A Lagrangian Relaxation Based Approach to Schedule Asset Overhaul and Repair Services

Peter B. Luh, *Fellow, IEEE*, Danqing Yu, Sada Soorapanth, Alexander I. Khibnik, and Ravi Rajamani

***Abstract*—Overhaul and repair services are important segments of the remanufacturing industry, and are characterized by complicated disassembly, repair and assembly process plans, stochastic operations, and the usage of rotable inventory. In view of today's time-based competition, effectively scheduling such services and managing rotable inventory and uncertainties are becoming imperative to achieve on-time deliveries and low overall costs. In this paper, a novel formulation for overhaul and repair services is presented where key characteristics, such as uncertain asset arrivals and operation processing times, and rotable parts are abstracted to model an overhaul center and multiple repair shops in a distributed framework to reflect organizational structures. Interactions between the overhaul center and repair shops are described by sets of coupling constraints across the organizations. Rotable inventory dynamics is formulated in terms of repair operation completion times and asset assembly beginning times to facilitate minimization of inventory holding costs through scheduling. A solution methodology combining Lagrangian relaxation, stochastic dynamic programming, and heuristics is developed to schedule operations in a coordinated manner to minimize total tardiness, earliness, and inventory holding costs. Additionally, penalty terms associated with coupling constraint violations are introduced to the objective function to improve algorithm convergence and schedule quality, and a surrogate optimization framework is used to overcome the inseparability difficulty caused by the penalty terms. Numerical testing results show that the new approach is computationally effective to handle rotable inventory and uncertainties, and provides high quality schedules with low overall costs for stochastic remanufacturing systems.**

***Note to Practitioners*—Overhaul and repair services for jet engines, helicopters, airplanes, are important segments of the remanufacturing industry, and are characterized by complicated disassembly, repair and assembly process plans, stochastic operations, and the usage of rotable inventory. In view of today's highly competitive business climate, effectively scheduling such services and managing rotable inventory and uncertainties are becoming critical to achieve on-time deliveries and low overall costs. In this paper, a novel formulation for overhaul and repair services is presented where key characteristics, such as uncertain asset arrivals and operation processing times, and rotable parts are abstracted to model an overhaul center and multiple repair shops in a distributed framework to reflect organizational structures. A solution methodology based on decomposition and coordination is developed to schedule operations to minimize total tardiness, earliness, and inventory holding costs. Numerical testing results show that the method is computationally efficient for managing rotable inventory and uncertainties, and generates high quality schedules with low overall costs. The value of rotable inventory to reduce tardiness costs and buffer uncertainties is demonstrated, and the robustness of the new method is evaluated by cases with different settings of machine utilization levels and uncertainty levels. The scalability of the method to solve large problems with hundreds of assets is also demonstrated.**

***Index Terms*—Inventory, Lagrangian relaxation, overhaul and repair services, surrogate optimization.**

Manuscript received May 18, 2004; revised November 2, 2004. This paper was recommended for publication by Associate Editor Y. Narahari and Editor N. Viswanadham upon evaluation of the reviewers' comments. This work was supported in part by National Institute of Standards and Technology's Advanced Technology Program under Cooperative Agreement 70NANB9H3023; the National Science Foundation under Grant DMI-0223443, and by a contract from the United Technologies Research Center (UTRC) to the University of Connecticut.

P. B. Luh and D. Yu are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: Peter.Luh@uconn.edu; danqing@engr.uconn.edu).

S. Soorapanth is with the Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333 USA (e-mail: bki2@cdc.gov).

A. I. Khibnik is with the United Technologies Research Center, East Hartford, CT 06108 USA (e-mail: KhibniAI@utrc.utc.com).

R. Rajamani is with Pratt & Whitney, East Hartford, CT 06108 USA (e-mail: ravi.rajamani@pw.utc.com).

Digital Object Identifier 10.1109/TASE.2005.844221

LIST OF SYMBOLS

a_e	Arrival time of asset e.
b_{ej}	Beginning time of asset overhaul operation (e,j).
$\tilde{b}^*_{ej}$	Optimal beginning time of asset overhaul operation (e,j).
b_{eij}	Beginning time of part repair operation (e,i,j).
$\tilde{b}^*_{eij}$	Optimal beginning time of part repair operation (e,i,j).
b^d_e	"Desired" disassembly beginning time of asset e.
c_{ej}	Completion time of asset overhaul operation (e,j).
c_{eij}	Completion time of part repair operation (e,i,j).
d_e	Due date of asset e.
e	Asset index.
(e,j)	j_{th} operation of asset e.
(e,i,j)	j_{th} operation of part (e,i).
E_e	Earliness of asset e, defined as $E_e = \max(0, b^d_e - b_{e1})$.
$\tilde{g}^n$	The SSG at iteration n.
h	Machine type index.
i	Part index.
$I_r(k)$	Inventory level of a rotable part type r at time k.
j	Operation index for both assets and parts.
J	Expected weighted penalties for tardiness, earliness, inventory costs, and violations of coupling constraints.
$J_{(e,i)}$	Number of operations for part (e,i).
$\bar{J}$	Expected cost of J estimated from Monte Carlo runs.
k	Time index, $0 \le k \le K-1$.
K	Time horizon of scheduling.
$\tilde{L}$	Lagrangian function.
$\tilde{L}_e$	Cost for asset subproblem e.
$\tilde{L}^*_e$	Optimal cost for asset subproblem e.

$\tilde{L}_{ei}$	Cost for part subproblem (e, i).
$\tilde{L}_{ei}^*$	Optimal cost for part subproblem (e, i).
M_{kh}	Number of type h machines available at time k.
n	Iteration index.
p_{ej}	Processing time of operation (e, j).
P_{ej}	The set of all possible processing times for operation (e, j).
$\|P_{ej}\|$	Cardinality of P_{ej}.
$\tilde{q}^n$	Surrogate dual value at iteration n.
q^*	Optimal dual value.
s_e	Required wait time of asset e (representing, e. g., transportation time).
s_{eij}	Required "time out" between operation (e, i, j) and $(e, i, j+1)$.
s^n	Step size at iteration n.
T_e	Tardiness of asset e defined as $T_e = \max(0, c_{e2} - d_e)$.
v_{ei}	Random variable describing the service scope of part (e, i).
$V_{ej}(b_{ej})$	Cumulative cost of stage j for beginning time b_{ej}.
w_c	Weight for penalties for coupling constraint violation.
w_e	Weight of tardiness penalty for asset e.
α_{e2k}^r	Assembly beginning indicator defined to be 1 if assembly operation $(e, 2)$ begins at time k and zero otherwise.
β_e	Weight of earliness penalty for asset e.
$\beta_{eiJ_ik}^r$	Repair completion indicator defined to be one if repair operation (e, i, J_i) of a part of rotable type r is completed at time $k-1$, and zero otherwise.
γ_r	Holding cost for each piece of rotable inventory of part type r per unit time.
δ_{ejkh}	0-1 operation variable which is one if operation (e, j) is performed on a machine type h at time k, and zero otherwise.
η_{ei}	Lagrangian multipliers for relaxing cross-organization precedence constraints of part (e, i).
μ_{kr}	Lagrangian multipliers of inventory level of type r at time k.
π_{kh}	Lagrangian multipliers of machine type h at time k.

I. Introduction

OVERHAUL and repair services are important segments of the remanufacturing industry.[1] Such services are traditionally characterized by complicated process plans, stochastic operations, and the usage of rotable inventory [8], [11], e.g.,

- highly variable operations including disassembly that disassembles an asset into parts, part repair operations and assembly that assembles repaired parts back into an asset. The disassembly operation is often time consuming and performed in conjunction with a "discovering" inspection process. The uncertain service scope is often a combined result of customer requirements and the inspection process.
- highly variable processing times for various operations.
- the existence of serial-number-specific-parts (SNS) required to be assembled into the original assets they belonged to, rotable parts, which are refurbished parts satisfying certain qualifications for general use, and the concomitant rotable inventory.

Pressed by customers' demand for fast deliveries [22], it becomes imperative for service providers to achieve short turn-around-times (TATs) and low overall costs by effectively scheduling overhaul and repair operations and managing rotable inventories. The above-mentioned characteristics, however, present major challenges, including how to formulate and manage rotable inventory and uncertainties in a computationally efficient manner for high quality solutions, how to model overhaul and repair operations to reflect organizational structures and facilitate the optimization process. With jet engine overhaul and repair as the background, this paper presents a novel formulation for overhaul and repair services and the corresponding solution methodology to address the above challenges.

After a brief review of the literature in Section II, a model is established in Section III to minimize the overall tardiness and earliness penalties and inventory holding costs. Key characteristics, such as uncertain asset arrivals and operation processing times, uncertain service scopes, serial-number-specific parts and rotable parts, are captured to describe an overhaul center and multiple repair shops in a distributed framework to reflect organizational structures. Interactions between the overhaul center and repair shops are described by sets of cross-organization coupling constraints (i.e., expected operation precedence constraints and inventory level constraints). The rotable inventory dynamics is formulated in terms of part repair completion times and asset assembly beginning times, and this facilitates minimization of inventory holding costs through scheduling.

In view that the above problem is NP hard but with a separable formulation,[2] Lagrangian Relaxation (LR) based methods can be used to decompose the problem into asset or part level subproblems by relaxing coupling constraints within and across organizations using Lagrange multipliers or shadow prices. The subproblems, which are not NP hard, can be easily solved, and coordination of subproblem solutions is performed through iterative price updating to reduce coupling constraint violations. In view of the size and the complexity of the problem, the standard LR-based method may converge slowly, and solutions may contain significant levels of constraint violation. To improve algorithm convergence and schedule quality, additional terms to penalize coupling constraint violation are introduced to the objective function motivated by a method developed for power systems [26]. In view that the penalty terms are not additive, the resulting objective function is inseparable, and this presents challenges to direct application of the Lagrangian Relaxation technique.

To overcome the inseparability difficulty caused by penalty terms, a surrogate optimization framework is developed in Section IV. The key idea is to "pull out" all the terms associated

[1]Remanufacturing is a process in which worn-out products are restored to like-new conditions through a series of disassembly, clean, refurbish, and assembly processes with the infusion of new parts as necessary in a factory environment.

[2]A problem is separable if it has an additive objective function and additive coupling constraints [2].

with one particular asset or part to form a subproblem. By keeping decision variables not belonging to it at their latest values, the subproblem can be efficiently optimized by using stochastic dynamic programming (SDP). Rotable inventory and uncertainties are handled at the subproblem level without excessive computational requirements. Although optimization is approximate, the surrogate subgradient method (SSGM [27]) will allow the algorithm to converge.

The main aim of this paper is to provide a proper formulation and corresponding solution methodology for scheduling overhaul and repair services, an important segment of remanufacturing industry. Numerical testing results presented in Section V show that the method is computationally efficient for managing uncertainties, and generates high quality schedules with low overall costs. The value of rotable inventory to reduce tardiness cost and buffer uncertainties is demonstrated, and the robustness of the new method is evaluated by cases with different settings of machine utilization levels and uncertainty levels. The scalability of the method to solve large problems with hundreds of assets is also demonstrated.

II. Literature Reiview

Remanufacturing and the associated repairable inventory theory have been summarized by Guide and his colleagues in a series of papers, including [6], [8], and [9]. Scheduling using various priority dispatching rules and the Drum-Buffer-Rope method were compared by using simulation and ANOVA analysis [6], [9]. However, it was concluded that a specifically designed framework and the corresponding models were lacking for the production planning and control of remanufacturing systems [9]. In addition, a critical feature of remanufacturing, i.e., the presence of rotable inventory, has not been well addressed, although the value of rotable inventory to buffer variability and to coordinate material flows has been identified [10].

With respect to stochastic scheduling, two categories of approaches have been developed: optimization-based methods and dispatching rules [20]. Optimization-based methods were developed to find optimized scheduling policies for a given problem context based on probabilistic analysis [12], [13], [24] or fuzzy analysis (e.g., [3]). Most results focused on single machine problems or for sequential tasks, with few available for manufacturing systems with complicate process plans or rotable inventories. With respect to dispatching rules, an extensive list of rules has been investigated, ranging from simple ones to complex combination of rules [4], [17], [23]. These rules are computationally efficient, however, the results are often of questionable quality, and there is no good way to systematically improve the results. Another approach is the scenario analysis that obtains "well-hedged" solutions by studying possible scenarios [15], [16]. In view that the number of possible scenarios grows exponentially as the number of uncertain events increases, the application of this method is limited to small problems.

Our recent work combines Lagrangian relaxation with stochastic dynamic programming to provide near-optimal scheduling policies for job shops with uncertain arrival times and pro-

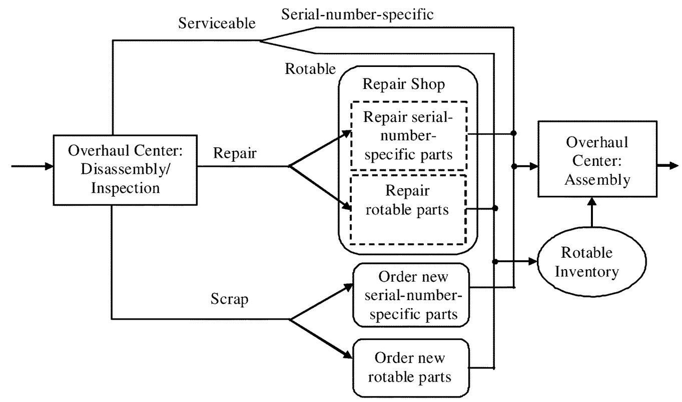

Fig. 1. Schematic of the overhaul and repair services.

cessing time with quantifiable quality [18]. More will be said in Section IV-A. However, the disassembly/assembly processes or rotable inventories have not been considered.

III. Problem Formulation

A. Problem Description

The overhaul center and repair shops could organizationally be divisions within one company or multiple companies; therefore a distributed model is used to reflect this organizational structure as shown in Fig. 1. For simplicity, only one repair shop is included in the figure, although in reality there could be multiple repair shops. Assets arrive at the overhaul center to be dissembled into individual parts, and for each asset, the arrival time is stochastic. Depending on the customer requirement and the outcome of the disassembly/inspection process, a part could be in a serviceable condition waiting to be reassembled, to be repaired in a particular repair shop with a specific routing or processing plan, or to be scrapped. For simplicity of consideration, scrapping a part triggers the ordering of a new part with an uncertain lead-time in this paper. A repair shop consists of one or multiple machine types each with one or multiple identical machines. Part repair times are generally stochastic. A part could be either serial-number-specific (SNS) or rotable. A SNS part after repair should be assembled to the asset it originally belonged to. A rotable part will be sent to the rotable inventory after repair, and to be assembled to an asset requiring the same type of part.

From the above description, it can be seen that overhaul and repair operations are subject to coupling constraints within individual organizations (e.g., machine capacity constraints), and linked through sets of coupling constraints across the organizations (e.g., precedence constraints and inventory level constraints). Effective scheduling and coordination are needed to achieve a shared goal of on-time asset deliveries and low inventory costs. In the following, overhaul center, repair shops, and the cross-organization relationships will be formulated, and finally the objective function will be presented. A list of symbols is provided for easy reference.

B. Formulations of the Overhaul Center

The overhaul center is formulated based on the model of [25] and consists of one or multiple machine types, with the capacity of type h machine at time k given and denoted as M_{kh}. After an asset arrives, it will go through a series of overhaul operations (i.e., disassembly and assembly) with each operation processed by a specific machine type h. The j_{th} overhaul operations of an asset e will be denoted by (e, j). Let $(e, 1)$ represent the aggregate operation of disassembling an asset into parts, and $(e, 2)$ the aggregate operation of assembling parts back into the asset. These operations are subject to the following constraints.

Processing Time Requirements: Operation (e, j) needs to be scheduled on a machine of the required type for a random amount of time p_{ej}

$$c_{ej} = b_{ej} + p_{ej} - 1 \quad \forall\,(e, j) \tag{1}$$

where b_{ej} is the operation beginning time, c_{ej} the operation completion time, and p_{ej} is a random variable with a given discrete distribution.[3]

Expected Machine Capacity Constraints: The number of active operations scheduled on a machine type h should be less than or equal to the capacity of that machine type at any time. Let δ_{ejkh} be an operation indicator defined to be one if operation (e, j) is active at time k on machine type h, and zero otherwise, then machine capacity constraints can be formulated as follows:

$$\sum_{ej} \delta_{ejkh} \leq M_{kh} \quad \forall\, k \;\; \forall\, h \text{ in the overhaul center.} \tag{2}$$

The above constraints are "coupling constraints" within the overhaul center as they couple decision variables belonging to different assets. They are generally difficult to handle, and are particularly complicated for stochastic scheduling in view of the multitude of possible realizations of random events. For this reason, they are approximated by the following expected machine capacity constraints:

$$E\left[\sum_{ej} \delta_{ejkh}\right] \leq M_{kh} \quad \forall\, k, \;\; \forall\, h \text{ in the overhaul center.} \tag{3}$$

These constraints are to be satisfied in the expected sense (as opposed to the sample path sense) in the core of the optimization algorithm, and to be strictly satisfied in the schedule implementation phase.

Arrival Time Constraints: Disassembling an asset cannot be started until the asset has arrived at the overhaul center, satisfying the arrival time constraints

$$a_e + s_e + 1 \leq b_{e1} \quad \forall\, e \tag{4}$$

where a_e is the arrival date of asset e and also is a discrete random variable with a given distribution, s_e is a required wait time, and b_{e1} is the beginning time of disassembly.

C. Formulations of Part Repair Shops

After the disassembly/inspection process, the distribution of a discrete random variable v_{ei} is obtained to describe whether part (e, i) is in serviceable condition, to be scrapped, or to be repaired with a particular routing, with each option represented by a discrete value with an assigned probability. The uncertain service scopes can be effectively described by using this random variables v_{ei}.

A part to be repaired will be assigned to a particular repair shop according to its specified processing plan. Similar to the overhaul center, a repair shop also contains multiple machine types each with a number of identical machines. Part (e, i) has to go through a series of $J_{(e,i)}$ operations with each operation processed by a specific machine type h. For both serial-number-specific parts and rotable parts, their repair operations are subject to the following constraints.

Processing Time Requirements: Each repair operation needs to satisfy the following processing time requirement:

$$c_{eij} = b_{eij} + p_{eij} - 1 \quad \forall\,(e, i, j). \tag{5}$$

If a part is scrapped, a replacement part will be ordered. In this case, (5) can be used with p_{eij} being the lead-time to order a new part. Similarly, if a part is in a serviceable condition, (5) can be used with p_{eij} being the required time-out (representing, e.g., transportation time).

Operation Precedence Constraints: Repair operation $(e, i, j+1)$ cannot be started until its preceding operation (e, i, j) has been completed, possibly plus a required time-out s_{eij} (representing, e.g., transportation time). This translates to the following operation precedence constraints:

$$c_{eij} + s_{eij} \leq b_{ei(j+1)} \quad \forall\,(e, i, j \neq J_i). \tag{6}$$

Expected Machine Capacity Constraints: Similar to (3), the expected machine capacity constraints for repair shops are given as

$$E\left[\sum_{eij} \delta_{eijkh}\right] \leq M_{kh} \quad \forall\, k, \;\; \forall\, h \text{ in the repair shop.} \tag{7}$$

For simplicity of the notation, it is assumed that the overhaul center and repair shops generally do not share same machine types, and the same symbol M_{kh} is used to denote the machine capacity. In addition, the uncertain service scope described by v_{ei} can be effectively handled by the above constraints.

D. Formulations of Cross-Organization Relationships

Overhaul and repair operations are linked through sets of operation precedence constraints and rotable inventory constraints as follows.

Expected Operation Precedence Constraints: The first repair operation of a part cannot be started until the asset it belongs to has been disassembled, plus a required time-out s_{e1} (representing, e.g., transportation time), i.e.,

$$c_{e1} + s_{e1} \leq b_{ei1} \quad \forall\,(e, i)$$
$$\text{or equivalently,} \quad b_{e1} + p_{e1} + s_{e1} \leq b_{ei1} \quad \forall\,(e, i). \tag{8}$$

Similarly, assembling asset e cannot be started until all its serial-number-specific parts have been repaired or their replacements have arrived, plus a required time-out, i.e.,

$$b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} \leq b_{e2} \quad \forall\, \mathrm{SNS}(e, i). \tag{9}$$

[3]For simplicity, it is assumed that the distributions of all random variables are independent of each other.

In addition, all required rotable parts should be available at the inventory pool, and this will be explained in detail later. In view that operation processing times are stochastic, the above coupling precedence constrains (8) and (9) are approximated by their expected forms

$$E[b_{e1} + p_{e1} + s_{e1} - b_{ei1}] \leq 0 \quad \forall\,(e,i) \tag{10}$$

$$E\left[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}\right] \leq 0 \quad \forall\,\mathrm{SNS}(e,i). \tag{11}$$

They are to be satisfied in the expected sense in the core of the optimization algorithm, and will be strictly satisfied in the schedule implementation phase.

Rotable Inventory: The inventory level of a particular rotable part type r at time k, i.e., $I_r(k)$, is affected by overhaul or repair operations. It is increased by one when the last repair operation for a part of rotable type r is completed or when a serviceable rotable part or a replacement of a scrapped part has arrived possibly plus a required time-out (assuming zero here for simplicity). It is decreased by one when such a part is needed for assembly. The inventory level is thus formulated in terms of repair completion times and assembly beginning times. To do this, two indicator variables are introduced: the repair completion indicator $\beta^r_{eiJ_ik}$ is defined to be one if repair operation (e,i,J_i) for a part of rotable type r is completed at time $k-1$, or when a serviceable part or a replacement of a scrapped part has arrived at time $k-1$ (the latter will not be explicitly represented here for the sake of simplicity), and zero otherwise. Similarly, the assembly beginning indicator α^r_{e2k} is defined to be one if assembly operation $(e,2)$ for a part of type r begins at time k and zero otherwise, as illustrated in Fig. 2.

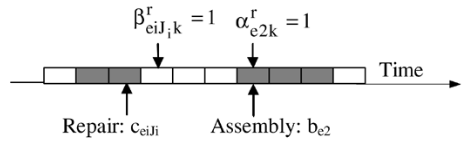

Fig. 2. Repair completion and assembly beginning indicators.

The inventory dynamics for a rotable part i of type r is then formulated as the following flow balance equation:

$$I_r(k+1) = I_r(k) + \sum_{ei} \beta^r_{eiJ_ik} - \sum_e \alpha^r_{e2k} \tag{12}$$

with a given initial inventory level $I_r(0) = I_{r0}$. Inventory level at k can then be recursively obtained in terms of repair completion times and assembly beginning times to facilitate the scheduling process

$$\begin{aligned} I_r(k+1) &= I_r(k) + \sum_{ei} \beta^r_{eiJ_ik} - \sum_e \alpha^r_{e2k} \\ &= I_{r0} + \sum_{ei}\left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa}\right] - \sum_e \left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa}\right]. \end{aligned} \tag{13}$$

Inventory Level Constraints: Rotable inventory is subject to the following nonnegativity constraints:

$$I_r(k+1) \geq 0 \quad \forall\, r \text{ and } k = 0, \ldots, K-1. \tag{14}$$

In view that an assembly cannot be started unless a rotable part of each required type is available in the inventory pool, and (14) imposes additional constraints on assembly operations. Similar to machine capacity constraints, the above are approximated by their expected forms

$$E\left[I_r(k+1)\right] \geq 0 \quad \forall\, r \text{ and } k = 0, \ldots, K-1 \tag{15}$$

or

$$\sum_e E\left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa}\right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa}\right] - I_{r0} \leq 0 \qquad \forall\, r \text{ and } k = 0, \ldots, K-1. \tag{16}$$

Note that inventory level constraints are now formulated in terms of repair completion times and assembly beginning times, and this facilitates the optimization process and represents a unique feature of this paper. In addition, the uncertain service scope described by v_{ei} can be effectively handled by (16).

E. The Objective Function

It is assumed that the overhaul center and repair shops share the same goal to achieve on-time asset delivery (with given required TATs) and low inventory costs. This goal is translated to minimizing the weighted sum of expected tardiness and earliness penalties plus expected inventory-holding costs. Assuming that each piece of rotable inventory of part type r incurs a holding cost of γ_r per unit time, the objective function is then modeled as shown in (17) at the bottom of the page. In the above equation, the summation over r is over all rotable inventory types.

In view of the existence of a large number of coupling constraints with various natures, the standard Lagrangian relaxation technique may not be effective to reduce the level of constraint violation and obtain a satisfactory solution within a short period of time. To improve algorithm convergence and schedule quality, additional terms to penalize coupling constraint violations are introduced motivated by a method developed for power systems [26] as shown in (18) at the bottom of the next page, where w_c is the weight for penalties for coupling constraint violation. The overall problem is to minimize J (18) subject to constraints within organizations, including operation processing requirements (1) and (5), arrival time constraints (4), operation precedence constraints (6), expected machine capacity constraints (3) and (7); and cross-organization constraints, including expected precedence constraints (10) and (11), and expected inventory level constraints (16) with given machine capacities $\{M_{kh}\}$ and initial inventory levels $\{I_{r0}\}$. The decision variables are operation beginning times

$$\begin{aligned} J' &\equiv E\left[\sum_e \left(w_e T_e^2 + \beta_e E_e\right) + \sum_{kr} \gamma_r I_r(k+1)\right] \\ &= E\left[\sum_e \left(w_e T_e^2 + \beta_e E_e\right) + \sum_{kr} \gamma_r \left(I_{r0} + \sum_{ei}\left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa}\right] - \sum_e \left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa}\right]\right)\right]. \end{aligned} \tag{17}$$

for disassembly, repair, and assembly operations. In view that the new penalty terms involve the max operator and are not additive, the resulting formulation is inseparable.

IV. Solution Methodology

A. Overview

Lagrangian relaxation (LR) is a powerful approach for constrained optimization of NP-hard problems with separable formulations [2], [18], [25]. To be consistent with the organizational structure, ideally a two-step relaxation is carried by the traditional LR. First, cross-organization constraints (10), (11), and (16) are relaxed by using sets of Lagrangian multipliers or "shadow prices," and the overall problem is decomposed into individual subproblems, one for each organization. The coupling constraints within individual organizations (3) and (7) are then relaxed, and organizational subproblems are further decomposed into asset or part-level subproblems. Subproblems can be solved by individual organizations in a distributed fashion based on their internal situations and inter-organizational prices. Coordination of subproblem solutions are achieved through the iterative updating of multipliers based on the "pricing" concept of market economy, and either by a coordinator in a synchronous fashion [5], or by individual organizations in an distributed and asynchronous fashion without a coordinator (e.g., [19]).

In this paper, for simplicity of derivation but without loss of mathematical correctness, coupling constraints across and within individual organizations are relaxed at the same time. The relaxed problem is inseparable, and the traditional LR-based approach cannot be directly applied. To overcome this difficulty, a surrogate optimization framework is established where all terms associated with one asset or part are "pulled out" from the Lagrangian to form an subproblem with decision variables belonging to other assets or parts kept at their latest available values. These asset or part subproblems are solved by individual organizations. For simplicity of presentation, it is assumed that there is a coordinator to coordinate individual subproblems and update all the multipliers in a synchronous fashion. Interested readers can, respectively, refer to [5] and [19] for more information about synchronous and asynchronous distributed coordination by individual organizations. The schematic of the approach is given in Fig. 3.

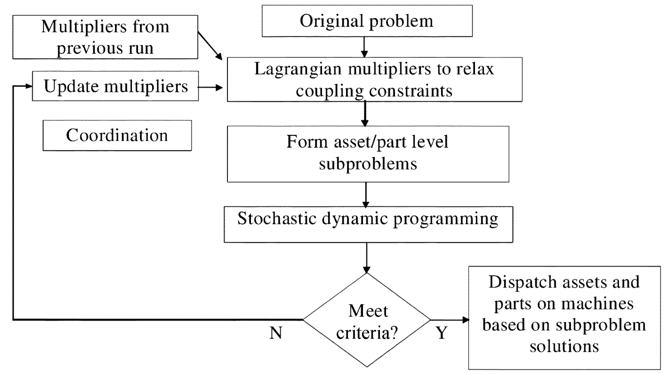

Fig. 3. Schematic of the new LR-based approach.

B. Forming the Relaxed Problem

For simplicity of derivation and presentation, coupling constraints, including cross-organization precedence constraints (10) and (11), inventory level constraints (16), and intra-organizational machine capacity constraints (3) and (7) are relaxed at the same time by using sets of Lagrangian multipliers $\{\eta\}$, $\{\mu\}$ and $\{\pi\}$. For simplicity of notation, the same symbol π_{kh} is used to relax machine capacity constraints (3) and (7). The relaxed problem is to minimize $\tilde{L}$ defined below

$$\begin{aligned}\tilde{L} \equiv J &+ \sum_{kh} \pi_{kh} \left[\sum_{ej} E(\delta_{ejkh}) - M_{kh} \right] \\ &+ \sum_{kh} \pi_{kh} \left[\sum_{eij} E(\delta_{eijkh}) - M_{kh} \right] \\ &+ \sum_{ei} \eta_{ei} E[b_{e1} + p_{e1} + s_{e1} - b_{ei1}] \\ &+ \sum_{ei} \eta_{ei} E[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}] \\ &+ \sum_{kr} \mu_{kr} \left[\sum_{e} E\left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa} \right] \right. \\ &\qquad \left. - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa} \right] - I_{r0} \right]. \end{aligned} \tag{19}$$

The minimization is subject to operation processing requirements (1) and (5), arrival time constraints (4), and operation

$$\begin{aligned} J = E &\left[\sum_e \left(w_e T_e^2 + \beta_e E_e \right) + \sum_{kr} \gamma_r \left(I_{r0} + \sum_{ei} \left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa} \right] - \sum_e \left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa} \right] \right) \right] \\ &+ w_c \sum_{kh} \max \left\{ 0, \sum_{ej} E(\delta_{ejkh}) - M_{kh} \right\} + w_c \sum_{kh} \max \left\{ 0, \sum_{eij} E(\delta_{eijkh}) - M_{kh} \right\} \\ &+ w_c \sum_{ei} \max \{ 0, E[b_{e1} + p_{e1} + s_{e1} - b_{ei1}] \} + w_c \sum_{ei} \max \{ 0, E[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}] \} \\ &+ w_c \sum_{kr} \max \left\{ 0, \sum_e E\left[\sum_{\kappa=0}^{k} \alpha^r_{e2\kappa} \right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta^r_{eiJ_i\kappa} \right] - I_{r0} \right\}, \end{aligned} \tag{18}$$

precedence constraints (6), with given machine capacities $\{M_{kh}\}$, initial inventory levels $\{I_{r0}\}$, and all multipliers. The decision variables are beginning times for disassembly, repair, and assembly operations. This relaxed problem is inseparable in view that the Lagrangian $\tilde{L}$ contains inseparable penalty functions.

C. Formulations and Resolution of Asset Subproblems

To overcome the inseparability difficulty, all terms associated with a particular asset e are "pulled out" from $\tilde{L}$ to form an asset subproblem while decision variables not belonging to e are kept at their latest available values. The subproblem for asset e is thus formulated as shown in (20) at the bottom of the page, subject to operation processing requirements (1) and arrival time constraints (4). The decision variables are beginning time b_{e1} for disassembly and b_{e2} for assembly, and are to be optimized by the overhaul center. It should be noted that disassembly and assembly are related to repair operations through the relaxed expected operation precedence constraints (10) and (11), and to rotable inventory through the relaxed expected inventory level constraints (16). The multipliers $\{\eta\}$ and $\{\mu\}$ therefore play a key role to coordinate these operations.

The subproblem is solved by using backward stochastic dynamic programming (SDP, [18]), where a stage corresponds to an operation, and a state corresponds to a possible operation beginning time. The SDP algorithm starts with the last stage (i.e., assembly) having the following expected terminal cost, as shown in (21) at the bottom of the page.

The expected cumulative cost at the first stage (i.e., assembly) as the algorithm moves backward is

$$\begin{aligned} V_{e1}(b_{e1}) = & E\left[\beta_e E_e + \sum_{k=b_{e1}}^{c_{e1}} \pi_{kh} + \left(\sum_i \eta_{ei}\right) b_{e1}\right. \\ & + w_c \sum_i \max\{0, E[b_{e1} + p_{e1} + s_{e1} - b_{ei1}]\} \\ & + w_c \sum_{kh} \max\left\{0, \sum_e E(\delta_{e1kh}) - M_{kh}\right\} \\ & \left. + \min_{\{b_{e2}\}} V_{e2}(b_{e2})\right]. \end{aligned} \tag{22}$$

The resulting optimal cost $\tilde{L}_e^*$ is obtained as the minimal expected cumulative cost at the first stage

$$\tilde{L}_e^* = \min_{\{b_{e1}\}} E[V_{e1}(b_{e1})] \tag{23}$$

subject to arrival time constraints (4). The computational complexity for solving subproblem e is $O(K \sum_j |P_{ej}|)$, where K is the time horizon, P_{ej} is the set of all possible processing times for operation (e, j), and $|P_{ej}|$ is its cardinality. The optimal beginning times can be determined by tracing forward the optimal SDP path based on the realization of random arrival and processing times. The result of SDP is a policy describing what to do under which circumstance, and will be used to construct feasible schedules based on the realization of random events. The above optimization is approximate in view that decision variables not belonging to the subproblem are kept at their latest available values.

$$\begin{aligned} \min \tilde{L}_e, \text{ with } \tilde{L}_e \equiv & E\left[w_e T_e^2 + \beta_e E_e + \sum_j \left(\sum_{k=b_{ej}}^{c_{ej}} \pi_{kh}\right)\right. \\ & \left. + \left[\left(\sum_i \eta_{ei}\right) b_{e1} - \left(\sum_i \eta_{ei}\right) b_{e2}\right] + \sum_{kr} (\mu_{kr} - \gamma_r) \left(\sum_{\kappa=0}^{k} \alpha_{e2\kappa}^r\right)\right] \\ & + w_c \sum_{kh} \max\left\{0, \sum_{ej} E(\delta_{ejkh}) - M_{kh}\right\} + w_c \sum_i \max\{0, E[b_{e1} + p_{e1} + s_{e1} - b_{ei1}]\} \\ & + w_c \sum_i \max\{0, E[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}]\} \\ & + w_c \sum_{kr} \max\left\{0, \sum_e E\left[\sum_{\kappa=0}^{k} \alpha_{e2\kappa}^r\right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta_{eiJ_ik}^r\right] - I_{r0}\right\} \end{aligned} \tag{20}$$

$$\begin{aligned} V_{e2}(b_{e2}) = & E\left[w_e T_e^2 + \sum_{k=b_{e2}}^{c_{e2}} \pi_{kh} - \left(\sum_i \eta_{ei}\right) b_{e2} + \sum_{kr} (\mu_{kr} - \gamma_r) \left(\sum_{\kappa=0}^{k} \alpha_{e2\kappa}^r\right)\right] \\ & + w_c \sum_{kh} \max\left\{0, \sum_e E(\delta_{e2kh}) - M_{kh}\right\} + w_c \sum_i \max\{0, E[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}]\} \\ & + w_c \sum_{kr} \max\left\{0, \sum_e E\left[\sum_{\kappa=0}^{k} \alpha_{e2\kappa}^r\right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta_{eiJ_i\kappa}^r\right] - I_{r0}\right\} \end{aligned} \tag{21}$$

D. Formulations and Resolution of Part Subproblems

By pulling out terms related to part (e, i) from the relaxed function $\tilde{L}$, the subproblem for (e, i) is formulated as shown in (24) at the bottom of the page, subject to operation processing requirements (5) and operation precedence constraints (6). The decision variables are beginning times for repair operations, and are to be optimized by the corresponding repair shop. Similar to an asset subproblem, this part subproblem is optimized by using SDP, and the optimal cost $\tilde{L}_{ei}^*$ is obtained as the minimal expected cumulative cost at the first repair stage

$$\tilde{L}_{ei}^*(\pi, \eta, \mu) = \min_{\{b_{ei1}\}} E\left[V_{ei1}(b_{ei1})\right]. \tag{25}$$

The computational complexity for solving part subproblem (e, i) is $O(K \sum_j |P_{eij}|)$, where P_{eij} is the set of all possible processing times for (e, i, j), and $|P_{eij}|$ is its cardinality.

E. Solving the High Level Dual Problem

Assuming the existence of a coordinator, the high level dual problem is to find an optimal set of multipliers to maximize the following dual function, i.e.,

$$\max_{\pi,\eta,\mu} \tilde{q}(\pi, \eta, \mu), \text{ with } \tilde{q}(\pi, \eta, \mu) \equiv \tilde{L}^* \tag{26}$$

where $\tilde{L}^*$ is $\tilde{L}$ in (19) evaluated at optimal beginning times obtained from subproblem solutions, and the optimization is subject to nonnegativity of multipliers. To solve (26), the surrogate subgradient (SSG, [27]) method is used. The key idea is that a proper search direction can be obtained under certain conditions without solving optimally all the subproblems. In fact, an approximate optimization of one or a few subproblems is needed to get a proper SSG direction. This surrogate optimization overcomes the inseparability difficulty while allowing more frequent multiplier updating as compared to the standard LR method.

It has been shown in [27] that if at iteration n, the following condition is satisfied:

$$\tilde{L}(\pi^n, \eta^n, \mu^n, b_{ei}^n, b_{eij}^n) < \tilde{L}(\pi^n, \eta^n, \mu^n, b_{ei}^{n-1}, b_{eij}^{n-1}) \tag{27}$$

then surrogate dual value is always less than the optimal dual value q^*, i.e.,

$$\tilde{q}^n < q^*. \tag{28}$$

Furthermore, the surrogate direction forms an acute angle with the direction toward optimal multipliers, and therefore is a proper direction. This property allows the new stochastic LR (SLR) method to converge to an optimal dual solution. The components of the SSG $\tilde{g}^n$ at iteration n is then given by

$$\tilde{g}_{kh}^{\pi n} = \sum_{ej} E\left(\delta_{ejkh}\right) - M_{kh} \quad \forall\, h \text{ in the overhaul center} \tag{29}$$

$$\tilde{g}_{kh}^{\pi n} = \sum_{eij} E\left(\delta_{eijkh}\right) - M_{kh}, \quad \forall\, h \text{ in the repair shop} \tag{30}$$

$$\tilde{g}_{ei}^{\eta n} = E\left[\tilde{b}_{e1}^* + p_{e1} + s_{e1} - \tilde{b}_{ei1}^*\right] \tag{31}$$

$$\tilde{g}_{ei}^{\eta n} = E\left[\tilde{b}_{eiJ_i}^* + p_{eiJ_i} + s_{eiJ_i} - \tilde{b}_{e2}^*\right] \tag{32}$$

$$\tilde{g}_{kr}^{\mu n} = \sum_{e} E\left[\sum_{\kappa=0}^{k} \alpha_{e2(\kappa+1)}^r\right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta_{eiJ_i\kappa}^r\right] - I_{r0} \tag{33}$$

where $\tilde{b}^*$ are optimal beginning times from subproblem solutions. It should be noted that in (29) and (30), the same symbol $\tilde{g}_{kh}^{\pi n}$ is used to represent the subgradients for multipliers π_{kh} in the overhaul center and repairs shops.

The multipliers are updated in the SSG direction

$$\pi_{kh}^{n+1} = \pi_{kh}^n + s^n \tilde{g}_{kh}^{\pi n} \; \forall\, h \text{ in overhaul centers or repair shops} \tag{34}$$

$$\eta_{ei}^{n+1} = \eta_{ei}^n + s^n \tilde{g}_{ei}^{\eta n} \tag{35}$$

$$\mu_{kr}^{n+1} = \mu_{kr}^n + s^n \tilde{g}_{kr}^{\mu n}. \tag{36}$$

The step size s^n at iteration n is set to be

$$s^n = \frac{\gamma^n (q^* - \tilde{q}^n)}{||\tilde{g}^n||^2}, \qquad 0 < \gamma^n < 1 \tag{37}$$

where q^* is the optimal dual value and $\tilde{q}^n$ the surrogate dual value obtained at iteration n. Since q^* is not known, the "Variable Target Value Method" (VTVM) for the subgradient optimization [14] is extended to the SSG method context to systematically provides an estimate of q^* to be used in (37).

To initialize SSGM, the objective function J' of (17) is used for the first iteration. In view that this problem is separable, the traditional LR approach is used and all subproblems are minimized to satisfy $\tilde{q}^0 < q^*$.

$$\begin{aligned} \min \tilde{L}_{ei}, \text{ with } \tilde{L}_{ei} \equiv & E\left[\sum_j \left(\sum_{k=b_{eij}}^{c_{eij}} \pi_{kh}\right) - \eta_{ei} b_{ei1} + \eta_{ei} b_{eiJ_i} + \sum_k (\gamma_r - \mu_{kr}) \left(\sum_{\kappa=0}^{k} \beta_{eiJ_i\kappa}^r\right)\right] \\ & + w_c \sum_{kh} \max\left\{0, \sum_{eij} E(\delta_{eijkh}) - M_{kh}\right\} + w_c \max\left\{0, E\left[b_{e1} + p_{e1} + s_{e1} - b_{ei1}\right]\right\} \\ & + w_c \max\left\{0, E\left[b_{eiJ_i} + p_{eiJ_i} + s_{eiJ_i} - b_{e2}\right]\right\} \\ & + w_c \sum_k \max\left\{0, \sum_e E\left[\sum_{\kappa=0}^{k} \alpha_{e2\kappa}^r\right] - \sum_{ei} E\left[\sum_{\kappa=0}^{k} \beta_{eiJ_i\kappa}^r\right] - I_{r0}\right\} \end{aligned} \tag{24}$$

F. On-Line Dispatching by Using Heuristics

Subproblem solutions, when put together, are generally infeasible since expected coupling constraints (3), (7), (10), (11), and (16) have been relaxed. Furthermore, in view that these constraints are in fact approximations of actual constraints (2), (8), (9), and (14), a schedule satisfying all the expected constraints is generally not implementable. A list scheduling heuristics is, therefore, developed based on the heuristics of [18], where a list of "assignable" operations is created at time 0 and updated at each subsequent time unit based on SDP solutions and the realization of random events. The procedure is illustrated as follows.

1) For all the operations ready to be dispatched at time 0, an operation list is created in the ascending order of their assigned optimal beginning times based on the SDP policy.
2) Operations are scheduled on the required machine types according to the operation list as machines become available
3) When multiple operations competed for a less number of machines at a particular time unit, incremental tardiness costs for delaying by one time unit are calculated. Operations are then assigned to machines in the descending order of their incremental tardiness costs until all available machines are used. The remaining operations are delayed by one time unit.
4) The process terminates if all operations are assigned to the required machine types. Otherwise, go to the next time unit.
5) If operation j is completed and it has a succeeding operation $j+1$, the beginning time for $j+1$ is determined by the SDP policy based on the realization of the processing time of j. The operation list is then updated by inserting succeeding operations in the ascending order of their beginning times. Then go to 2).

The overall complexity of the algorithm including Lagrangian relaxation, stochastic dynamic programming, and heuristics, is mainly determined by the total number of subproblems solved during the process, and increases as the size of the problem increases.

G. Performance Evaluation

To evaluate the performance of our method, Monte Carlo simulation has been performed based on subproblem solutions. Random variables were realized based on their discrete distributions. After N runs, the sample mean of independent realizations J_n, with $n = 1, \ldots, N$ provides an estimate of the expected cost J

$$J \approx \bar{J} \equiv \left(\frac{1}{N}\right) \sum_{n=1}^{N} J_n. \tag{38}$$

The associated sample variance is given by

$$Var_{\bar{J}} = \frac{1}{N^2} \sum_{i=1}^{N} (J_n - \bar{J})^2. \tag{39}$$

To compare the performance of different algorithms for a particular case, the same number of Monte Carlo runs was performed using the same set of random variables for each method. Then the sample mean and sample variance of the cost differences were calculated, and optimality comparison technique is used based on hypothesis testing to derive the significance of comparison [1].

In view that the surrogate dual cost $\tilde{q}$ is a lower bound to the optimal dual cost q^* [27] and q^* is a lower bound to the expected feasible cost J [18], $\tilde{q}$ is a lower bound to the expected feasible cost. The relative duality gap $(J - \tilde{q})/\tilde{q}$ or its approximation $(\overline{J} - \tilde{q})/\tilde{q}$ is used as a measure of schedule optimality.

V. NUMERICAL RESULTS

The method and the simulation environment were implemented by using Matlab on a Pentium IV 2 GHz PC with 512 SDRAM. Numerical testing has been performed based on a test bed developed by the United Technologies Research Center in conjunction with Pratt & Whitney and i2 Technologies under the project "Condition-Based Maintenance" supported by the National Institute of Standards and Technology's Advanced Technology Program. In the test bed, each asset has two rotable parts of different types, and each part requires a series of two repair operations. The overhaul center has two machine types, one for disassembling assets into parts, and the other for assembling parts into assets. There is a single repair shop that contains two machine types, one for the first repair operations, and the other for the second repair operations. The asset disassembly and assembly times are deterministic. The asset arrival times and part repair times are uncertain with their distributions randomly selected from sets of given three-value distributions with specific variances. These distributions are symmetric, taking values at $k-2$, k and $k+2$ with probabilities $0.5(1-p)$, p, and $0.5(1-p)$, respectively. By adjusting k and p, the mean k and variance $4(1-p)$ can be separately controlled.

In the following, three examples are examined. The first is a small problem with 12 assets under different settings of initial inventory. The performance of the new method is compared with that of the "mean method," where all random variables are replaced by their means, and the converted problems are solved by using the deterministic LR approach with additional penalty terms. The second example is a medium-sized problem with 100 assets under different levels of machine utilization, and the performance of the new method is compared with that of the traditional SLR method without additional penalty terms and the FIFO rule. The third example considers problems with 100, 200, and 300 assets to demonstrate the scalability of the new method.

For all the examples, assets arrive randomly at the overhaul center in-between day 5 and day 35 to go through all disassembly, repair, and assembly operations. The due date of an asset is its expected arrival time plus a required turn-around-time, which equals the expected total processing time plus a slack time uniformly distributed within the interval $[5, 16]$. The tardiness and earliness weights of each asset are 1 and 0.05, respectively, and the weight for additional penalty terms w_c is 1. The inventory holding cost is 0.1 per day for both rotable part types. For LR-based methods, the multipliers are initialized at

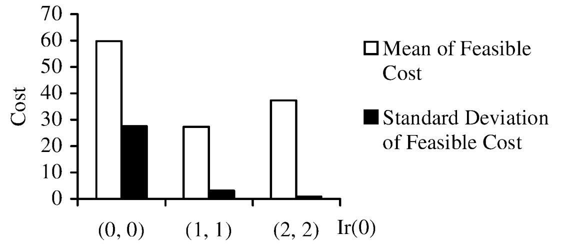

Fig. 4. Performance of the new method under different initial inventories.

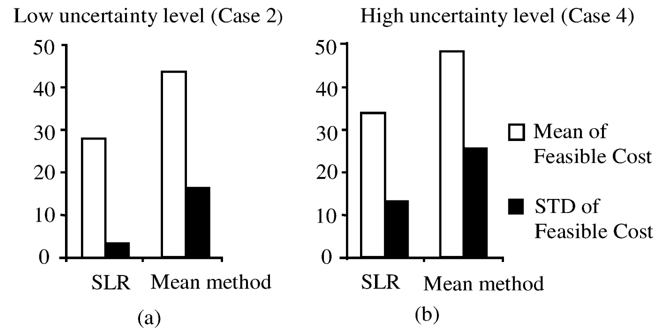

Fig. 5. Performance comparison between the new SLR method and the mean method under different uncertainty levels.

zero, the surrogate subgradient (SSG) method is used to update the multipliers, and the algorithms terminated after a fixed amount of computation time. The data sets and testing results are available at *www.engr.uconn.edu/msl.*

Example 1: Twelve assets containing 24 parts with a total of 72 operations are to be scheduled on four machine types over a time horizon of 90 days. The number of machines per type is two for the overhaul center, and four for the repair shop. Three cases with different levels of initial rotable inventory are considered, with $I_r(0) = (I_{10}, I_{20})$ for Cases 1, 2, and 3 equals (0, 0), (1, 1) and (2, 2), respectively. For all the cases, the variances of asset arrival times and repair times are set to be 0.8 (i.e., $p = 0.8$).

The results of the new SLR method after 32 s of optimization and 500 Monte Carlo simulation runs are summarized in Fig. 4. The duality gap by using the new method for solving Cases 1, 2, and 3 is 8.19%, 3.40% and 1.48%, respectively. It can be seen that with the increase of the initial inventory from zero in Case 1 to one in Case 2, the mean feasible cost decreases significantly. In view that the assembly can be started early by using rotable parts available in the inventory without waiting for original parts to be repaired, assets are assembled in a timely fashion and the mean tardiness cost is reduced significantly from 53.81 in Case 1 to 13.66 in Case 2, leading to the decrease of mean feasible cost from 59.84 to 27.38. The standard deviation of the feasible cost also significantly decreases from 27.53 to 3.03, implying more predictable asset deliveries. Further increase of the initial inventory from one in Case 2 to two in Case 3 has less impact on delivery performance. The increase of the mean inventory holding cost from 13.72 in Case 2 to 26.43 in Case 3, exceeds the decrease of the mean tardiness and earliness cost from 13.66 to 11.33, and this leads to an increase of the mean feasible cost from 27.38 to 37.76. The asset deliveries, however, become more reliable and the standard deviation of the feasible cost decreases further from 3.03 to 0.88. By examining the results obtained for different levels of initial inventory, a good balance between on-time delivery and low rotable inventory cost can be achieved.

To examine the impact of uncertainties, Case 4 is considered where the situation is the same as that of Case 2 except that the variances of uncertain parameters were increased from 0.8 to 1.6 (i.e., $p = 0.6$). The performance of the new SLR method is compared with that of the mean method for Cases 2 and 4 with low and high levels of uncertainties, respectively. Both algorithms were terminated after 32 s and the results are summarized in Fig. 5.

From the figure, it can be seen that the standard deviation of total feasible cost for Case 4 is higher than that for Case 2 for both methods as expected. Comparing the performance of the new SLR method with that of the mean method, the new SLR method generates better schedules with lower feasible costs and lower standard deviations, and the comparison of the mean feasible cost is statistically significant with 80% of confidence for Case 2 when the uncertain level is low, and 85% of confidence for Case 4 when uncertainty level is high. The duality gap by using the new SLR method is 3.4% for Case 2 and 13.34% for Case 4. The duality gap by using the mean method is 45.10% for Case 2, and 60.11% for Case 4. The above demonstrates that the new SLR method can effectively handle uncertainties and obtain high quality schedules without excessive computational requirements.

Example 2: In this example, 100 assets (containing 200 rotable parts) with 600 operations are to be scheduled on four machines types over a time horizon of 90 days. Two cases with different levels of machine utilization were tested to compare the performances of the new SLR method, the traditional SLR method without additional penalty terms, and the FIFO dispatching rule. For both cases, the number of machines per type in the overhaul center is set to be 10. The number of machines per type in the repair shop is set to be 24 for Case 1, and 20 for Case 2, to represent low and high levels of machine utilization, respectively. The initial rotable inventory level $I_r(0)$ is set to 5 for both rotable part types, and the variances of uncertain parameters are set to 0.8 (i.e., $p = 0.8$). The LR-based algorithms were terminated after the 375 s of optimization, and then 100 Monte Carlo simulation runs were performed.

The results are summarized in Table I, where the "constraint violation" is the sum of coupling constraint violations of subproblem solutions. For example, the violation of machine capacity constraints (3) is calculated by $\sum_{kh} \max\{0, \sum_{ej} E(\delta_{ejkh}) - M_{kh}\}$. As the machine utilization level increases, the mean feasible cost obtained by using the new method increases by 31.31% since more assets are delayed when competing for a less number of machines.

The new SLR method outperforms the traditional SLR method by reducing the feasible cost by 28.8% for Case 1, and 39.1% for Case 2. The comparisons are statistically significant with 99% confidence, demonstrating that the new method with additional penalties improves the feasibility of subproblem solutions and leads to high-quality schedules. The new method also significantly outperforms the FIFO dispatching rule with 99% confidence.

TABLE I
PERFORMANCE OF THE METHODS UNDER DIFFERENT MACHINE UTILIZATION LEVELS

Machine utilization	Compared items	SLR (new)	SLR (traditional)	FIFO
Low (Case 1)	Surrogate dual cost	470.2	242.3	/
	CPU time (s)	375	375	/
	Mean feasible cost	554.1	778.0	845.1
	STD of feasible cost	91.5	112.7	110.8
	Duality gap (%)	17.8	221.1	/
	Constraint violation	55.9	632.4	/
High (Case 2)	Surrogate dual cost	626.3	386.1	/
	CPU time (s)	375	375	/
	Mean feasible cost	727.6	1194.8	1338.3
	STD of feasible cost	129.0	157.6	210.3
	Duality gap (%)	16.2	209.5	/
	Constraint violation	59.8	943.5	/

Example 3: Three cases are considered, having 100, 200, and 300 assets to be scheduled over a time horizon of 90 days to demonstrate the scalability of the new method. The initial inventory levels $I_r(0)$ is scaled to be 5 for both part types for Case 1, 10 for Case 2, and 15 for Case 3. The machine utilization levels are set to be approximately equal for the three cases, with the number of machines per type in the overhaul center set to be 10 for Case 1, 20 for Case 2, and 30 for Case 3; and the number of machines per type in the repair shop set to be 20 for Case 1, 40 for Case 2, and 60 for Case 3. The variances of the uncertain parameters are set to be 0.8 for all the cases. In addition, three problem instances were randomly generated and tested for each case. The new SLR algorithms were terminated after 375, 587, and 744 seconds for cases with 100, 200 and 300 assets, respectively. One hundred Monte Carlo runs were then performed and the numerical results are summarized in Table II.

It is observed that high quality schedules are generated by the new method for problems of considerable sizes within a reasonable amount of CPU time. The feasible costs from the method are significantly lower than those from the FIFO dispatching rule with 90% confidence for Subcases 1.3 and 3.2, and 99% confidence for others. The results thus demonstrate the scalability of the method, implying that large problems can be effectively solved without excessive computational requirements.

In the above cases, the algorithms were terminated after a fixed amount of computation time. To further examine the convergence of the new method terminated after different computational times, Subcase 3.3 was solved with the algorithm terminated after 3, 6, 9, and 12 min. One hundred Monte Carlo runs were then performed for each to obtain the sample means of the feasible cost. Testing results are summarized in Fig. 6.

It can be seen from the figures that the sample mean of the feasible cost decreases while the surrogate dual cost increases as the time increases, resulting in a decrease in the duality gap. High quality schedules can be obtained within a reasonable amount of CPU time (9 minutes in this case).

TABLE II
TESTING RESULTS FOR LARGE-SIZED PROBLEMS

Case	CPU Time (s)	Subcase	Surrogate dual cost	Mean feasible cost (SLR)	Gap (%)	Mean feasible cost (FIFO)
1-100 assets	375	1.1	626.3	727.6	16.2	1338.3
		1.2	665.7	779.2	16.7	1848.7
		1.3	524.2	594.1	13.3	815.0
2-200 assets	587	2.1	1084.5	1106.4	2.0	1827.5
		2.2	1307.0	1369.8	4.8	2265.3
		2.3	1241.9	1377.7	10.9	2936.7
3-300 assets	744	3.1	2551.9	2726.7	6.8	5504.5
		3.2	2509.8	2791.4	11.2	3262.7
		3.3	2317.0	2473.6	6.8	4868.1

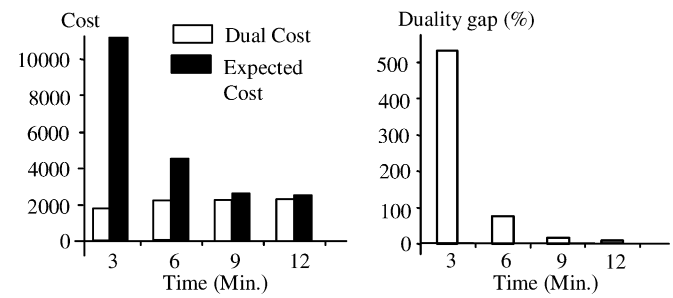

Fig. 6. Testing results for Subcase 3.3 (300 asset problem).

As mentioned at the beginning of this Section, a test bed developed by the United Technologies Research Center was used for testing. A realistic problem may consist of hundreds of assets each with hundreds of major parts, and each part may require tens of repair operations. We have not been able to get realistic data sets since the information needed for optimization is not readily available. The current practice of aerospace aftermarket is relatively traditional in view of how inventory is planned and resources are scheduled. It is hopeful that such implementation barriers will be removed with the advancement of IT infrastructure in the future.

VI. CONCLUSION

In this paper, a novel formulation and the corresponding solution methodology have been established to schedule overhaul and repair services, an important segment of the remanufacturing industry. Rotable inventory are effectively considered in the schedule optimization process through a novel formulation of inventory dynamics in terms of operation beginning and completion times. Adding penalties on coupling constraint violation improves algorithm convergence, and can be extended to other problems having a large number of coupling constraints of various types. The problem is then solved by using LR within the surrogate optimization framework—a new direction for inseparable optimization. Testing results supported by simulation demonstrate that the new method generates high quality schedules with low expected costs and variances, and that the method is robust with respect to different levels of initial inventory, uncertainty, machine utilization, and problem size. The value of rotable inventory to reduce the mean and variance of tardiness cost at the expense of inventory holding cost has also been demonstrated.

ACKNOWLEDGMENT

The authors would like to thank Dr. R. N. Tomastik and Dr. Y. Lin of UTRC for their support of the work.

REFERENCES

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: Theory, Algorithms and Software*. New York: Wiley-Interscience, 2001.

[2] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[3] L. M. M. Custodio, J. J. S. Sentieiro, and C. F. G. Bispo, "Production planning and scheduling using a fuzzy decision system," *IEEE Trans. Robot. Autom.*, vol. 10, no. 2, pp. 160–167, Apr. 1994.

[4] G. W. DePuy and G. E. Whitehouse, "A simple and effective heuristic for the resource constrained project scheduling problem," *Int. J. Prod. Res.*, vol. 39, no. 14, pp. 3275–3287, 2001.

[5] L. Gou, P. B. Luh, and Y. Kyoya, "Holonic manufacturing scheduling: Architecture, cooperation mechanism, and implementation," *Comp. Ind.*, vol. 37, pp. 213–231, 1998.

[6] V. D. R. Guide Jr, "Scheduling using drum-buffer-rope in a remanufacturing environment," *Int. J. Production Research*, vol. 34, no. 4, pp. 1981–1091, 1996.

[7] ——, "Production planning and control for remanufacturing: Industry practice and research needs," *J. Oper. Manage.*, vol. 18, pp. 467–483, 2000.

[8] V. D. R. Guide Jr, V. Jayaraman, and R. Srivastava, "Production planning and control for remanufacturing: A state-of-the-art survey," *Robot. Comput. Integr. Manuf.*, vol. 15, no. 3, pp. 221–230, 1999.

[9] V. D. R. Guide Jr, M. E. Kraus, and R. Srivastava, "Scheduling policies for remanufacturing," *Int. J. Prod. Econ.*, vol. 48, no. 2, pp. 187–204, 1997.

[10] V. D. R. Guide Jr and R. Srivastava, "Inventory buffers in recoverable manufacturing," *J. Oper. Manage.*, vol. 16, no. 5, pp. 551–568, 1998.

[11] V. D. R. Guide Jr, R. Srivastava, and M. E. Kraus, "Product structure complexity and scheduling of operations in recoverable manufacturing," *Int. J. Prod. Res.*, vol. 35, no. 11, pp. 3197–3199, 1997.

[12] W. Jang, "Dynamic scheduling of stochastic jobs on a single machine," *Eur. J. Oper. Res.*, vol. 138, pp. 518–530, 2002.

[13] W. Jang and C. M. Klein, "Minimizing the expected number of tardy jobs when processing times are normally distributed," *Oper. Res. Lett.*, vol. 30, no. 2, pp. 100–106, 2002.

[14] H. A. S. Kim and S. Cho, "Variable target value subgradient method," *Math. Programm.*, vol. 49, pp. 359–369, 1991.

[15] P. Kouvelis, R. L. Daniels, and G. Vairaktarakis, "Robust scheduling of a two-machine flow shop with uncertain processing times," *IIE Trans.*, vol. 32, pp. 421–432, 2000.

[16] E. Kutanoglu and I. Sabuncuoglu, "Experimental investigation of iterative simulation-based scheduling in a dynamic and stochastic job shop," *J. Manuf. Syst.*, vol. 20, no. 4, pp. 264–279, 2001.

[17] S. R. Lawrence and E. C. Sewell, "Heuristics, optimal, static and dynamic schedules when processing times are uncertain," *J. Oper. Manage.*, vol. 15, pp. 71–82, 1997.

[18] P. B. Luh, D. Chen, and L. S. Thakur, "An effective approach for job-shop scheduling with uncertain processing requirements," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 328–339, Apr. 1999.

[19] P. B. Luh, M. Ni, H. Chen, and L. S. Thakur, "A price-based approach for activity coordination in a supply network," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 335–346, Apr. 2003.

[20] H. Matsuura, H. Tsubone, and M. Kanezashi, "Sequencing, dispatching and switching in a dynamic manufacturing environment," *Int. J. Prod. Res.*, vol. 31, no. 7, pp. 1671–1688, 1993.

[21] M. L. Pinedo, *Scheduling -Theory, Algorithms and Systems*. Englewoood Cliffs, NJ: Prentice-Hall, 1995.

[22] P. J. Rondeau, M. A. Vonderembse, and T. S. Ragu-Nathan, "Exploring work system practices for time-based manufacturers: Their impact on competitive capabilities," *J. Oper. Manage.*, vol. 18, no. 5, pp. 509–529, 2000.

[23] R. Shafaei and P. Brunn, "Workshop scheduling using practical data: Part 1: The performance of heuristics scheduling rules in a dynamic job shop environment using a rolling time horizon approach," *Int. J. Prod. Res.*, vol. 37, no. 17, pp. 3913–3925, 1998.

[24] H. M. Soroush, "Optimal sequence in stochastic single machine shops," *Comput. Oper. Res.*, vol. 23, no. 7, pp. 705–721, 1996.

[25] J. Wang, P. B. Luh, X. Zhao, and J. Wang, "An optimization-based algorithm for job shop scheduling," *SADHANA (a Journal of Indian Academy of Sciences on Competitive Manufacturing Systems)*, pt. 2, vol. 22, pp. 241–256, 1997.

[26] Q. Z. Zhai, X. H. Guan, and J. Cui, "Unit commitment with identical units: Successive subproblem solving method based on Lagrangian relaxation," *IEEE Trans. Power Syst.*, vol. 17, no. 4, pp. 1250–1257, Nov. 2002.

[27] X. Zhao, P. B. Luh, and J. Wang, "The surrogate gradient algorithm for lagrangian relaxation method," *J. Optimiz. Theory Appl.*, vol. 100, no. 3, pp. 699–712, Mar. 1999.

Peter B. Luh (S'77–M'80–SM'91–F'95) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1973, M.S. degree in aeronautics and astronautics engineering from M.I.T., Cambridge, MA, in 1977, and Ph.D. degree in applied mathematics from Harvard University in 1980.

Since 1980 he has been with the University of Connecticut, and currently is the SNET Professor of Communications and Information Technologies in the Department of Electrical and Computer Engineering, and a Visiting Professor at the Center for Intelligent and Networked Systems, the Department of Automation, Tsinghua University, Beijing, China. He is interested in planning, scheduling, and coordination of design, manufacturing, and supply chain activities; and schedule, auction, and portfolio optimization and load/price forecasting for power systems.

Dr. Luh is the founding Editor-in-Chief of the newly created IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, an Associate Editor of *IIE Transactions on Design and Manufacturing*, an Associate Editor of *Discrete Event Dynamic Systems*, and was the Editor-in-Chief of IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION (1999–2003).

Danqing Yu received the B.S. degree in automation engineering from Shanghai Jiaotong University, China, in 2000 and the M.S. degree in electrical and computer engineering from the University of Connecticut, in 2003. She is currently working toward the Ph.D. degree in electrical and computer engineering, University of Connecticut, Storrs.

Her research interests are focused on application of optimization theory and algorithms in manufacturing systems and supply chains.

Sada Soorapanth received the B.S. degree from Chulalongkorn University, Thailand, in chemical engineering, the M.S. degree from the University of Houston in industrial engineering, and the Ph.D. degree from the University of Michigan at Ann Arbor in industrial and operations engineering.

She is currently an Industrial Engineer and Prevention Effectiveness Fellow at the Center for Disease Control and Prevention (CDC), Atlanta, GA. Prior to joining CDC, she was a senior research engineer at United Technology Research Center, East Hartford, CT. Her main research interests are in mathematical modeling of system dynamic, discrete-event simulation, decision analysis, and economic evaluation; with applications in manufacturing and public health. She has published papers in academic journals such as *Management Science,* and a *Handbook of Operations Research/Management Science Applications in Health Care.*

Alexander I. Khibnik received the M.S. degree in mathematics from Moscow State University, Moscow, U.S.S.R., in 1973, and Ph.D. degree in mathematics and physics from the Institute of Biological Physics, Russian Academy of Sciences (RAS), in 1985. Since 1973 he has been with the Institute of Mathematical Problems in Biology (formerly Research Computing Center) of RAS in Pushchino, Russia.

Since 1993 he has been with the Theory Center at Cornell University, Ithaca, NY. During that period he concentrated in the area of applied and computational nonlinear dynamical systems and chaos theory, with applications in physics, chemistry, biology and neuroscience. In 1997 he joined United Technologies Research Center (UTRC) where he currently is a Staff Scientist. He is interested in application of dynamical systems, probabilistic, data-driven modeling and complexity reduction methods to problems arising in structural prognostics and health management, and reliability assessment of aerospace structures, life extending controls, condition based maintenance, asset utilization planning and supply chain optimization.

Ravi Rajamani received the B.S. degree in mechanical engineering from the Indian Institute of Technology (IIT), Delhi, India, the M.S. degree in automation from the Indian Institute of Science (IISc), Bangalore, India, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis. Recently he received the MBA degree from the University of Connecticut.

He has been with United Technologies Corporation, East Hartford, CT, for five years, the last two with its Pratt & Whitney division. Prior to this he was with the General Electric Company for ten years. While his interests have varied, his primary focus has been in the area of controls and diagnostics of gas turbines for aerospace and industrial applications. He currently leads the Diagnostics/Prognostics & Health Management area at Pratt & Whitney. He has published more than 35 papers in refereed journals and conference proceedings and is named as an inventor or co-inventor on 12 patents.