7-2010

Exploiting metaheuristics to strategize on performance-based logistics contracts for MRO services

Arnd SCHIRRMANN

Elaine WONG

Zhichao ZHENG
Singapore Management University, DANIELZHENG@smu.edu.sg

Exploiting Metaheuristics to Strategize on Performance-Based Logistics Contracts for MRO Services

Conference Paper · July 2010

3 authors, including:

Arnd Schirrmann
Airbus Group Innovations, Hamburg

Zhichao Zheng
Singapore Management University

Exploiting Metaheuristics to Strategize on Performance-Based Logistics Contracts for MRO Services

Arnd Schirrmann *· Elaine Wong †· Zhichao Zheng ‡

June 3, 2010

Abstract

An inherent challenge of using Performance-Based Logistics (PBL) contracts for aircraft maintenance, repair and overhaul (MRO) services is pricing. As with traditional bricks and mortar services, under-priced contracts cannot cover costs, while overpriced contracts lose out to competition. Furthermore MRO services have an additional element of uncertainty. Performance uncertainties arise due to the inability to accurately forecast demand of spare parts, while cost uncertainties are a result of globally distributed operations subjected to fluctuating economic conditions. Previous work to solve this contracting problem adopted the principal-agent model, obtaining an optimal solution from the perspective of both risk-averse parties (i.e., a price-sensitive customer and a profit-driven service provider). This work presents a model that extends existing models by incorporating integrality constraints. Using the metaheuristics, Simulated Annealing algorithm, we show how the new non-linear mixed integer programming model can be efficiently solved by implementing appropriate algorithms for (a) initial solution derivation, (b) next solution (neighbour) generation, and (c) worse solution acceptance criterion. The algorithm has been tested with real operational data and a graphical representation of the results will be provided and analyzed.

Keywords: Metaheuristics; Programming, Integer ; Inventory Management ; Airline Applications;

1 Introduction

Heighten competition in the commercial aviation industry has spurred airlines to outsource maintenance, repair and overhaul (MRO) activities in an effort to streamline operations. Expectedly, such a move would not only bring to the airlines cost savings, but also increased fleet utilization and reduced logistics delays. On the other hand, outsourcing this vital function

*European Aeronautic Defence and Space Company Innovation Works.

†European Aeronautic Defence and Space Company Innovation Works. Email: elaine.wong.kl@eads.net

‡Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore 117591. Email: zhichao@nus.edu.sg

has unavoidably expose both parties (airlines and MRO service provider) to greater risks as a result of (a) complexity of performance-based contracts (contractual terms include a fixed price component, a penalty term as well as an optional cost sharing), (b) unpredictability of attainable service levels due to unscheduled spares parts demand (due to the sporadic nature of failures), and (c) uncertain economic outlook for the duration of the contract with significant impact on operating costs.

There have been notable attempts to model and determine risk-management strategies for performance-based contracts (PBC) by applying the contract theory. A contract theory problem is built on the principal-agent model with the aim to analytically determine the optimal incentives in face of adverse selection (or hidden information) and moral hazard (or hidden action) (Bolton & Dewatripont (2005)). The challenges in applying contract theory on performance-based contracting problems can be broadly classified into two groups: modeling and computation. Modeling challenges include identification of control variables (i.e., what are the decision variables and who sets them?), nature of interactions (i.e., who initiates the contracts and is negotiation allowed?), quantification of risks, and prediction of the other party's responses. Computational aspect, although given less attention, also poses significant challenges, namely general nonconvex objective, nonlinear constraints, and integer decision variables. The inclusion of integer variables (overlooked in current works) are necessary for representing inventory levels. The drawback of mixed-integer programming (MIP) is the optimization problem becomes nonconvex and combinatorial in nature, implying (i) local optimal solutions complicate the search for a global optimal solution , and (ii) the time required to determine an optimal solution increases exponentially with the number of variables. Addressing either challenge has led to the compromise of the other. For example employing the branch and bound method, though able to accurately determine the global optimal solution, is computationally expensive. When supporting strategic business decision making, time-consuming methods (notwithstanding quality of results) is impractical and ineffective.

Metaheuristics have been found to derive good approximation of the global optimal solution for MIP problems efficiently (Osman & Kelly (1996)). A distinguishing feature of metaheuristics is its adaptation of processes observed in physical, biological and statistical sciences. And successful exploitation of the latter determines the success of metaheuristics to solve difficult optimization problems. One class of metaheuristics is the simulated annealing algorithm, which unlike other metaheuristics, offers flexibility in deployment (sequential or parallelized), nonconvexity tolerance tuning (acceptance threshold for worse solutions), and convergence criterion (specify when a solution is good enough).The flexibility provided by simulated annealing makes it suited for solving complex MIP problems of varying scale, such as the contracting problem for performance-based MRO services with integer constraints enforced. The remaining of this paper presents our exploitation of simulated annealing algorithm and the assessment of its performance.

1.1 Literature Review

Optimizing spare parts allocation from a MRO service contract perspective originated from Sherbrooke's system-approach as opposed to then commonly used item-approach (Sherbrooke (1968)). Sherbrooke proposed and proved that a system is most cost-effective when spare parts allocation and system availability is considered on for the entire set of parts. There have been several extensions to the mathematical model for optimizing spare parts allocation, for example considering a multi-echelon system (Muckstadt (1973)), using of emergency transshipment (Axsäter (1990)), etc. For the interest of space, we will not detail the works here but refer readers to the excellent review paper by Kennedy, et al. (2002).

Apart from the MRO research, the idea of performance-based contracting (PBC) has become increasingly popular in many fields, like public transportation (Hensher & Stanley (2002)), healthcare services (Lu, et al. (2003)), etc. The primary objective of PBC is to change the behavior of contractors to focus more on performance and thus to alleviate the moral hazard problem. Some assessment on the effectiveness of PBC can be found in Lawrence (2005) and Petersen, et al. (2006).

The first attempt to integrate spare parts allocation optimization with PBC was presented by Kim, et al. (2007). Based on the original mathematical model for a single location, the work applied contract theory (based on the classical principal-agent model), to solve optimal contractual parameters for a performance-based logistics (PBL) contract. The focus of the work was to address modeling challenges, specifically, (a) defining service provider's control parameters to include not only spares allocation, but also a private cost reduction effort, (b) quantifying risks using Markovitz theorem, and (c) predicting service provider's strategy by determining the optimal turning point based on first and second derivatives. Some attention was also given to the computational aspect, namely the use of Lagrangian relaxation technique to handle nonlinear constraints. However, the incorporation of integer constraints have not been considered. Nowicki, et al. (2008) developed an optimization model for spares provisioning under a multi-item, multi-echelon scenario, where the objective of the optimization model is to maximize the profit to the supplier under a PBL contract. The authors demonstrated that one can not guarantee an optimal collection of spares without considering the associated revenue stream. However, the issue of contract design is not studied in that paper.

Computationally, spare parts allocation problems have been largely solved using classical optimization techniques, for example, Lagrangian multipliers method, subgradient optimization, Markov chain type recursive computing, etc. In recent years, metaheuristics such as genetic algorithms (GA) and simulated annealing (SA) have also been applied to allocation problems. For example, Batchoun, et al. (2003) attempted to use GA to determine the optimal allocation of aircraft spare parts that minimizes the cost of delay caused by unexpected failure. Ilgin & Tunali (2007) went a step further to apply GA the optimization of spare parts inventory and maintenance policies. Many other industrial application problems have also been solved using metaheuristics, for example, scheduling in manufacturing, transportation routes, multi-factor design. A complete review and bibliography of metaheuristics can be found in Osman & Laporte (1996). Besides GA, the use of SA established by Kirkpatrick, et al. (1983), has also been widely applied in many optimization problems (cf. Haddock & Mittenthal (1992), Ganesh & Punniyamoorthy (2004)). However, it is demonstrated that only with carefully tuned parameters (Granville, et al. (1994)) can SA converge to the global optimum.

1.2 Contributions

This objective of this work is to exploit the flexibility of the simulated annealing algorithm to efficiently solve the contracting problem for performance-based logistics contracts for MRO services. The mathematical model is an extension of Kim et al Kim, et al. (2007) incorporating multiple part essentialities and target service levels, and scrapping of spare parts in addition to repairs. We apply the model to a commercial aviation scenario and investigate different strategies to improve on the simulated annealing algorithm's convergence, namely (a) choice of initial solution, (b) generation of next solutions, and (c) acceptance of worse solutions. Finally, to demonstrate the use of the algorithm, we apply the algorithm on a set of possible business scenarios and discuss managerial insights drawn from the results.

2 Model Formulation

This section first gives an overview of the problem, and then proceeds to define the notations and assumptions adopted for the model parameters, and finally describes the exact formulation of the problem with additional constraints to be considered.

2.1 Problem Description - A Principal-Agent Framework

The contracting problem builds on a principal-agent model. The interaction between the two parties resembles a Stackelberg game, with the principal (customer) initiating the contract offer to the agent (MRO service provider). We assume no negotiations, so if the service provider agrees to the terms of contract, the service will be performed with a deliberated effort, and measured according to a performance metric. It is assumed that the service provider is a rational decision maker, and acts to maximize her own utility (i.e., minimize disutility). Also, the actions (i.e., spares allocation and cost reduction effort) of the service provider is hidden from the customer, hence the onus lies in the customer to design a contract that incentivizes the agent to exert effort level that will maximizes her own utility. The sequence of the events can be summarized as follows:

- Step 1: The customer offers the service provider take-it-or-leave-it contract.

- Step 2: The service provider accepts or rejects the contract. If the service provider rejects the contract, the process stops here. Otherwise,
  - Step 3: The service provider decides on the cost reduction actions and the base stock levels for the spare parts.
  - Step 4: The service provider realizes operating costs and performance metrics are evaluated at the end of the period.
  - Step 5: Payment is made according to the contract terms.

Next, we proceed to define the notations and modelling assumptions and mathematically formulate the above problem.

2.2 Notations and Assumptions

We consider the case where an MRO service is provided for a commercial airlines with a fleet of $\mathcal{N}$ aircrafts operating an annual fleet flight hours, $\mathcal{H}$. The contract covers a total of n spare parts ($\mathcal{P} = \{p_1, ..., p_n\}$). We use i to index the spare parts. We focus on unscheduled removals which would account for the unpredictable demand of spare parts.

2.2.1 Part Specifications

The following notations will be used for the specifications of spare parts.

τ_i	Mean time between unscheduled removals of p_i. Failure of p_i is assumed to occur at a Poisson rate $1/\tau_i$, independently from failures of other parts.
c_i	Unit price of p_i
$\mathcal{Q}_i$	Number of units of p_i installed on an aircraft (i.e., Quantity per aircraft (QPA).)
ξ_i	Spare part class (SPC) code of p_i Parts with different SPC codes will undergo different MRO processes. (a) $\xi_i = 1$: The unserviceable part must be scrapped (Expendable) (b) $\xi_i = 2$: The unserviceable part is sent for repair and returned to inventory upon completion (Rotable) (c) $\xi_i = 6$: Based on the condition of the unserviceable part, it might be scrapped with a certain probability (scrap rate, κ_i), or otherwise repaired before returning to inventory (Repairable)
η_i	Essentiality (ESS) code of p_i (a) $\eta_i = 1$: A serviceable part is required for aircraft to fly (NO-GO) (b) $\eta_i = 2$: Based on other aircraft/flight conditions, the aircraft might be allowed to fly even with an unserviceable part (GO-IF) (c) $\eta_i = 3$: Aircraft is allowed to fly even with an unserviceable part (GO)
$\mathcal{F}_i$	Number of failures per year of part i Following the assumption on τ_i, $\mathcal{F}_i$ is a Poisson random variable with mean $\mu_i = N \cdot \mathcal{H} \cdot Q_i/\tau_i$

2.2.2 MRO Process

We assume that the MRO service provider employs a one-for-one base stock policy, where an unserviceable part is immediately replaced by a serviceable part from the supplier's inventory. If a replacement is unavailable, a backorder occurs. The replenishment of the used inventory depends on the spare part class code. An unserviceable rotable/repairable will be repaired while an expendable/repairable will be replenished with a new one so as to bring the inventory position back to the base stock levels. Assuming that the repair and repurchase times are independent of each other, and the repair facility and new parts supplier has ample capacity, the number of unscheduled removals arising during the replenishment duration is equivalent to the occupancy of an $M/G/\infty$ queue. Based on Palm's theorem (Feeney & Sherbrooke (1966)) we can use Poisson to represent the steady-state distribution of the unscheduled removals. The notations used to describe the MRO process is provided below.

μ_i	Average number of unscheduled removals per year for p_i $\mu_i = \mathcal{N} \cdot \mathcal{H} \cdot \mathcal{Q}_i / \tau_i$
t_i^p	Purchase lead time of p_i
t_i^a	Average administration time for purchasing p_i
t_i^r	Average repair turn around time of p_i
t_i^t	Average transit time for repairing p_i
t_i	Average replenishment lead time of p_i (a) If $\xi_i = 1$, then $t_i = t_i^p + t_i^a$ (b) If $\xi_i = 2$, then $t_i = t_i^r + t_i^t$ (c) If $\xi_i = 6$, then $t_i = \kappa_i \left(t_i^p + t_i^a\right) + (1 - \kappa_i)\left(t_i^r + t_i^t\right)$
C_i	Expected compensation for each backorder of p_i
B_i	Number of backorders expected for p_i B_i is a random variable observed at a random point in time after steady state is reached. B_i depends on the probability that the number of unscheduled removals during the replenishment interval exceeds the inventory level.
s_i	Spare stocking level of p_i
O_i	Pipeline (on-order) inventory for p_i O_i is a stationary random variable, and it links B_i and s_i through $B_i = (O_i - s_i)^+$ Palm's Theorem states that O_i is Poisson distributed for any lead time distribution with mean $\mu_i t_i$ (cf. Feeney & Sherbrooke (1966)).

Similar to Kim, et al. (2007), we use normal approximation for the Poisson distributions of B_i and O_i in the model derivation, which is well-justified in that paper.

2.2.3 Cost Structure

A simplified MRO service operating cost (c_i^t) is assumed, consisting of inventory holding costs and replenishment costs. Inventory holding costs cover capital costs, overheads (such as insurance, obsolescence, depreciation and taxes), labour costs, and warehouse space leasing costs. Replenishment costs covers transportation, tests and repairs, and administrative overheads. The cost is expressed as

$$c_i^t = c_i^h s_i + (1 - a_i) c_i^r \mathcal{F}_i + \epsilon_i \tag{1}$$

where

c_i^h	Average annual holding cost of p_i
c_i^r	Average repair cost per event of p_i
s_i	Spare part inventory level for p_i
a_i	Supplier's discretionary (cost reduction) effort for p_i By exerting such effort, the supplier can reduce the repair cost, and the effective repair cost of p_i would be $(1 - a_i)c_i^r$. Furthermore, we assume that such effort is unobservable to the customer, hence not contractable.
$\psi(a_i)$	Supplier's disutility for exerting a_i $\psi(a_i)$ is increasing convex, i.e., $\psi'(a_i) > 0$, $\psi''(a_i) > 0$. In this paper, we assume a quadratic form $\psi(a_i) = k_i a_i^2/2$ with $k_i > 0$.
ϵ_i	Additive cost variance of p_i ϵ_i is a random variable with assumed zero mean and finite variance, and it is independent between different parts. Furthermore, we assume it is independent from backorders.

2.2.4 Contract Value

The MRO service contract pricing model comprises an annual fixed payment (ω), a variable component based on a predefined share of the service provider's cost (α), and a penalty for each backorder encountered (ν). The total contractual value is expressed as

$$\mathcal{V} = \omega + \alpha \sum_{i=1}^{n} c_i^t - \nu \sum_{i=1}^{n} B_i \tag{2}$$

2.2.5 Expected Utility

We assume that the contracting parties are risk-averse, and the utility function takes the form of an expected mean-variance utility expressed as $\mathbf{E}[U(X)] = \mathbf{E}[X] - rVar(X)/2$, where X denotes a lottery (i.e., a random payoff), and $U(X)$ is the utility of X, and $r \geq 0$ is the risk aversion factor (a larger r reflects greater risk-averseness). The service provider and customer's utility function can thus be expressed as:

$$\begin{aligned} \mathbf{E}[U_s(\mathcal{V} - \sum_{i=1}^{n} c_i^t) - \psi(\boldsymbol{a}) | \boldsymbol{a}, \boldsymbol{s}] = \ & \mathcal{V} - \sum_{i=1}^{n} c_i^t - \sum_{i=1}^{n} \psi(a_i) \\ & - \tfrac{1}{2} r_s \nu^2 Var(B_i | s_i) - \tfrac{1}{2} r_s (1 - \alpha)^2 Var(\epsilon_i), \end{aligned} \tag{3}$$

and

$$\mathbf{E}[U_c(\mathcal{V})|\boldsymbol{a},\boldsymbol{s}] = \quad -\mathcal{V} - \tfrac{1}{2}r_c(C_i-\nu)^2 Var(B_i|s_i) - \tfrac{1}{2}r_c\alpha^2 Var(\epsilon_i), \tag{4}$$

respectively, where r_s is the risk-aversion factor for the supplier and r_c is the risk-aversion factor for the customer, and capital lowercase letters are used to represent the vectors.

Both parties are concerned with the expected contractual value ($\mathcal{V}$), and risks associated with performance (second last term) as well as cost variations (last term). In addition, the service provider is also affected by the operating costs (c_i^t) and cost-reduction effort ($\psi(a_i)$).

2.3 Problem Formulation

With the notations defined above, we obtain the mathematical formulation of the principal's problem described at the beginning of this section:

$$\begin{array}{ll} \min_{\omega,\alpha,\nu} & -\mathbf{E}[U_c(\mathcal{V})|\boldsymbol{a}^*,\boldsymbol{s}^*] \\ s.t. & (\boldsymbol{a}^*,\boldsymbol{s}^*) \in \arg\min \ -\mathbf{E}[U_s(\mathcal{V}-\sum_{i=1}^{n} c_i^t) - \psi(\boldsymbol{a})|\boldsymbol{a}\geq \mathbf{0},\boldsymbol{s}\geq\mathbf{0}] \\ & 0\leq\alpha\leq 1,\ \omega\geq 0,\ \nu\geq 0 \end{array} \tag{5}$$

where the first constraint corresponds to the agent's problem that maximizes her own utility. This is called an incentive compatibility (IC) constraint, which will be discussed with two additional constraints in the following sections.

2.3.1 Essentiality-Based Performance (EP)

The set of n parts consists of three subsets, grouped according to the essentiality code ($\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_3$), where $\mathcal{P}_j$ consists of p_i with $\eta_i = j$. For each set $\mathcal{P}_j$, the performance of the service is measured according to the fill rate (β_j), which is the percentage of part replacements that can be fulfilled. We define fill rate as

$$\beta_j = \frac{\sum_{p_i\in\mathcal{P}_j}(\mu_i \times F_i(s_i))}{\sum_{p_i\in\mathcal{P}_j}\mu_i} \tag{6}$$

where $F_i(s_i)$ is the estimated fill rate of p_i when the stocking level is s_i. The essentiality-based performance (EP) constraint can thus be expressed as

$$\beta_j \geq \hat{\beta}_j,\ j = 1,2,3 \tag{EP}$$

where $\hat{\beta}_j$ is the target fill rate for $\mathcal{P}_1$ parts.

2.3.2 Individual Rationality (IR)

A rational service provider would not only maximize her utility, but also ensure that the contract offer is advantageous. Without loss of generality, we assume the supplier's reservation utility (corresponding to exogenous opportunity) is normalized to zero. Then the rationality constraint is defined as

$$\mathbf{E}[U_s(\mathcal{V} - \sum_{i=1}^{n} c_i^t) - \psi(\boldsymbol{a})|\boldsymbol{a}, \boldsymbol{s}] \geq 0 \tag{IR}$$

Assuming that it is beneficial for the principal to keep the contract value as low as possible, it necessitates the above constraint to bind at the optimum, i.e., the customer will choose an annual fixed payment, ω^* that will leave the supplier with exactly zero expected utility.

$$\omega^* = [(1-\alpha)\sum_{i=1}^{n} c_i^t + \nu\sum_{i=1}^{n} B_i + \sum_{i=1}^{n}\psi(a_i) + \frac{1}{2}r_s\nu^2 Var(B_i|s_i) + \frac{1}{2}r_s(1-\alpha)^2 Var(\epsilon_i)]^+, \tag{IRa}$$

where $(x)^+ = \max\{0, x\}$, and the last two terms represent the risk premiums to be paid by the customer.

2.3.3 Incentive Compatibility (IC)

The incentive compatibility (IC) constraint defined in equation (5) can be assumed to yield unique interior optimal solution under some mild conditions: $\alpha < 1$ and $\nu[1 - F_i(0)] \geq (1-\alpha)c_i$ [1]. Moreover, $(\boldsymbol{a}^*, \boldsymbol{s}^*)$ can be expressed as:

$$a_i^* = \frac{(1-\alpha)c_i^r \mu_i}{k_i} \tag{ICa}$$

$$\nu\sum_{i=1}^{n}[1 - F_i(s_i^*)] + r_s\nu^2\sum_{i=1}^{n}(F_i(s_i^*)\mathbf{E}[B_i|s_i^*]) = (1-\alpha)\sum_{i=1}^{n} c_i^h \tag{ICb}$$

Instead of deriving s_i^*, it is more convenient to use equation (ICb) to determine ν, which yields

$$\nu = \begin{cases} \dfrac{(1-\alpha)\sum_{i=1}^{n} c_i^h}{\sum_{i=1}^{n}\left[1-F_i(s_i^*)\right]} & \text{if } r_s = 0 \\ \dfrac{-\sum_{i=1}^{n}\left[1-F_i(s_i^*)\right] + \sqrt{\left(\sum_{i=1}^{n}\left[1-F_i(s_i^*)\right]\right)^2 + 4r_s\sum_{i=1}^{n}\left(F_i(s_i^*)\mathbf{E}[B_i|s_i^*]\right)(1-\alpha)\sum_{i=1}^{n} c_i^h}}{2r_s\sum_{i=1}^{n}\left(F_i(s_i^*)\mathbf{E}[B_i|s_i^*]\right)} & \text{if } r_s > 0 \end{cases} \tag{ICc}$$

With these additional constraints, the contracting problem (5) can be reduced to an optimization problem with only two decision variables, α and $\boldsymbol{s}$.

[1]The conditions will be violated only under extreme parameter settings, e.g., when the customer's risk aversion is orders of magnitude greater than that of the supplier (cf. Kim, et al. (2007)). Hence, it is generally safe to assume these conditions are satisfied, which is the case under our model settings.

2.3.4 Integrality (IG)

Integrality constraints have been introduced for both variables. As discussed earlier, representing inventory levels as integers is unequivocal. In addition, we propose a discretization of the cost sharing parameter (α).

$$\boldsymbol{s}_i \in \mathbb{Z} \quad \boldsymbol{s}_i \geq 0 \tag{IGa}$$

$$\alpha = \tilde{\alpha}/100, \tilde{\alpha} \in \mathbb{Z} \quad 0 \leq \tilde{\alpha} \leq 100 \tag{IGb}$$

Discretizing α accomplishes two things. Firstly, the granularity of α becomes limited, making the solution more realizable. For example having a share of cost equals to 15% as opposed to 14.391%. Secondly, the original MIP problem is transformed into a pure IP problem, which is comparatively easier to solve using our algorithm.

Till now, we have defined the complete problem, which is Problem 5 with (EP), (IR), (IC) and (IG) constraints. Unfortunately, the objective function of Problem (5) is generally not quasiconvex and hence not necessarily unimodal. The analytical specification of $\boldsymbol{s}^*$ is intractable even when α is fixed (cf. Kim, et al. (2007)). Furthermore, the size of the problem could be extremely large in the context of aircraft spares management. In view of these difficulties, we thus explored the use of metaheuristics to numerically solve the problem.

3 The Algorithm

Figure 1 summarizes the flow of the simulated annealing algorithm used to solve the contracting problem. We focus our attention on tuning four points in the flow (marked with * in the flow chart).

3.1 Simulated Annealing - Cooling Scheme

3.1.1 Initial Solution — [ew] to be modified..

The performance of simulation annealing, like other metaheuristics based on random searches, depends on the initialization procedure. Providing a good initial solution would improve the convergence of the algorithm. The challenge however is ensuring that the solution is feasible. Individual rationality (IR), incentive compatibility (IC) and integrality have been discussed in the previous section. To meet the essentiality performance (EP) constraint, a brute-force approach is to set $s_i = \hat{S}$, where $\hat{S}$ is an arbitrary large number. However the convergence for such methods would be slow. We propose a process to derive a better feasible initial solution.

Proposition 1 *Given set of n spare parts ($\mathcal{P} = \{p_1, ..., p_n\}$) consisting of three subsets grouped according to the essentiality code ($\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_3$), with corresponding performance constraints*

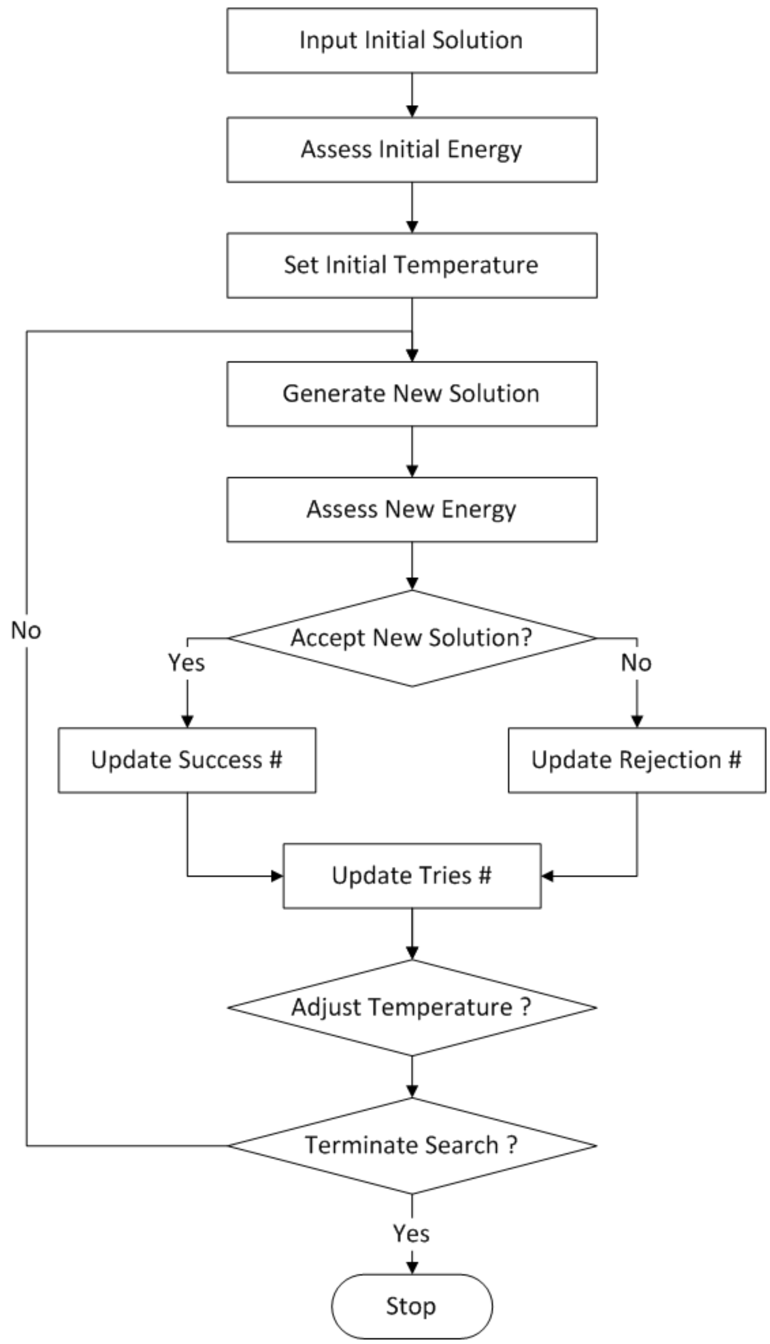


Figure 1: Flow chart of the simulated annealing algorithm

($\hat{\beta}_1,\ \hat{\beta}_2,\ \hat{\beta}_3$), there is a feasible solution for the contracting problem that solves the following constrained integer problem:

$$\begin{array}{lll} min_{s_i} & s_i & \\ & F_i(s_i) \geq \hat{\beta}_j & \forall i \in \mathcal{P}_j, \quad j = 1,2,3 \\ & s_i \in \mathbb{Z} & \end{array} \tag{CIP}$$

CIP is a one-dimensional problem, and hence it can be easily solved by starting from a zero initial solution and progressively incrementing s_i until the constrained is met.

3.1.2 New Solution Generation — [ew] to be modified..

Another key step in the simulated annealing algorithm is generating a new solution using the current one. Naturally, the generation algorithm should ensure that the entire feasible solution space is reachable. We propose a random alteration process governed by two features: the size of the randomly selected set (m), and the step-size(ϕ) which is small but randomly chosen.

Proposition 2 *Suppose m variables are selected each time from the set of $n+1$ decision variables (n spare parts in set $\mathcal{P} = \{p_1, ..., p_n\}$ and a cost-sharing parameter α). Then if $n >> m$ and $0 < |\phi| \leq \hat{\phi}$, the generation efficiency is dependent on the probability of alteration (π) and the repetition coefficient (θ) defined as*

$$\pi \approx \frac{m}{n+1} \qquad , \qquad \theta = 1/\binom{2\hat{\phi}+m+1}{m} \tag{7}$$

The probability of alteration (π) determines whether all combinations of the solution space would be reached. If π is too large it is improbable because too many variables change each time. Reversely, if π is too small, the progress is slow. The repetition coefficient (θ) is inversely proportionate to the binomial coefficient of a multi-set of size $2\hat{\phi}$. If θ is close to one, it is likely that the step-size will be repeated, otherwise it is unlikely.

3.1.3 Acceptance Probabilities

The simulated annealing algorithm overcomes local minima by accepting worse solutions (”uphill” move) in an attempt to arrive at an even better one. Worse solutions are accepted based on a exponentially distributed probability function that depends on the difference in objective values[2] as well as the current temperature. Given extreme variations in energy levels, we propose a scaling function using current energy levels.

Proposition 3 *Let u be a uniformly distributed random variable between 0 and 1, and E_k, E_{k+1} represent the current and new energy levels respectively. If $E_k < E_{k+1}$ and the acceptance condition is $\hat{u}+u < \mathcal{U}(E_k, E_{k+1}, T)$, increasing minimum acceptance threshold ($\hat{u}$) can positively impact the convergence when acceptance probability is defined as*

$$\mathcal{U}(E_k, E_{k+1}, T) = e^{-\frac{E_{k+1}-E_x}{(E_k \cdot T)/n}} \tag{8}$$

Equation (8) uses the fraction of change and temperature as a basis for acceptance. In addition, the change is scaled according to the problem size n.

[2]In simulated annealing, the objective value of a feasible solution is referred to as the energy level of that point.

3.2 Test Setup and Results

We tested our implementation of the simulated annealing algorithm on a commercial aviation use case involving AirAsia's fleet of A320s operating more than 300,000 annual flight hours. We consider an MRO service contract covering 60 landing gear parts (x ESS 1, x ESS 2, and x ESS 3). Unit price ranges from \$2,000 to \$52,000, and MTBUR/QPA ranges from 625 to 345,000.

Table 5 compares three performance measures, namely final energy level, the total number of new solutions generated, and the total number of solutions accepted. Since the search of the simulated annealing is random, we only report the magnitude of the sample averages. Note that in most of the cases, the algorithm converged at the same final solution. A plausible reason is the contracting problem is unimodal under the selected settings and the final solution is indeed a global optimum. Although the final solution is identical except under some extreme parameter settings, the number of solutions generated and the number accepted are different. The number of solutions generated corresponds to the number of random trials taken to reach the final solution, and hence a smaller number reflects faster convergence. The number of accepted solutions provide an additional aspect. Consider these scenarios:

- **Generated: no change; Accepted: decreases** - less efficient because the algorithm takes the same number of trials, but encounters more inferior solutions.
- **Generated: decreases; Accepted: no change** - more efficient because the algorithm requires fewer number of trials to find the same number of superior solutions.
- **Generated: increases; Accepted: decreases** - less efficient because the algorithm require more trials, yet encounters fewer superior solutions.

3.2.1 Impact of Initial Solution

The results show that initial solution derived using s^{CIP} outperforms the brute force approach, which requires more than double the number of generated solutions. For further insight into the impact of the initial solution on the algorithm, we traced the entire history of solutions (including rejected solutions) explored by the algorithm. Customized solution trace graphs (Figures 2a and 2b) were created to visualize the results for initial solutions generated from s^{CIP} and the brute force approach respectively.

The lines represent variable values plotted against the left axis. The first variable is the cost sharing parameter (α)[3] followed by inventory levels (s_i) sorted by the part's unit price. The line colors are indicative of the resulting energy (obtained from the Energy Color Chart). The color in the middle (dark green) represents the initial energy, the extreme left and right colors (dark red and dark blue) represent the lowest and the highest energy respectively.

[3] The cost sharing parameter is scaled by 50 for ease of comparison.

	Final Energy	# Generated (x10^3)	# Accepted (x10^3)
Initial Solution			
($m=1$, $\hat{\phi}=1$, $\hat{u}=0$)			
$s_i = s_i^{CIP}$, $\alpha=0$	2.1768×10^6	35.0	3.0
$s_i = 50$ (Brute force), $\alpha=0$	2.1768×10^6	40.0	5.5
New Solution Generation			
($s_i = s^{CIP}$, $\alpha=0$, $\hat{u}=0$)			
$m=1$, $\hat{\phi}=1$, $\pi=1.64\%$, $\theta=0.50$	2.1768×10^6	35.0	3.0
$m=1$, $\hat{\phi}=2$, $\pi=3.28\%$, $\theta=0.33$	2.1768×10^6	35.0	2.5
$m=2$, $\hat{\phi}=1$, $\pi=1.64\%$, $\theta=0.25$	2.1768×10^6	40.0	2.5
$m=2$, $\hat{\phi}=2$, $\pi=3.28\%$, $\theta=0.10$	2.1768×10^6	40.0	2.5
$m=3$, $\hat{\phi}=1$	2.1769×10^6	55.0	2.0
$m=4$, $\hat{\phi}=1$	2.1782×10^6	55.0	2.0
$m=3$, $\hat{\phi}=2$	2.1768×10^6	55.0	2.0
$m=4$, $\hat{\phi}=2$	2.1778×10^6	55.0	1.5
$m=1$, $\hat{\phi}=3$	2.1771×10^6	55.0	1.5
Acceptance Probabilities			
($s_i = s^{CIP}$, $\alpha=0$, $m=1$, $\hat{\phi}=1$)			
$\hat{u}=0.5$	2.1768×10^6	35.0	2.0
$\hat{u}=1$	2.1768×10^6	20.0	0.1

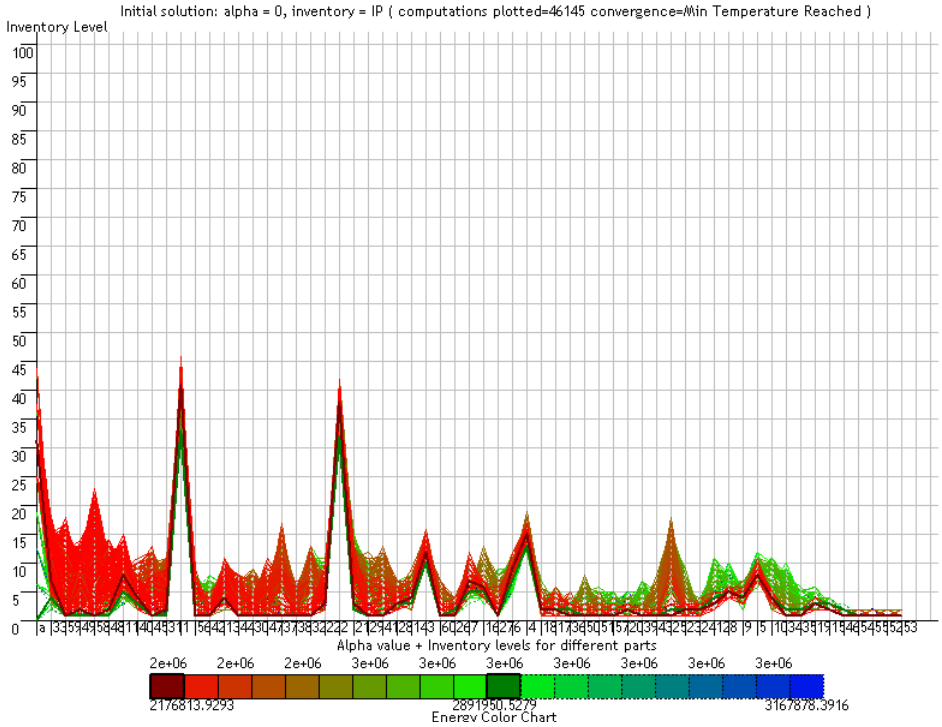
Table 5: Performance of SA under different settings

Comparing Figure 2a with Figure 2b, it is clear that the brute force approach searches a larger space (colored area) to arrive at the identical final solution (dark red line). We highlight two other observations: (i) similar unexplored spaces (white area) below the final solution, and (ii) smaller search area above the final solution (Figure 2a) towards right side of the graph.

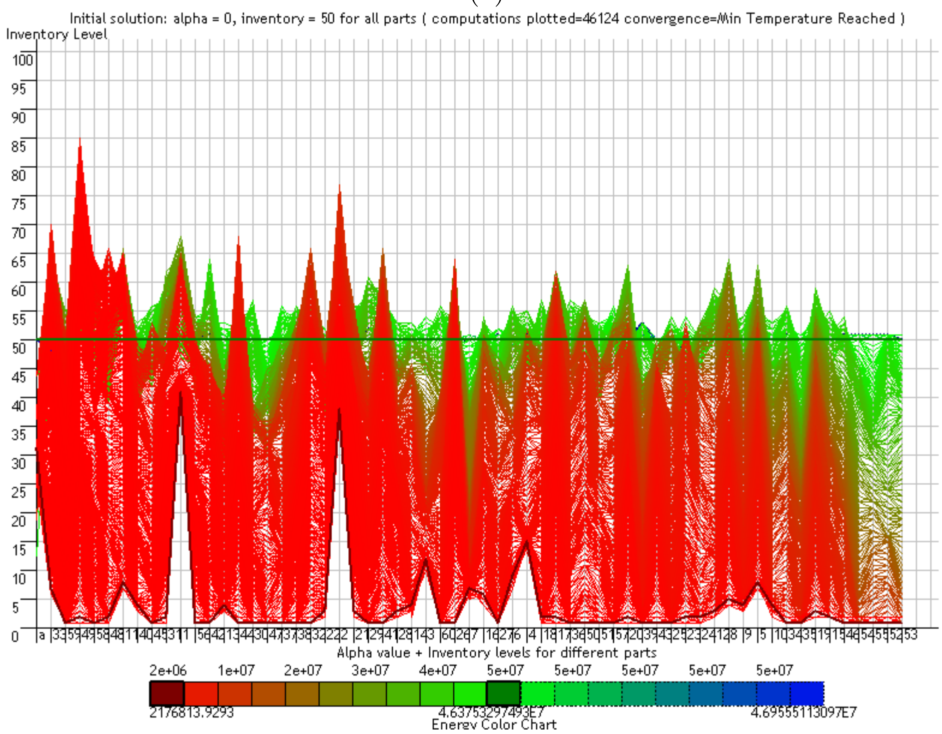
The unexplored spaces below the final solution is observed for parts with a higher annual number of unexpected removals (i.e., parts requiring more inventory). Reducing inventory for such parts beyond a certain level lowers the fill rate significantly. And in order to meet the essentiality performance constraint, the shortfall must be compensated by increasing inventory on another part. However such a move would compromise on the objective as observed by the green lines (indicating an inferior solution). For this reason, we observe that the algorithm identifies a lower bound for inventory levels (a non-optimal state) and subsequently returns to a higher (but superior) inventory level.

The inventory level variables are sorted according to unit price (more expensive parts on the right). A large space remains unexplored above the final solution for expensive parts. This is because increasing inventory on such parts, which although would increase fill rate, would increase MRO service providers costs significantly leading to a higher contract value, and a negative impact on the customers utility (i.e., an inferior solution). For this reason, the upper bound for inventory levels for expensive parts is close to the final solution.

The lower and upper bounds identified by the simulated annealing algorithm controls the exploration around the final solution. This confined but strategic search provide a strong basis



(a)



(b)

Figure 2: Solution trace for initial solution generated from (a) s^{CIP} and (b) brute force approach

for deducing that the final solution is a global optimum.

3.2.2 Impact of Solution Generation Approach

[ew] need more results to comment Figure 3 on the page 16 depict one example of the converge of the algorithm starting from the initial point, $s_i = s_i^{CIP}$ and $\alpha = 0$.

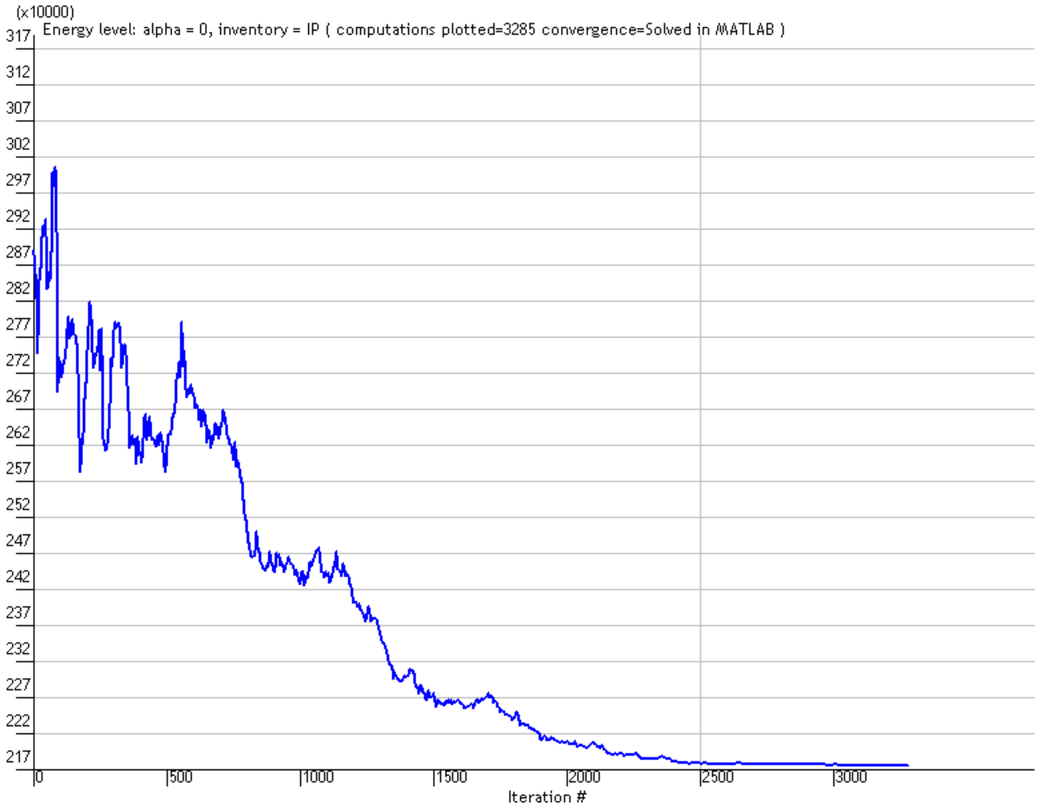


Figure 3: Convergence of SA for m=1, $\hat{\phi} = 1$

It is observed that the algorithm arrives at the minimum energy much earlier than it is terminated. This is because the conservative parameter settings we adopt to guarantee that the solution found by the algorithm is indeed local optimum. Another observation is that we do see the random walk pattern of the energy changes dual the search, which confirms with the unique feature of simulated annealing that it will not be trapped in certain local optima.

3.2.3 Impact of Acceptance Probabilities

Results in Table 5 report fewer generated solutions (i.e., better efficiency) with increasing acceptance threshold $\hat{u}$. A small $\hat{u}$ increases the probability of an inferior solution being accepted. Accepting inferior solutions is only beneficial if the consequence of not accepting results in the algorithm being trapped in a local minimum. However if the threat of local minimum traps is low, rejecting inferior solutions allows the algorithm to move quickly to the global optimum by avoiding inferior spaces. The results reveal that this is indeed the case for the contracting problem.

[ew] need to compare with brute force

3.2.4 Computational Time

The simulated annealing algorithm was implemented in MATLAB 7.8 environment and tested on a Dell desktop with Intel® Core™2 Duo CPU E8400 @ 3.00GHz 2.99GHz and 3.21GB of RAM. On average, it takes less than 7 minutes to solve the problem of sixty parts. Compared with the time cost of solving only ten parts using Lagrangian multipliers and subgradient methods using branch and bound type approach for integrality constraints, which is more than two hours, the proposed simulated annealing is extremely fast while the quality of the solution is also guaranteed. Taking into the account that much of the time is spent during the local search after reaching the optimum, we conclude that the efficiency and accuracy of the proposed simulated annealing algorithm are very promising for our problem.

4 Business Case Study and Analysis

As observed in the previous section, the annual number of unexpected removals (which is dependent on $\tau/\mathcal{Q}$) impact the shape of the optimal solution. In the previous test case, we studied a contract covering parts with $\tau/\mathcal{Q}$ ranging from 625 to 345,000. The question is how would the final solution change if the range is smaller or bigger? To assess the results obtained from the algorithm in face of different contracts comprising parts with different range of unexpected removals, we modified the original test scenario by scaling the mean time between unscheduled removals (τ) for all[4] parts (details in Table 7).

Change in MTBUR	Lowest $\tau/\mathcal{Q}$	Highest $\tau/\mathcal{Q}$	Range in $\tau/\mathcal{Q}$
(20%)	750	414,000	413,250
(10%)	718	396,750	396,032
(15%)	687	379,500	378,813
(5%)	656	362,250	361,594
original	625	345,000	344,375
(-5%)	594	327,750	327,156
(-10%)	562	310,500	309,938
(-15%)	531	293,250	292,719
(-20%)	500	276,000	275,500

Table 7: Performance of SA in face of different business scenarios

Figure 4 shows the percentage change in inventory investment levels for the initial solution (derived using s^{CIP}) and the final solution. We observe that in the most extreme case (-20%),

[4]Improving (reducing) unscheduled removals can be achieved in practice through the adoption of health monitoring or maintenance assistive technologies. Improvements typically affect selected parts rather than all parts. However for this test, we have assumed improvements on all parts to simplify the case study.

the change in inventory investment for the initial solution is substantial (almost 50%), yet the change in inventory investment for the final solution is significantly less (less than 10%). Operationally, this is a significant result for the service provider because it implies lesser impact on operating costs in spite of changes in unexpected removals. This is dampened changes in costs can be seen in (Figure 5). By exploiting on the risk-averseness of the customer, the service provider can potentially get higher profits even though unexpected removals increases (i.e., MTBUR decreases).

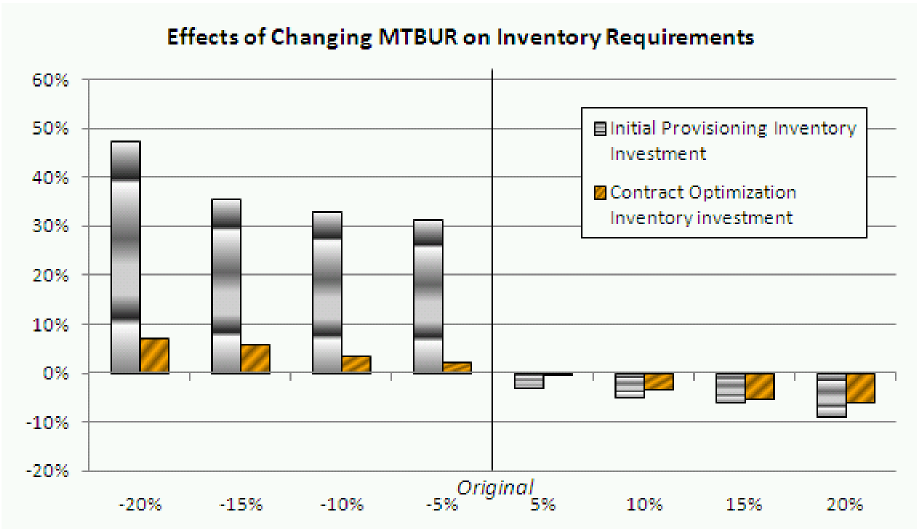


Figure 4: Effects of changing MTBUR on inventory requirements

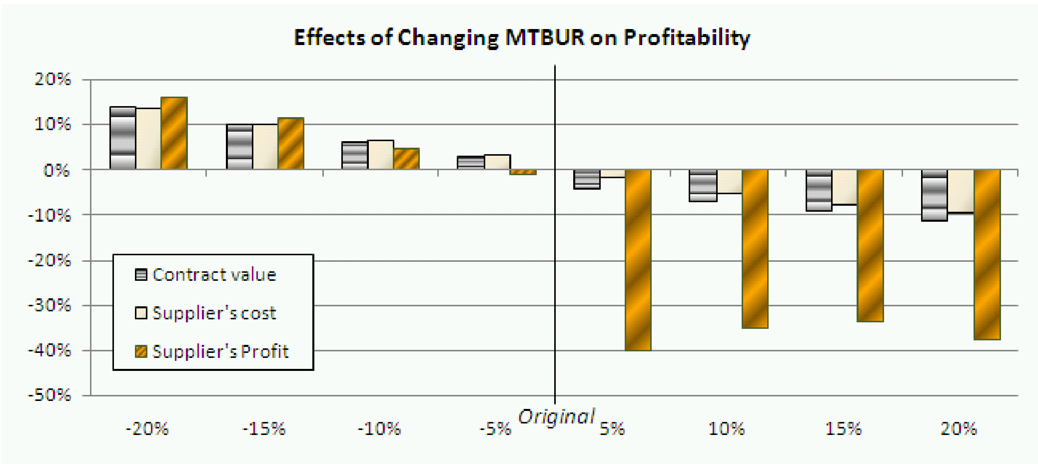


Figure 5: Effects of changing MTBUR on profitability

Figure 6 on shows the percentage change in pricing (in terms of fixed payment, cost sharing, and penalty) as a result of changes in MTBUR. Decreasing MTBUR generally increases the contract value by increasing the fix payment and cost sharing amount. Since the nominal value

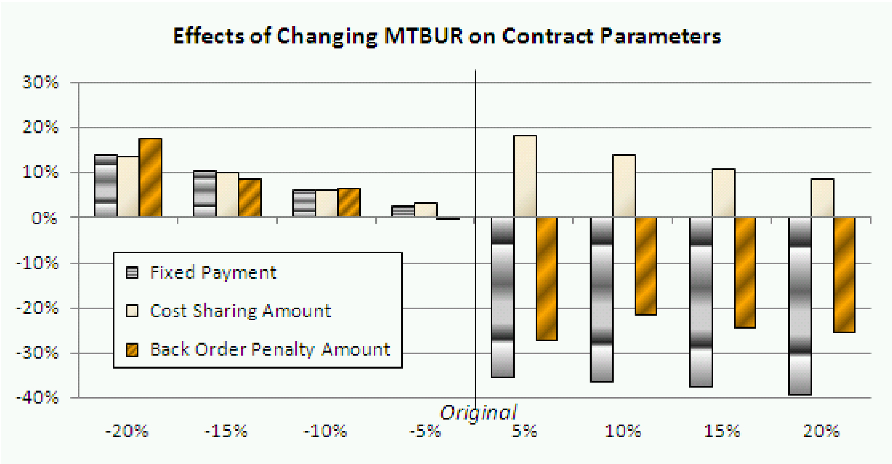


Figure 6: Effects of changing MTBUR on contractual parameters

of back order penalty amount is relatively very small compared to the other two components, the increase in penalty amount is not going to offset the increase in contract value. On the other hand, increasing MTBUR indicates the parts become less frequent to fail, so both of the fix payment and penalty amount decrease and the customer would like to share more cost since the risk is low.

5 Conclusion

Integrality constraints are important for realistic spares parts allocation modelling. However solving the resulting mixed integer programming problem is computationally challenging hence previous works have relaxed the integrality constraints. We present an alternative approach to address the computational challenge by introducing an additional discretized variable to replace the continuous cost-sharing parameter, and then implementing the metaheuristics, simulated annealing algorithm to solve the problem. Given the sensitivity of the algorithm's performance on the implementation of the metaheuristics, we proposed algorithms for (a) deriving initial solutions, (b) generating new solutions, and (c) criterion for accepting worse solutions. The implementation was tested on Air Asia's short-range commercial aviation scenario and we found that ... **([ew] to add once latest results are out)**

The simulated annealing algorithm presented in this work holds potential for further extensions. One aspect is the extension of the model to consider inventory allocations in multiple locations, the use of time-based service levels as a contractual term, and incorporating different part replacement policies (Eg. replace with different part number, repair at different facility). Another aspect is extending the size of the problem to consider more parts, locations, and/or customers. A third aspect is extending the simulated annealing algorithm for better performance by parallelizing, or employing hybrid approaches (Eg. tabu search).

References

S. Axsäter (1990) *Modelling emergency lateral transshipments in inventory systems*, Management Science, Vol. 36, No. 11, pp. 1329–1338.

N. E. Abboud, J. N. Daigle (1997) *A Little's result approach to the service constrained spares provisioning problem for repairable item*, Operations Research, Vol. 45, pp. 577–583.

P. Batchoun, J. A. Ferland, R. Cléroux (2003) *Allotment of aircraft spare parts using genetic alorithms*, Pesquisa Operacional, Vol. 23, No. 1, pp.141–159.

P. Bolton, M. Dewatripont (2005) *Contract Theory*, Cambridge and London, MIT Press.

M. J. Carrillo (1991) *Extensions of Palm's theorem: A review*, Management Science, Vol. 37, No. 6, pp. 739–744.

R. W. Eglese (1990) *Simulated annealing: A tool for operational research*, European Journal of Operational Research, Vol. 46, pp. 271–281.

G. J. Feeney, C. C. Sherbrooke (1966) *The (s-1,s) inventory policy under compound Poisson demand*, Management Science, Vol. 12, No. 6, pp. 391–411.

K. Ganesh, M. Punniyamoorthy (2004) *Optimization of continuous-time production planning using hybrid genetic algorithms-simulated annealing*, The International Journal of Advanced Manufacturing Technology, Vol 26, pp. 148–154.

V. Granville, M. Kvanek, J. P. Rasson (1994) *Simulated annealing: a proof of convergence*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 6, pp. 652–656.

S. Graves (1985). *A multi-echelon inventory model for a repairable item with one for one replenishment*, Management Science, Vol. 31, pp. 1247–1256.

J. Haddock, J. Mittenthal (1992) *Simulation optimization using simulated annealing*, Computers & Industrial Engineering, Vol. 22, No. 4, pp. 387–395.

H. Szu, R. Hartley (1987) *Fast simulated annealing*, Physics Letters A, Vol. 112, No. 3, pp. 157–162.

D. A. Hensher, J. Stanley (2002) *Performance-Based Quality Contracts in Bus Service Provision*, Institute of Transport Studies, University of Sidney.

M. A. Ilgin, S. Tunali *Joint optimization of spare parts inventory and maintenance policies using genetic algorithms*, The International Journal of Advanced Manufacturing Technology, Vol. 34, pp. 594–604.

W. J. Kennedy, J. W. Patterson, L. D. Fredendall (2002) *An overview of recent literature on spare parts inventories*, International Journal of Production Economics, Vol. 76, pp. 201–215.

S. H. Kim, M. A. Cohen, S. Netessine (2007) *Performance contracting in after-sales service supply chains*, Management Science, Vol. 53, No. 12, pp. 1843–1858.

S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi (1983) *Optimization by simulated annealing*, Science, New Series 220, No. 4598, pp. 671–680.

M. L. Lawrence (2005) *erformance-based contracting for human services: Does it work?*, Administration in Social Work, Vol. 29, pp. 63–77.

L. H. Lee, E. P. Chew, S. Teng, Y. Chen (2008) *Multi-objective simulation-based evolutionary algorithm for an aircraft spare parts allocation problem*, European Journal of Operational Research, Vol. 189, pp. 476–491.

M. Lu, C. A. Ma, L. Yuan (2003) *Risk selection and matching in performance-based contracting*, Health Economics, Vol. 12, pp. 339–354.

J. Muckstadt (1973) *A model for multi-item, multi-echelon, multiindenture inventory system*, Management Science, Vol 20, pp. 472–481.

D. Nowicki, U. D. Kumar, H. J. Steudel, D. Verma (2008) *Spares provisioning under performance-based logistics contract: Profit-centric approach*, Journal of the Operational Research Society, Vol. 59, pp. 342–352

I. H. Osman, J. P. Kelly (1996) *Metaheuristics: Theory and Applications*, Kluwer, Boston (1996).

I. H. Osman, G. Laporte (1996) *Metaheuristics: A bibliography*, Annals of Operations Research, Vol. 63, pp. 513–623.

C. Palm (1938) *Analysis of the Erlang traffic formula for busy signal arrangements*, Ericsson Technology, Vol. 4, pp. 39–58.

L. A. Pertersen, L. D. Woodard, T. Urech, C. Daw, S. Sookanan (2006) *Does pay-for-performance improve the quality of health care?*, Annals of Internal Medicine, Vol. 145, pp. 265–272.

R. Sarker, A. Haque (2000) *Optimization of maintenance and spare provisioning policy using simulation*, Applied Mathematical Modelling, Vol. 24, pp. 751–760.

J. A. Shaunty, V. C. Hare, Jr. (1960) *An airline provisioning problem*, Management Technology, Vol. 1, No. 2, pp. 66–84.

C. C. Sherbrooke (1968) *METRIC: a multi-echelon technique for recoverable item control* Operations Research, Vol. 16, pp. 122–141.

C. C. Sherbrooke (2004) *Optimal Inventory Modeling of Systems*, Kluwer Academic Publishers, Boston.

N. Z. Shor (1998) *Nondifferentiable Optimization and Polynomial Problems*, Springer, ISBN-10: 0792349970, ISBN-13:978-0792349976.

K. Zabm, A. Sa (1996) *A stocking policy for spare part provisioning under age based preventive replacement*, European Journal of Operations Research, Vol. 90, pp. 171–181.