1	7.7	< 1	7.1
D2	< 1	27.9	< 1	27.8	< 1	22.7
E1	< 1	11	< 1	10.1	< 1	9.4
E2	< 1	39.1	< 1	38.6	< 1	39
Average	<1	13.05	<1	13.06	<1	12.33

6. Conclusion

This paper considers production planning problem with unreliable machine in the context of multi-state system. The problem is important because it considers the load versus failure rate relationship while optimizing the production planning system. The proposed model is motivated by the fact that combining load distribution decisions with production planning may reduce the total expected cost. The objective of the production planning is to minimize the total production cost while the objective of the load distribution planning is to maximize the production planning rate which might lead to the high cost of unused capacity. If the load distribution and production planning decisions are performed sequentially then it might lead towards non-optimal decisions. Therefore, the integrated load distribution and production planning decisions are considered to find the best tradeoff between the load distribution and production planning.

We propose two heuristics to solve the problem. The first heuristic consists in a three-phase based approach, while the second is based on tabu search metaheuristic. The effectiveness of each proposed heuristic is evaluated by solving the randomly generated problem instances. It is found that the two heuristics are efficient in terms of solution quality and computation time. Although the three-phase based heuristic provides solutions that are slightly superior, it is worth mentioning that another purpose of introducing tabu search based heuristic is to provide a framework for its implementation. Within this framework of tabu search based heuristic, we are currently improving our tabu search based heuristic to efficiently solve large scale problems under high computation time constraints. On the one hand, introducing exact algorithm to solve the production planning problem is an important step towards improvement of tabu search based heuristic. On the other hand, new types of neighbor solutions should be tested. In the current tabu search based heuristic, we use the neighbor solution which changes the machine load for only one machine (except neighbor solution 6) and production quantity for only one product. A new neighbor solution can be designed to change the load of two machines at a time. Similarly a new neighbor solution can be designed to change the two products at a time. Finally, the proposed heuristics are used for systems with parallel machines. We intend to extend this work to deal with other configurations, such as series-parallel, networks and buffered systems.

Acknowledgement

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Asper School of Business at University of Manitoba (Canada).

References

[1] M. Nourelfath and F. Yalaoui, “Integrated load distribution and production planning in series-parallel multi-state systems with failure rate depending on load”, *Reliability Engineering and System Safety*. vol. 106, pp. 138–145, 2012.

[2] R.K Iyer and D.P. Rosetti, “A measurement-based model for workload dependency of CPU errors,” *IEEE Trans. Comput.*, vol. C-35, pp. 511–9, 1986.

[3] G. Levitin and S.V. Amari, “Optimal load distribution in series-parallel systems,” *Reliability Engineering and System Safety,* vol. 94, pp. 254–260, 2009.

[4] W. Nelson, *Accelerated testing: statistical models, test plans, and data analysts*: New York-Wiley, 1990.

[5] M.C. Pike, “A method of analysis of a certain class of experiments in carcinogenesis,” *Biometrics*, vol. 22, pp. 142–61, 1966.

[6] A. Lisnianski and G. Levitin, Multi-state system reliability: Assessment, optimization and applications: *World Scientific*, 2003.

[7] G. Levitin, Universal generating function in reliability analysis and optimization: Berlin-Springer, 2005.

[8] J. F. Shapiro, “Mathematical programming models and methods for production planning and scheduling,” in *Handbooks in Operations Research and Management Science, Volume 4, Logistics of Production and Inventory*, S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin, Eds: North-Holland, 1993.

[9] D. Sipper and R. Bulfin, *Production: planning, control and integration*: McGraw-Hill, 1997.

[10] L.A. Wolsey, “Solving multi-item lot sizing problems with MIP solver using classification and reformulation,” *Management Science*, vol. 48, no. 12, pp. 1587–1602, 2002.

[11] J. Filus, “A problem in reliability optimization,” J. Oper. Res. Soc., vol. 37, pp. 407–12, 1986.

[12] W. Kuo and M.J. Zuo, *Optimal reliability modeling*: New York-Wiley, 2003.

[13] D.A. Collier, "A comparison of MRP lot sizing methods considering capacity change costs," *Journal of Operations Management*, vol. 1, no. 1, pp. 23-29, 1980.

[14] Z. Sebestyén and V. Juhasz, "The impact of the cost of unused capacity on production planning of flexible manufacturing systems," *Periodica Polytechnica Ser. Soc. Management Scence*, vol. 11, no. 2, pp. 185–200, 2003.

[15] S. Eiamkanchanalai and A. Banerjee, "Production lot sizing with variable production rate and explicit idle capacity cost," *International Journal of Production Economics*, vol. 59, no. 1-3, pp. 251-259, 1999.

[16] M. Nourelfath, "Service level robustness in stochastic production planning under random machine breakdowns," *European Journal of Operational Research*, vol. 212, no. 1, pp. 81-88, 2011.

[17] I. Maatouk, E. Chatelet and N. Chebbo, "Availability Maximization and Cost Study in Multi-State Systems", *Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual,* pp. 1-6, 2013.

[18] A. Chamseddine, D. Theilliol, I. Sadeghzadeh, Y. Zhang and P. Weber, " Optimal reliability design for over-actuated systems based on the MIT rule: Application to an octocopter helicopter testbed", *Reliability Engineering and System Safety,* vol. 132, pp. 196-206, 2014.

[19] G. Levitin, L. Xing and Y. Dai, "Optimal component loading in 1-out-of-N cold standby systems", *Reliability Engineering and System Safety,* vol. 127, pp. 58-64, 2014.

[20] Y. Gajpal and M. Nourelfath, "*A heuristic approach to solve production planning problem with machine failure in a series parallel multi-state system*", Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management Bali, Indonesia, January 7 – 9, 2014

[21] J.K. Filus and L. Z. Filus, "Parameter Dependence in Stochastic Modelling – Multivariate Distributions", *Applied Mathematics,* vol. 5, pp. 928-940, 2014.

[22] K.A. Francie, K. Jean-Pierre, D. Pierre, S. Victor and P. Vladimir, " Stochastic optimal control of manufacturing systems under production-dependent failure rates", *International Journal of Production Economics*, vol. 150, pp. 174-187, 2014.

[23] R. Jans and Z. Degraeve, "Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches" *European Journal of Operational Research*, vol. 177, pp. 1855–1875, 2007.

[24] A. A. Menezes, A. Clark and B. Almada-Lobo, "Capacitated lot-sizing and scheduling with sequence-dependent, period-overlapping and non-triangular setups", *Journal of Scheduling*, vol. 14, pp. 209–219, 2011.

[25] C. Almeder D. Klabjan, R. Traxler and B. Almada-Lobo, "Lead time considerations for the multi-level capacitated lot-sizing problem", *European Journal of Operational Research*, vol. 241, Issue 3, pp. 727–738, 2015.

[26] M. Zhang, S. Kucukyavuz and H. Yaman "A Polyhedral Study of Multiechelon Lot Sizing with Intermediate Demands", *Operations Research,* vol. 60 issue 4, pp. 918-935, 2012.

[27] D. Gade and S. Küçükyavuz "Formulations for dynamic lot sizing with service levels", *Naval Research Logistics,* vol. 60, Issue 2, pp. 87–101, 2013.

[28] F. Glover, "Heuristics for integer programming using surrogate constraints", *Decision Science*, vol. 8, pp. 156–66, 1977.

[29] F. Glover, "Future paths for integer programming and links to artificial intelligence", *Computer and Operations Research*, vol. 13, pp. 533–49, 1986.

[30] F. Glover and M. Laguna, "Tabu search", Boston: Kluwer Academic Publishers; 1997.

Highlights

- The expected performance of the system is a non-monotonic function of its load.
- We consider the integration of load distribution and production planning decisions.
- The paper proposes three phase and tabu search based heuristics to solve the problem.
- Lower bound has been developed for checking the effectiveness of the heuristics.
- The efficiency of the heuristic is tested through randomly generated instances.