Dynamic Surgery Assignment of Multiple Operating Rooms With Planned Surgeon Arrival Times

Zheng Zhang, Xiaolan Xie, Na Geng

Dynamic Surgery Assignment of Multiple Operating Rooms with Planned Surgeon Arrival Times

Zheng Zhang, Xiaolan Xie *Senior Member IEEE* and Na Geng *Member IEEE*

Abstract—This paper addresses the dynamic assignment of a given set of surgeries to multiple identical operating rooms (ORs). Surgeries have random durations and planned surgeon arrival times. Surgeries are assigned dynamically to ORs at surgery completion events. The goal is to minimize the total expected cost incurred by surgeon waiting, OR idling, and OR overtime. We first formulate the problem as a multi-stage stochastic programming model. An efficient algorithm is then proposed by combining a two-stage stochastic programming approximation and some look-ahead strategies. A perfect information-based lower bound of the optimal expected cost is given to evaluate the optimality gap of the dynamic assignment strategy. Numerical results show that the dynamic scheduling and optimization with the proposed approach significantly improve the performance of static scheduling and First Come First Serve (FCFS) strategy.

Note to Practitioners—Hospitals usually consider ORs as the most critical resources and are used to manage patients and human resources to avoid OR idle time. As surgeons are becoming another critical resource, providing surgeons with reliable surgery starting times becomes more and more important. Unfortunately, uncertainties of surgery durations make the problem of setting reliable surgery starting times challenging. We propose an efficient algorithm for dynamic surgery assignment in a multi-OR system by taking into account given surgeon arrival times while best balancing surgeon waiting time and OR idle and overtime. The proposed approach significantly outperforms benchmark strategies. Various sensitivity analyses provide valuable managerial insights to hospital decision makers.

Index Terms—Operating room, appointment scheduling, dynamic scheduling, stochastic programming.

I. Introduction

A SURGERY needs different resources, including an operating room (OR), a post-anesthesia care unit, an intensive care unit, ward, diagnostic equipment and human resources such as surgeons, nurses, anesthesiologists, and other staff. Surgery planning and scheduling need the coordination of all the resources [1]. The OR is one of the most critical and most expensive resources in most hospitals [2]. Hospital managers usually pursue the high utilization of ORs, which

All authors are with the Department of Industrial Engineering and Logistics Management, Shanghai Jiao Tong University, Shanghai, China. E-mail: {zhang, xie, gengna} @sjtu.edu.cn.

Xiaolan Xie is also with the Centre for Biomedical and Healthcare Engineering, UMR CNRS 6158 LIMOS, Ecole Nationale Supérieure des Mines, Saint Etienne, France. xie@emse.fr.

may result in wasted time of other resources.

We have performed a field study in a large university teaching hospital. It is one of the best hospitals in China and is hence overcrowded with patients from all over China. In such a hospital, human resources, especially surgeons, become more and more busy. This hospital uses an open-booking system for surgery planning in which each medical unit is given an amount of OR time each day. At the end of a day, each medical unit submits to an OT its list of surgeries for the next day and the OT is responsible for the surgery-to-OR assignment and surgery scheduling. Currently, top surgeons are scheduled to operate at the beginning of the day in order to avoid waiting. Other surgeons must wait for availability of the OR. As surgery durations are often random, these surgeons usually do not know their surgery starting times and have to wait for the call from the OT. This blind waiting negatively impacts the availability and service quality of these surgeons for their other duties. Therefore, it is very important for the OT to provide surgeons with appointment of arrivals in advance. This can be achieved by two complementary solutions: (i) surgery starting time planning by taking into account uncertainties of surgery durations, (ii) dynamic assignment scheduling to avoid extra waiting of surgeons.

This field study raises two new OR planning/scheduling problems in an OT with multiple ORs shared among different medical units. The first problem, called proactive problem in this paper, consists of determining surgeon arrival times by taking into account OR capacities and random surgery durations. The second problem, called dynamic assignment problem or reactive problem, is the reassignment of surgeries to ORs during the course of a day as surgeries progress by taking into account planned surgeon arrival times. Dynamic scheduling has been shown to contribute a larger portion of the cost reduction in many systems such as surgery scheduling [3] and patient flow [4, 5], which motivates us to focus on the reactive problem.

This paper addresses the dynamic assignment problem of a multi-OR system for a given set of surgeries with planned surgeon arrival times. Dynamic decisions are made to determine the next surgery in the OR released at each surgery completion event. The objective is to minimize the total expected cost incurred by surgeon waiting, OR idling, and OR overtime. A multi-stage stochastic programming model is formulated. A solution strategy based on a two-stage stochastic integer programming (2–SIP) approximation is proposed to deal with the dynamic assignment of surgeries to ORs. A mathematical formulation is given for the proactive problem to determine

surgeon arrival times. Extensive numerical experiments are performed to show the performance of the proposed methods.

Our contributions are multifold. First, we propose a multi-stage stochastic programming formulation of the dynamic assignment problem by allowing decision revision at all surgery completion events. It is more realistic than the model of [3] which allows only one given decision revision point. Second, we propose a dynamic assignment algorithm based on a 2−SIP approximation, look-ahead strategies and some valid inequalities. Third, we propose an efficient algorithm for surgeon arrival times determination based on local search and a new optimal rule for sequencing two surgeries in an OR. Note that this paper is an extended version of a preliminary work [6] which proposed a dynamic assignment model for small-size problems based on mixed-integer programming.

The remainder of this paper is organized as follows. Section II is a brief review of relevant literature. Section III gives a detailed description of the dynamic assignment problem and proposes a multi-stage stochastic programming model and a lower bound of the optimal cost. Section IV proposes various 2−SIP approximations of the dynamic assignment problem. Section V addresses the optimization of the proactive problem. Numerical results are presented in Section VI. Section VII concludes the paper.

II. LITERATURE REVIEW

This section reviews existing work on appointment scheduling (AS). This issue has been widely studied in the health care service literature including surgery scheduling [7] and clinic patient scheduling [8, 9]. A complete review of operating room scheduling is given by [10], and detailed reviews of AS in health care systems are provided by [11, 12]. We limit ourselves to work that take into account uncertainties. In this section, we use "customer" and "surgery" interchangeably.

Earlier work on AS concerns single-server systems. A newsvendor model was proposed in [13] for the two-customer case. An analytical solution of optimal appointment times was derived, and convex ordering was proved to give an optimal customer sequence under the condition of equal unit waiting cost. Based on [13], a stochastic linear program was proposed in [14] for the case of multiple customers. The optimal customer allowances (i.e., scheduled inter-arrival times) were characterized by a dome shape with allowances first increasing then decreasing.

Several extensions were proposed by considering customer sequencing, discrete AS and robust AS, etc. Several heuristics were proposed in [15] for sequencing customers, including smallest variance first (SVF), smallest mean service duration first, smallest coefficient of variation first and an interchange sequence heuristic. Numerical experiments showed that SVF performed the best among these heuristics. A 2−SIP formulation was proposed in [16] for joint optimization of customer sequencing and arrival times. It was proved that the sample average approximation problem belongs to NP-complete when the waiting costs differ among customers. Problem instances

with up to 12 customers were solved by using L-shaped algorithm similar to [14], and this method was shown to be significantly better than SVF rule and a perturbed sort by variance rule. Under similar conditions as in [13], convex ordering was also proved in [17] to give optimal ordering between two blocks of customers of i.i.d. service durations within each block.

Dynamic AS with random service duration and random number of customers was considered in [18, 19]. In [18], customers were scheduled dynamically, one at a time under the condition of FCFS, as they requested appointments. The value of stochastic solution is significant in the dynamic environment. The assumption of FCFS was further relaxed in [19] to jointly optimize the sequence and arrival scheduling in dynamic environment. A 2−SIP model was proposed and properties of optimal customer sequence were derived for the two-customer case.

A so-called discrete AS problem with discrete service duration distribution was addressed in [20] for a single server with a given sequence of customers. It was proved that there exists an optimal integer arrival schedule that can be found in polynomial time on mild condition of cost functions. Closed-form formulas were proposed in [21] for the subdifferential of the objective function. A bi-criteria appointment sequencing problem was considered in [22] and a set of Pareto solutions were derived using genetic algorithm. Robust optimization of AS was considered in [17, 23]. The service durations were random with limited information of mean and covariance estimates. A convex conic optimization model was formulated with the objective to minimize the expected cost under the worst distribution. Specifically, [23] considered a robust arrival scheduling problem and [17] considered a robust appointment sequencing problem.

We now present the literature of AS of multi-OR systems. An assignment scheduling problem in a multi-OR setting was considered in [24] by considering OR opening costs and overtime costs. Decisions include the number of ORs to open each day and assignments of surgeries to ORs. A 2−SIP model was proposed under the full knowledge of surgery duration distributions and a robust counterpart under limited knowledge (i.e., upper bounds and lower bounds on surgery durations).

The 2−SIP model of [24] was extended in [3] by allowing a one-time re-assignment scheduling in dynamic multi-OR setting. Specifically, surgeries not yet started at the re-assignment point could be revised. The joint optimization of the surgery assignment and re-assignment was very complicated due to the extremely large state space. A special case with the re-assignment decision made under perfect information was solved by using L-shaped algorithm and a progressive hedging-based heuristic. The re-assignment was shown to significantly reduce the cost especially when the overtime cost is high but the benefit depends on the surgery sequence. The LEPT (the longest mean service duration first) sequence was shown to perform the best among several heuristics.

An OR scheduling and surgery AS problem was considered in [7] in a multi-OR setting. Rescheduling of surgery assignment was not taken into account. Surgeries could be

performed in parallel in a multi-OR operating theater by a primary surgeon with the assistance of other staff. The decisions included the number of ORs to open each day, the assignment of surgeries to ORs, the sequence of surgeries within each OR, and the arrival time for each surgeon. A 2–SIP model was proposed and solved by an L-shaped-based branch-and-cut algorithm. Numerical experiments showed that the benefit of OR pooling is significant.

Compared with the aforementioned literature, this paper is the first to consider the multi-stage dynamic assignment problem in a multi-OR setting. It will be shown through numerical experiments that the proposed dynamic assignment scheduling brings a significant cost reduction.

III. PROBLEM DESCRIPTION AND MATHEMATICAL MODELS

In this section, we describe and formulate the dynamic assignment problem as a multi-stage stochastic program. A mixed integer programming model is then proposed to provide a lower bound of the optimal cost for dynamic assignment problem.

A. Problem setting

This paper considers the problem of the dynamic assignment scheduling for a multi-OR operating theater and for a given day. The operating theater is composed of M identical and interchangeable ORs. A given set of N surgeries are planned for the day under consideration. Surgery durations are independent random variables with known distributions, and are independent of ORs and the dynamic scheduling.

Each surgery is associated with a surgeon and each surgeon is assumed to have only one surgery in a day. As a result, we will use interchangeably the terms of surgery and surgeon. Each surgeon has been given a planned surgery starting time that was called promised surgery starting time in our preliminary work [6] and called surgeon arrival time in this paper. Each surgeon is assumed to arrive on time. Patients are assumed to be available for surgery.

The operation of the operating theater during the day is as follows. At the beginning of the day, say at time $t = 0$, the first M surgeons arrive and immediately start their surgery separately in the different ORs. When one surgery completes and the OR is released, the dynamic assignment strategy selects and assigns an unassigned surgery i to be the next surgery in the released OR. The surgery becomes the state of waiting for the surgeon or in process if the surgeon arrives. The OR is not available for other surgeries till the completion of i. As a result, there are $N-M$ decision epochs corresponding to the first $N-M$ surgery completion events among N such events.

The problem is to determine a dynamic assignment strategy with the objective to minimize a weighted linear combination of OR idle time, OR overtime and surgeon waiting time. The OR idle time is incurred if an OR is assigned to a surgery but the surgeon is not yet arrived. The OR overtime is incurred if an OR is released by its last surgery after a regular opening time called OR session. The surgeon waiting time is incurred if no OR is available at the arrival of a surgeon.

Note that we restrict ourselves to event-based strategies, i.e., the assignment decisions are made at the occurrence of surgery completion events. Once an unassigned surgery is assigned to the released OR, the decision will not change. This restriction is quite reasonable in practices.

Although this paper focuses on the dynamic assignment scheduling, we will provide an efficient algorithm for determination of surgeon arrival times in Section V.

We now justify major assumptions of our model through a series of remarks.

Remark 1: Each surgeon is assumed to have only one surgery a day. Results of this paper directly apply if multiple surgeries per surgeon are allowed but all surgeries of each surgeon are performed in the same OR without interruption. In this case, surgeries of each surgeon can simply be considered as a single aggregated surgery. An interesting future research is the dynamic assignment scheduling with fixed order constraint for all surgeries of each surgeon. In this case, it is necessary to take into account this precedence constraint, waiting of the same surgeon for different surgeries.

Remark 2: ORs are assumed to be identical. This assumption is often made in OR planning and scheduling literature [7, 24]. To increase their usage, ORs tend to be housed in an integrated OT to share material/human resources and to serve different medical units. ORs with special equipment are often dedicated to some medical units such as cardio-surgery and are managed separately. However, our approach can be extended to the general case where each surgery can only be performed in a subset of ORs.

Remark 3: Patients are assumed to be available for surgery. No-shows due to surgery cancellations or surgeon absenteeism are not considered. Other uncertainties such as emergency demands and delayed patient preparations also disturb the operation of the OR. Literature on OR planning and scheduling with taking into account these uncertainties is scarce. The dynamic scheduling approach of this paper can be extended to take into account these uncertainties by allowing rescheduling at extra events such as emergency arrival and the patient's no-show being notified. The proactive problem of planning surgeon arrival by taking into these uncertainties in a multi-OR setting is an interesting future research issue.

Remark 4: We consider surgeon waiting instead of the usual patient waiting costs in most healthcare scheduling literature. This is appropriate as most surgery patients are inpatients and the focus of this paper is to release surgeons from blind waiting. For this reason, a surgery cannot start before the planned arrival time of the surgeon. However, surgeon waiting costs and patient waiting costs lead to the equivalent formulation. The primary difference is in their estimation.

B. Notation

To summarize, the dynamic assignment problem is characterized by the following notations.

N number of surgeries or surgeons;

M number of identical ORs with $M < N$;

J set of surgeries or surgeons indexed by i, j;

E set of ORs indexed by r;

A_i arrival time of surgeon i with $A_i = 0, \forall i \in \{1, \ldots, M\}$ and $A_{i-1} \le A_i, \forall i \in \{M+1, \ldots, N\}$;

T OR session, i.e., the regular capacity of each OR;

α unit OR overtime cost;

β_i unit waiting time cost of surgeon i;

p_i random duration of surgery i;

$\bar{p}_i, \sigma_i$ mean and standard deviation of p_i;

$F_i(\cdot)$ probability distribution function of p_i.

The unit OR idle time cost is assumed, without loss of generality, to be 1. Otherwise, the overtime cost α and the surgeon waiting cost β_i can always to be scaled to make this assumption true.

Let Ω be the set of all possible realizations also called scenarios of the surgery duration vector. Let $p_i(\omega)$ denote the duration of surgery i under scenario ω. Each scenario can be represented by a random vector $\boldsymbol{p}(\omega) = \{p_1(\omega), \ldots, p_N(\omega)\}$.

C. A multi-stage stochastic programming modeling

Additional notations are needed to rigorously characterize the dynamic assignment problem. Each event is characterized as follows:

$e(n)$ nth completed surgery;

t_n completion time of $e(n)$ with $t_0 = 0 \le t_1 \le \cdots \le t_N$.

The decision made at event n is denoted as:

$i(n)$ surgery to be performed next in the OR released at event n.

The states of the system at event n after decision $i(n)$ are characterized by the following state variables:

$J(n)$ set of completed or on-going surgeries including $i(n)$ at event n with $J(0) = \{1, 2, \ldots, M\}$;

$\bar{J}(n)$ set of on-going surgeries at event n with $\bar{J}(0) = J(0)$;

$a_i(n)$ elapsed time or age of surgery i at event n;

s_i starting time of surgery i;

C_i completion time of surgery i.

The decision made at each event n is according to some strategy π_n based on state information available at this point: $J(n-1), \bar{J}(n-1), \boldsymbol{a}(n), e(n), t_n$ and fulfilling the following conditions:

$$i(n) = \pi_n(J(n-1), \bar{J}(n-1), \boldsymbol{a}(n), e(n), t_n) \tag{1}$$

$$i(n) \in J \backslash J(n-1) \tag{2}$$

The scheduling problem at each decision epoch k for $k = 1, \ldots, N\text{-}M$ can now be formulated as follows.

$$V_k = \min_{\pi_k, \cdots, \pi_{N-M}} g_k + \sum_{\tau=k+1}^{N} E_\omega(g_\tau) \tag{3}$$

subject to constraints (1) and (2) for $n = k, \ldots, N\text{-}M$ and

$$s_{i(n)} = \max(t_n, A_{i(n)}), \forall n \le H \tag{4}$$

$$g_n = (t_n - A_{i(n)})^+ + \beta_{i(n)}(A_{i(n)} - t_n)^+, \forall n \le H \tag{5}$$

$$J(n) = \{i(n)\} \cup J(n-1), \forall n \le H \tag{6}$$

$$\bar{J}(n) = \{i(n)\} \cup \bar{J}(n-1) \backslash \{e(n)\}, \forall n \tag{7}$$

$$C_i = s_i + p_i, \forall i \in J \tag{8}$$

$$e(n+1) = \arg\min_{i \in \bar{J}(n)} C_i, \forall n \tag{9}$$

$$t_{n+1} = \min_{i \in \bar{J}(n)} C_i, \forall n \tag{10}$$

$$g_n = \alpha(t_n - T)^+, \forall n > H \tag{11}$$

where $(x)^+ = \max(0, x)$, and $H = N\text{--}M$.

The objective function V_k includes two parts: the stage cost for assigning the surgery $i(k)$ at the current event k and the total expected cost related to remaining surgeries and OR overtime. It depends on the state observed at event k. We also extend the definition of V_k to $k = 0$ and $k > H$ with the convention $\{i(k)\} = \varnothing$ and $g_0 = 0$. As a result, V_0 is the total expected cost of the dynamic assignment problem of the day.

Constraints (4) determines the starting time of each newly assigned surgery at each event. Constraint (5) is the stage cost of assigning surgery $i(n)$ at event n, including both OR idle cost and surgeon waiting cost for the first $N\text{--}M$ events. Constraint (6) and (7) update $J(n)$ and $\bar{J}(n)$. Constraint (8) determines the completion time of each on-going surgery. Constraints (9) and (10) determine the next event and the event time. Constraint (11) defines the stage cost at event $n > N\text{--}M$, i.e., the overtime cost of each OR for the last M events.

D. A lower bound cost

This subsection provides a lower bound on the optimal cost of the dynamic assignment problem V_0. The lower bound cost is obtained by applying the *wait-and-see* or perfect information solution [25]. Assuming that all surgeries durations are known immediately at time 0, surgery assignment and sequencing are determined independently for each scenario ω to minimize the total cost $\theta(\omega)$. $E_\omega[\theta(\omega)]$ is the lower bound cost proposed in this section. Similar lower bound of V_k can also be defined.

Each scenario ω is associated with the following variables:

$x_{ir}(\omega)$ binary variable equal to 1 if surgery i is assigned to OR r;

$y_{ij}(\omega)$ binary variable equal to 1 if surgery i precedes surgery

j in the same OR ($j \neq i$);

$C_{ir}(\omega)$ completion time of surgery i in OR r;

$I_r(\omega)$ total idle time of OR r;

$O_r(\omega)$ overtime of OR r;

$W_i(\omega)$ waiting time of surgeon i;

$R_r(\omega)$ completion time of surgery started at time 0 in OR r also called release time of OR r. Without loss of generality, $R_r(\omega) = p_r(\omega)$.

The lower bound problem is formulated as

$$LB = E_\omega[\theta(\omega)] \leq V_0 \tag{12}$$

where

$$\theta(\omega) = \min \sum_{r \in E} I_r(\omega) + \alpha \sum_{r \in E} O_r(\omega) + \sum_{i \in \hat{J}} \beta_i W_i(\omega) \tag{13}$$

subject to:

$$\sum_{r \in E} x_{ir}(\omega) = 1 \tag{14}$$

$$y_{ij}(\omega) + y_{ji}(\omega) \geq x_{ir}(\omega) + x_{jr}(\omega) - 1 \tag{15}$$

$$C_{ir}(\omega) \leq M x_{ir}(\omega) \tag{16}$$

$$C_{ir}(\omega) \geq \big(R_r(\omega) + p_i(\omega)\big) x_{ir}(\omega) \tag{17}$$

$$C_{jr}(\omega) \geq C_{ir}(\omega) + p_j(\omega) - M\big(1 - y_{ij}(\omega)\big) - M\big(2 - x_{ir}(\omega) - x_{jr}(\omega)\big) \tag{18}$$

$$C_{ir}(\omega) \leq I_r(\omega) + R_r(\omega) + \sum_{j \in \hat{J}} p_j(\omega) x_{jr}(\omega) \tag{19}$$

$$O_r(\omega) \geq C_{ir}(\omega) - T \tag{20}$$

$$O_r(\omega) \geq R_r(\omega) - T \tag{21}$$

$$\sum_{r \in E} C_{ir}(\omega) = A_i + W_i(\omega) + p_i(\omega) \tag{22}$$

$$I_r(\omega), O_r(\omega), W_i(\omega) \geq 0, x_{ir}(\omega), y_{ij}(\omega) \in \{0,1\} \tag{23}$$

$\forall i \in \hat{J}, r \in E, \omega \in \Omega$ where M is a large number and $\hat{J} = J \setminus J(0)$.

The objective function (12) defines the lower bound of V_0 as the assignment and sequencing of any feasible dynamic assignment strategy are also a feasible solution fulfilling (14)-(23) for any scenario. Constraint (14) ensures that each surgery is assigned to exactly one OR. Constraint (15) ensures a precedence relation between two surgeries assigned to the same OR. Constraints (16) and (17) link surgery completion times to OR assignment. Constraint (18) links completion times to precedence relation. Constraint (19) determines OR idle times. Constraint (20) and (21) define OR overtimes. Constraint (22)

defines surgeon waiting times. Constraint (23) defines the domains of variables.

Note that any realization of a dynamic assignment strategy fulfilling (1)-(11) is a solution of the lower bound problem and hence fulfills (14)-(23). Constraint (14) obviously holds as any surgery is either assigned at time 0 to an OR or assigned at an event to the OR released. As an unassigned surgery is assigned to an OR only after the completion of the on-going surgery, (15) always holds. Constraints (16)-(22) are used to define surgery completion time, surgeon waiting time, OR idle time and OR overtime and hence trivially hold for any realization of a dynamic assignment strategy.

IV. TWO STAGE STOCHASTIC INTEGER PROGRAMMING APPROXIMATION

As the dynamic assignment problem is very difficult due to extremely large state space, we propose in this section various 2–SIP approximations to the multi-stage stochastic programming problem V_k at event $k = 1, \ldots, N–M$.

Roughly speaking, the 2-SIP approximation consists of selecting a surgery l to minimize the first stage cost related to l and the remaining cost. Two methods are proposed to estimate the remaining cost. The first one in Section IV.B simply approximates the remaining cost by the stage cost at next event. The second one in Section IV.C uses some simple scheduling rules to estimate the remaining cost. We choose these methods for their short computational time and their good numerical performances.

A. A general approximation scheme

The 2-SIP approximation relies on the following form of the dynamic scheduling cost function:

$$V_k = \min_{l \in J \setminus J(k-1)} \hat{g}_{lk} + \Theta_{lk}$$

$\hat{g}_{lk}$ is the first stage cost if surgery l is selected at event k, i.e., $i(k) = l$. It is related to the waiting time of surgeon l and idle time of the OR released at event k for the arrival of surgeon l. Its formal definition is as follows:

$$\hat{g}_{lk} = (A_l - t_k)^+ + \beta_l(t_k - A_l)^+ \tag{24}$$

Θ_{lk} is the *expected value of the recourse function* also called the second stage cost related to all other unassigned surgeries and OR overtime, i.e., related to $\sum_{\tau=k+1}^{N} E_\omega(g_\tau)$ of (3). The exact evaluation of $\sum_{\tau=k+1}^{N} E_\omega(g_\tau)$ is again a complex multi-stage stochastic programming problem.

The evaluation of Θ_{lk} strongly depends the termination of the on-going surgeries or equivalently OR release times $R_r^{lk}(\omega)$ with assignment of surgery l at event k defined as follows:

$$R_r^{lk}(\omega) = max\big(A_{i_{rlk}}, t_k\big) - a_{i_{rlk}}(k) + p_{i_{rlk}}(\omega) \tag{25}$$

where i_{rlk} is the on-going surgery in OR r at event k. The probability distribution of each on-going surgery i becomes the following conditional probability depending on its age $a_i(k)$:

$$P\big(p_i \le x | p_i \ge a_i(k)\big) = \frac{F_i(x) - F_i\big(a_i(k)\big)}{1 - F_i\big(a_i(k)\big)} \tag{26}$$

The first stage decision $i(k)$, i.e., the selection of next surgery for the OR $r(k)$ released at event k, is made by exhaustive search by algorithm 1. The best candidate surgery with minimal total two-stage cost is selected as the next surgery in the released OR. The basic steps to solve the two-stage approximation problem at a given event k are described as the following algorithm.

Algorithm 1: *Surgery assignment at event k.*
Step 1. Initialization: $z^* \leftarrow \infty$;
Step 2. Update OR release times $R_r^{lk}, \forall r \ne r(k)$ by (25);
Step 3. For each remaining surgery l
- Calculate $\hat{g}_{lk}$ according to (24);
- Update release times $R_{r(k)}^{lk}$ according to (25);
- Evaluate the second stage cost Θ_{lk};
- If $z^* > \hat{g}_{lk} + \Theta_{lk}, i(k) \leftarrow l, z^* \leftarrow \hat{g}_{lk} + \Theta_{lk}$;
Step 4. Assign surgery $i(k)$ as the next surgery in OR $r(k)$.

The main difficulty of this two-stage approximation approach is the evaluation of the second stage cost. A straightforward approach consists in evaluating as in the lower bound problem by assuming that all random outcomes, i.e., surgery durations of all on-going and unassigned surgeries, are known immediately for all scenarios. More specifically,

$$\Theta_{lk} \approx \min E_\omega \left\{ \sum_{r \in E} I_r(\omega) + \alpha \sum_{r \in E} O_r(\omega) + \sum_{i \in J^{lk}} \beta_i W_i(\omega) \right\} \tag{27}$$

subject to constraints (14) – (23) with $\hat{J}$ replaced by the set $J^{lk} = J \backslash J(k-1) \backslash \{l\}$ and $R_r(\omega)$ by $R_r^{lk}(\omega)$.

The expected recourse function (27) is still too time-consuming to solve directly. In order to estimate efficiently the second stage cost Θ_{lk}, we propose a one-period look ahead heuristic and a rule based multi-period look ahead heuristic in subsection B and subsection C, respectively.

B. One-period look ahead (OPLA) heuristic

OPLA heuristic approximates Θ_{lk} by the average stage cost observed at the next event $k+1$ and hence selects the surgery $i(k)$ that minimizes the total cost observed at stage k and stage $k+1$.

More specifically, with OPLA,

$$\Theta_{lk} = \min_{j \in J \backslash J(k-1) \backslash \{l\}} \theta_{jlk}$$

where θ_{jlk} is the expected cost of stage $k+1$ conditioned on the state observed at event k and surgery l selected at event k and surgery j selected at event $k+1$. It is defined as follows:

$$\theta_{jlk} = E_\omega \left[\big(A_j - t_{k+1}^l(\omega)\big)^+ + \beta_j \big(t_{k+1}^l(\omega) - A_j\big)^+ \right] \tag{28}$$

where $t_{k+1}^l(\omega)$ is the epoch of event $k+1$ if l is selected at event k and hence $t_{k+1}^l(\omega) = \min_{r \in E} R_r^{lk}(\omega)$. The above expectation is taken under the condition of the state observed at event k.

Proposition 1. $\theta_{jlk} \ge \underline{\theta}_{jlk}$ with

$$\underline{\theta}_{jlk} = \big(A_j - E_\omega[t_{k+1}^l]\big)^+ + \beta_j \big(E_\omega[t_{k+1}^l] - A_j\big)^+ \tag{29}$$

Proof. Since the function $f(x) = (a - x)^+ + b(x - a)^+$ is convex, by Jensen's inequality [25], $E\left[f\big(t_{k+1}^l(\omega)\big)\right] \ge f\big(E[t_{k+1}^l(\omega)]\big)$, which completes the proof. Q.E.D.

Algorithm 2 presents the main steps to estimate the remaining cost by OPLA heuristic.

Algorithm 2: *Approximation of Θ_{lk} with OPLA*
Step 1. (Initialization): $\Theta_{lk} \leftarrow \infty$;
Step 2. (next event time): determine the next event time $t_{k+1}^l(\omega) = \min_{r \in E} R_r^{lk}(\omega)$ for each scenario ω and its mean $E_\omega[t_{k+1}^l]$;
Step 3. Determine $\underline{\theta}_{jlk}$ using (29);
Step 4. For each remaining surgery $j \in J \backslash J(k-1) \backslash \{l\}$
- If $\underline{\theta}_{jlk} + \hat{g}_{lk} \ge z^*$, go to Step 4 for the next j;
- Calculate θ_{jlk} using (28), $\Theta_{lk} \leftarrow \min(\Theta_{lk}, \theta_{jlk})$;
- If $\hat{g}_{lk} + \Theta_{lk} < z^*, z^* \leftarrow \hat{g}_{lk} + \Theta_{lk}$.

Algorithm 2 is used jointly with Algorithm 1 in order to select the surgery l at event k that minimizes the total cost $\hat{g}_{lk} + \Theta_{lk}$ with Θ_{lk} evaluated by OPLA with Algorithm 2. The lower bound from Proposition 1 is used to avoid the time-consuming computation of θ_{jlk} when $\underline{\theta}_{jlk} + \hat{g}_{lk} \ge z^*$ where z^* is the current minimal total cost.

C. Rule-based multi-period look ahead (MPLA) heuristic

The second method approximates the second stage cost Θ_{lk} with the total cost incurred by all remaining surgeries and OR overtime under some given dynamic assignment strategies. For the sake of computational complexity, it is based on the following simple scheduling rules.

Rule 1: Remaining surgeries are assigned in the scenario-independent order of minimal expected first stage cost, i.e., the surgery selected at event $n > k$ minimizes $E_\omega\big[\hat{g}_{jn}(\omega)\big]$ among all surgeries in $J \backslash J(n-1)$ with

$$\hat{g}_{jn}(\omega) = \big(A_j - t_n(\omega)\big)^+ + \beta_j \big(t_n(\omega) - A_j\big)^+ \tag{30}$$

Rule 2: Remaining surgeries are selected in non-decreasing order of their surgeon arrival times A_i at any event $n > k$.

A better but more time-consuming alternative of Rule 1 is the scenario-dependent ordering determined by OPLA heuristic at each event epoch $t_n(\omega)$.

Algorithm 3 presents the main steps to estimate the remaining cost. At each time, we obtain two remaining costs by using rule 1 and rule 2, respectively. We will choose the smallest one as final remaining cost.

Algorithm 3: *Approximation of Θ_{lk} by MPLA heuristic*

Step 1. (Initialization): Set the event counter $n=k+1$;
Step 2. Determine next event time $t_n(\omega) = \min_{r \in E} R_r^{lk}(\omega)$;
Step 3. Determine $\hat{g}_{jn}(\omega), \forall j \in J \backslash J(n-1)$ by (30);
Step 4. Determine the surgery $i(n)$ according to Rule 1 and the stage cost $g_n = E_\omega[\hat{g}_{i(n),n}(\omega)]$;
Step 5. Assign $i(n)$ to the OR $r(n, \omega)$ released at event n and update OR release times $R_r^{lk}(\omega)$;
Step 6. $n = n+1$, repeat steps 2-5 till $n = N-M$;
Step 7 (Second stage cost):

$$\Theta_{lk}^{R1} = \sum_{m=k+1}^{N-M} g_m + E_\omega \left[\alpha \sum_{r \in E} (R_r^{lk}(\omega) - T)^+ \right]$$

Step 8. Repeat steps 1-7 with Rule 1 replaced by Rule 2 and let Θ_{lk}^{R2} be the second stage cost;
Step 9. $\Theta_{lk} = \min(\Theta_{lk}^{R1}, \Theta_{lk}^{R2})$.

As for Algorithm 2, Proposition 1 can also be used to accelerate the determination of the next surgery $i(n)$ in Steps 3 and 4. We also stop the simulation of Rule 1 or Rule 2 when the following holds:

$$\hat{g}_{lk} + \sum_{m=k+1}^{n} g_m \geq z^*$$

(31)

where z^* is the current minimum total cost. The use of the lower bound of Proposition 1 and condition (31) cuts by 3 the computation time needed for dynamic assignment scheduling with Algorithms 2 and 3.

V. THE PROACTIVE PROBLEM

This section provides a mathematical formulation and an algorithm for the proactive problem, i.e., the optimization of surgeon arrival times. The goal is to determine surgeon arrival times such that the dynamic assignment strategies of the previous sections perform better.

We propose a two-stage stochastic programming approximation of this proactive problem. The first stage determines the surgeon arrival times A_i, surgery-to-OR assignment x_{ir} and surgery sequencing y_{ij}. The second stage evaluates the cost when all random durations are realized.

The proactive problem denoted $P(S)$ is formulated as follows:

$$\min E_\omega \left\{ \sum_{r \in E} I_r(\omega) + \alpha \sum_{r \in E} O_r(\omega) + \sum_{i \in J} \beta_i W_i(\omega) \right\}$$

(32)

subject to constraints (14) – (23) with $\hat{J}$ replaced by J, $x_{ir}(\omega)$ by x_{ir}, $y_{ij}(\omega)$ by y_{ij}, and $R_r = 0$ and

$$0 \leq A_i \leq T, \qquad \forall i \in J$$

(33)

When the assignment and sequencing decisions are given, the proactive problem $P(S)$ becomes a linear program (LP) model, which can be efficiently solved. However, the problem of joint optimization of the arrival times as well as the assignment and sequencing decisions is a stochastic integer program (SIP), which is quite difficult to solve. The L-shaped algorithm was used in [7, 16] to solve SIP. However, the algorithm is quite time consuming.

In order to efficiently solve the proactive model, we proposed a heuristic approach. We first propose sequencing heuristic for the single OR problem denoted SOR. We then propose a local search heuristic for the optimization of the multi-OR problem. Local search improves the assignment decision and each complete surgery assignment is evaluated by the heuristic of the SOR problem.

The SOR problem is solved by the following heuristic.

SOR heuristic: Sequencing surgeries in the decreasing order of the following index:

$$\varphi_i = \beta_i / (\beta_i + 1 + \epsilon\alpha)\sigma_i$$

where $\epsilon = P\left(\sum_{i \in J} p_i > T\right)$ is the overtime probability without OR idle time.

The SOR heuristic is motivated by the following results for optimal sequencing of two surgeries in a single OR.

Proposition 2. Consider an SOR problem with two surgeries of uniformly distributed durations $p_i = U(l_i, u_i)$ such that $u_i \leq T$. If $u_1 + u_2 \leq T$ or $l_1 + l_2 \geq T$, the sequence $\{1, 2\}$ is optimal if and only if $\varphi_1 > \varphi_2$.

Proof. Consider first the sequence $\{1, 2\}$ and hence $s_1 = 0$. It can be easily shown that $s_2 \in [l_1, u_1]$.

Consider first the case $u_1 + u_2 \leq T$. The total cost $z_{12}(s_2)$ of the sequence $\{1, 2\}$ is

$$z_{12}(s_2) = E[(s_2 - p_1)^+] + \beta_2 E[(p_1 - s_2)^+]$$

By taking derivative of $z_{12}(s_2)$ and letting it equal to 0,

$$s_2^* = l_1 + \beta_2 \Delta_1 / (\beta_2 + 1)$$

$$z_{12}(s_2^*) = \beta_2 \Delta_1 / 2(\beta_2 + 1)$$

where $\Delta_i = u_i - l_i$. Similarly, the optimal cost of sequence $\{2, 1\}$ under the solution s_1^* is $z_{12}(s_1^*) = \beta_1 \Delta_2 / 2(\beta_1 + 1)$. Since

$\varphi_i = 2\sqrt{3}\,\beta_i/(\beta_i + 1)\Delta_i$, the sequence $\{1, 2\}$ is optimal if and only if $\varphi_1 > \varphi_2$.

Consider now the case $l_1 + l_2 \geq T$. Let $\bar{p}_i = E[p_i]$.

$$z_{12}(s_2) = E[(s_2 - p_1)^+] + \beta_2 E[(p_1 - s_2)^+] + \alpha(\bar{p}_1 + \bar{p}_2 + E[(s_2 - p_1)^+] - T)$$

leading to:

$$s_2^* = l_1 + \beta_2 \Delta_1/(\beta_2 + 1 + \alpha)$$

$$z_{12}(s_2^*) = \beta_2 \Delta_1 (1 + \alpha)/2(\beta_2 + 1 + \alpha) + \alpha(\bar{p}_1 + \bar{p}_2 - T)$$

Since $\varphi_i = 2\sqrt{3}\,\beta_i/(\beta_i + 1 + \alpha)\Delta_i$, the sequence $\{1, 2\}$ is optimal if and only if $\varphi_1 > \varphi_2$. Q.E.D.

SOR heuristic becomes the SVF rule of [15] when all surgeons have equal waiting cost, i.e., $\beta_1 = \cdots = \beta_N$. It becomes highest waiting penalty first (HPF) rule (which are widely used in the hospitals we observed) when all surgeries have equal variances, i.e., $\sigma_1 = \cdots = \sigma_N$, SOR heuristic is close to SVF when the waiting time costs of surgeons are much higher than that of OR idle time and OR overtime such that $\beta_i \gg 1 + \alpha, \forall i$.

For multi-OR problems, we propose a simple two-phase heuristic. The first phase determines an initial surgery assignment in order to best balance the expected workloads $\sum_{i \in J} x_{ir} \bar{p}_i / T$ of different ORs, i.e.,

$$\min \max_{r \in E} \sum_{i \in J} x_{ir} \bar{p}_i$$

subject to $\sum_{r \in E} x_{ir} = 1, \forall i \in J$ and $x_{ir} \in \{0, 1\}$. It can be easily solved by any standard MIP solver.

The second phase improves the initial assignment by local search. The local search is performed by OR pairs. For each OR pair, pair-wise swaps are considered for surgeries of the two ORs. The best improving swap is accepted, and the process continues until no improving swap exists for the OR pair under consideration. Local search then tries to improve the assignment of other OR pairs. In the whole process, when the set of surgeries assigned to an OR changes, the SOR heuristic is applied to evaluate its new cost. This local search algorithm is termed ASB (All OR-pairs, Swap-based Best improving).

The main steps of ASB are as follows where SOR(P) denotes the cost of the single OR problem with the set P of surgeries. It is solved by the SOR heuristic to determine the sequence and an LP solver to determine surgeon arrival times.

Algorithm 4 *(ASB heuristic)*
Step 1. Select an OR pair with surgery sets (P, Q).
- Compute SOR(P), SOR(Q);
- Set $\theta^* \leftarrow$ SOR(P) + SOR(Q);
Step 2 (Identify the best swap). For all $i \in P$ and $j \in Q$:
- *Set* $\hat{P} \leftarrow P \cup \{j\} \setminus \{i\}$ and $\hat{Q} \leftarrow Q \cup \{i\} \setminus \{j\}$;
- Solve SOR($\hat{P}$) and SOR($\hat{Q}$) problems;

- If $\theta^* >$ SOR($\hat{P}$) + SOR($\hat{Q}$), $\theta^* \leftarrow$ SOR($\hat{P}$) + SOR($\hat{Q}$), $i^* \leftarrow i, j^* \leftarrow j$;
Step 3. If $\theta^* <$ SOR(P)+SOR(Q), swap (i^*, j^*), go to Step 2;
Step 4. Otherwise, go to Step 1 until no OR-pair can be improved.

The SOR and ASB heuristics are tested numerically. The SOR heuristic is evaluated on randomly generated single OR problem instances with 2 to 6 surgeries for weakly or heavily loaded OR with 75% or 125% of the regular OR capacity. Over 10,000 tested instances, SOR performs the best with 69% best solutions compared to 60% for SVF and 39% for HPF.

The ABS heuristic is compared with an "optimal" assignment obtained by partial enumeration with no more than 5 or 6 surgeries per OR and with optimal surgery sequencing and surgeon arrival times. 80 instances of 2-OR problem and 40 instances of 3-OR problem are generated as in Section VI. The average optimality gap of ASB is less than 1% and the worst 8.6%.

VI. NUMERICAL RESULTS

This section presents numerical results to evaluate the performance of the proposed methods. Our purpose is to investigate the value of dynamic scheduling, the value of dynamic scheduling optimization, the optimality gap of the proposed scheduling strategies, and the value of proactive decision, i.e., surgeon arrival times.

A. Experimental setting

Three dynamic assignment strategies are considered and compared.
- Strategy FCFS which assigns surgeries with the least surgeon arrival time when an OR is released;
- Strategy I which is a two-stage approximation approach with second stage cost estimated by OPLA heuristic;
- Strategy II, another two-stage approximation approach with second stage cost estimated by MPLA heuristic.

The dynamic assignment strategies are also compared with a so-called static strategy with surgery assignment and sequencing being those of the proactive solution and the solution of the lower bound problem to check the optimality gap.

Except where otherwise noted, surgeon arrival times are determined with ASB and SOR heuristics. However, we have also considered the impact of proactive decision optimization. Solutions obtained with ASB and SOR heuristics are compared with some reasonable rules for surgeon arrival time determination.

All algorithms are coded in Microsoft Visual Studio .NET 2010 and CPLEX 12.4. Experiments are conducted on Intel Core i5 PC running at 3.10 GHz with 4 GB memory. For solving the proactive problem, the initialization of the ASB heuristic is solved to within 1% optimality using Cplex MIP solver, and all linear programs of single OR problems are solved to optimality.

The proactive problem is solved with a common set Ω of K =1000 randomly generated scenarios, i.e., randomly generated surgery durations vectors. The expected criterion function (32) is replaced by the arithmetic average of their values in different scenarios, i.e., the expectation operation replaced by $1/K$. The corresponding proactive model then becomes mixed-integer linear programming models. This sample size is large enough for these optimization problems as it has been shown in [16] that 500 scenarios are able to provide a reasonable approximation with similar mean and variance of surgery durations.

Dynamic assignment strategies and the lower bound cost are evaluated by a number of independent simulation runs. A common set Ω' of K' =5000 randomly generated scenarios is used for the simulation of dynamic assignment strategies and for evaluation of the lower bound cost (12). It will be shown that, this sample size is large enough as the 95% confidence intervals of the costs under different strategies are well separated. With the use of the same scenarios, equation (12) is indeed a lower bound cost of any feasible dynamic assignment strategy.

At any event k of the dynamic assignment scheduling, a common set Ω'' of K'' =1000 scenarios of OR release times and surgery durations is used in Algorithm 1 for evaluation of the second-stage costs. For the MPLA heuristic, at each future event $n > k$, another set Ω''' of K''' =1000 scenarios of OR release times is used to evaluate the one-period look ahead cost by equations (28) and (29).

B. Problem instances

The problem instances of this paper are derived from our field observations of the OT in the Ruijin Hospital, Shanghai. The hospital has an integrated OT with 21 ORs. 9 ORs are laminar air flow rooms, and are dedicated to special surgeries, such as cardiosurgery, neurosurgery, etc. The 12 remaining ORs are identical and served for different general surgeries such as gastric surgery.

The regular OR session is 480 minutes, and the overtime cost is 50% higher than the OR idle time cost (set to 1). That is, the value of α is estimated to be 1.5. Mean surgery duration varies from 60 minutes to 300 minutes.

Surgeons in the hospital are divided as three classes: top surgeons, senior surgeons and ordinary surgeons. In the observed hospital, the partition of surgeons in the three classes is 1/4, 2/4, 1/4. The following waiting costs are considered reasonable: 2.5 for top surgeons (sum of unit OR idle and overtime cost), 1.5 for senior surgeons (equally important as OR overtime), 0.5 for ordinary surgeon (less important than OR idle).

The workload of the OT ($\rho = \sum_{i \in J} \bar{p}_i /MT$) varies during the week with Monday being the busiest day and Friday the least busy day. The total average surgery duration of Monday accounts for 125% of the regular OT capacity, whereas that of Friday accounts for 75% of the OT capacity.

In this paper, the following parameters are used in the generation of problem instances: $M \in \{3, 6, 12\}$, $T = 480$, $\alpha = 1.5$. Surgery durations are assumed to follow log-normal distribution. The mean surgery duration $\bar{p}_i$ is randomly generated with uniform distribution U(60, 300). Surgeries are

TABLE I
RESULTS OF 12 RANDOMLY SELECTED INSTANCES

M.	N	$(\eta,\rho\%)$	$cost^{FCFS}$	$cost^{I}$	$cost^{II}$	$CPU^{II}(sec.)$
3	7	(0.3,75)	119±3	119±3	116±3	0.01
3	6	(0.7,75)	302±6	292±5.6	286±5	0.01
3	11	(0.3,125)	1087±9	1076±9	1061±9	0.06
3	11	(0.7,125)	942±12	929±11	915±11	0.06
6	11	(0.3,75)	202±4	198±4	191±4	0.01
6	13	(0.7,75)	281±6	273±6	263±6	0.01
6	24	(0.3,125)	1561±13	1555±14	1527±13	0.52
6	22	(0.7,125)	1801±17	1801±17	1718±15	0.37
12	24	(0.3,75)	307±5	299±5	285±5	0.13
12	28	(0.7,75)	554±8	558±8	495±7	0.41
12	38	(0.3,125)	3095±8	3306±8	2970±7	2.43
12	38	(0.7,125)	3219±8	3286±8	3018±7	2.34

generated randomly one after another until the total average workload exceeds a target workload ρMT, i.e., until $\sum_{i \in J} \bar{p}_i \geq \rho MT$ with $\rho = 0.75$ (low workload case) or 1.25 (high workload case).

The other parameters of the surgeries are generated as follows. Surgeon waiting cost β_i is randomly generated and is equal to 2.5, 1.5, or 0.5 with probability 1/4, 2/4 and 1/4 respectively. The coefficient of variation, defined as $CV_i = \sigma_i/\bar{p}_i$, is randomly generated and is equal to 0.2 or 0.5 with probability $1-\eta$ and η respectively and with $\eta = 0.7$ (high variance case) or 0.3 (low variance case).

C. Accuracies of the simulation results

Trial runs are performed in this section to validate the simulation results by comparing the confidence intervals of the costs for different strategies.

TABLE I is the results of 12 different problem instances with different combinations of M, η, ρ. For each of the three dynamic assignment strategies (FCFS, I, II), Table I gives its average cost and the 95% confidence half-width derived from its 5000 independent simulation runs. The CPU time for each simulation run of strategy II is also given as the CPU times for other strategies are sufficiently small.

From these results, the confidence intervals are sufficiently small for faithful comparison of different strategies. Further, even for a large instance for 12-OR, the dynamic assignment strategies are able to make timely decisions in short computational time.

D. Value of dynamic scheduling

This subsection evaluates the contribution of the dynamic scheduling. We implemented the static assignment strategy for comparison purpose, which assign surgeries according to the optimal assignment and sequencing decisions of the proactive model. As a result, each surgery is assigned to the same OR for all simulation runs under static assignment strategy.

The value of dynamic scheduling is defined as follows:

TABLE II
VDS OF 140 INSTANCES

M	($\eta,\rho\%$)	VDS^I (%)			VDS^{II} (%)		
		Ave.	Min.	Max.	Ave.	Min.	Max.
3	(0.3,75)	9.7	2.6	22.7	10.6	2.6	22.9
	(0.7,75)	14.2	5.4	26.9	14.8	5.5	26.9
	(0.3,125)	6.2	1.9	14.0	7.4	3.9	14.1
	(0.7,125)	9.7	5.6	14.9	11.1	5.7	15.5
	Ave.	9.9	3.9	19.6	11.0	4.4	19.9
6	(0.3,75)	23.7	18.5	31.4	25.4	18.7	31.6
	(0.7,75)	25.3	19.1	38.3	29.2	24.7	39.9
	(0.3,125)	9.2	4.6	13.9	11.1	7.1	15.5
	(0.7,125)	15.6	10.4	20.9	19.1	12.8	24.1
	Ave.	18.5	13.2	26.1	21.2	15.8	27.8
12	(0.3,75)	30.4	27.3	33.9	33.6	30.1	37.9
	(0.7,75)	29.2	22.0	34.7	36.0	28.9	42.1
	(0.3,125)	10.4	1.8	14.7	18.6	17.2	20.4
	(0.7,125)	18.6	14.4	21.6	26.1	23.9	30.1
	Ave.	22.2	16.4	26.2	28.6	25.0	32.6

TABLE III
VOS OF 140 INSTANCES

M	($\eta,\rho\%$)	VOS^I(%)			VOS^{II}(%)		
		Ave.	Min.	Max.	Ave.	Min.	Max.
3	(0.3,75)	1.8	-2.8	14.4	2.8	0.0	14.4
	(0.7,75)	4.6	-0.7	28.8	5.4	0.0	26.5
	(0.3,125)	1.0	-0.1	4.5	2.3	0.0	7.0
	(0.7,125)	1.5	-0.8	6.4	3.1	0.0	10.2
	Ave.	2.2	-1.1	13.5	3.4	0.0	14.5
6	(0.3,75)	3.3	-0.1	12.6	5.4	-0.1	13.6
	(0.7,75)	0.9	-2.7	3.2	6.0	-0.1	11.3
	(0.3,125)	0.9	-0.8	3.7	2.9	0.0	5.0
	(0.7,125)	0.9	-1.5	3.2	5.0	0.6	8.7
	Ave.	1.5	-1.3	5.7	4.8	0.1	9.6
12	(0.3,75)	2.5	0.5	3.2	7.0	5.8	7.8
	(0.7,75)	-0.5	-2.1	2.2	9.3	6.1	11.8
	(0.3,125)	-4.5	-13.8	0.2	5.0	3.4	6.8
	(0.7,125)	-3.0	-6.7	-0.8	6.4	4.7	9.2
	Ave.	-1.4	-5.5	1.2	6.9	5.0	8.9

$$VDS^X = \frac{cost^{Static} - cost^X}{cost^{Static}} \times 100\%$$

where $cost^{Static}$ denotes the average cost under the static assignment strategy and $cost^X$ that of a dynamic strategy. From the definition, the greater the value of VDS^X, the higher the benefit of dynamic scheduling is.

We tested 140 instances with 20 instances for 3-OR, 10 instances for 6-OR and 5 instances for 12-OR for all the combinations of η and ρ. TABLE II presents the average, min and max of VDS^X for all instances of each combination of (M, η, ρ).

The following observations can be made:

- Dynamic scheduling always works as VDS is always positive.

- The benefit of dynamic scheduling is more important for a larger OT. The average VDS is 10.46% for 3-OR instances, 19.83% for 6-OR instances and 25.37% for 12-OR instances.

- Dynamic scheduling is able to efficiently cope with surgery uncertainties as VDS increases when the variance of surgery duration increases.

- VDS decreases as the workload of the OT increases.

E. Value of dynamic scheduling optimization

This subsection evaluates the importance of dynamic scheduling optimization and compares the dynamic assignment strategies with a default FCFS strategy. For this purpose, we define the value of optimized solution VOS as follows:

$$VOS^X = \frac{cost^{FCFS} - cost^X}{cost^{FCFS}} \times 100\%$$

Positive VOS^X means the strategy X outperforms FCFS strategy, whereas negative VOS^X implies the opposite.

TABLE III presents the VOS of the 140 instances for different combination of (M, η, ρ). VOS^{II} is non-negative except for two instances for which it is close to 0. This implies that strategy II outperforms FCFS. VOS^{II} is consistently better than VOS^I which implies that multi-period look ahead pays.

The value of VOS depends on problem parameters and is explained in the following.

Impact of the number M of ORs: VOS^{II} increases as M increases. Compared to FCFS, strategy II is able to dynamically revise the surgery sequencing in order to improve the solution. The benefit becomes greater for a larger OT where the opportunities of dynamically matching surgery assignment and surgery sequencing become greater.

In contrast, VOS^I decreases as M increases. VOS^I is even always negative for the case (M, η, ρ) = (12, 0.7, 1.25). The myopic nature of OPLA heuristic makes it unable to identify appropriate surgery assignment, especially for a large OT.

Impact of the variances: VOS^{II} increases as η increases, i.e., the variance of surgery durations increases. This implies that strategy II is able to efficiently cope with the uncertainties through appropriate dynamic surgery assignment. VOS^I increases as η increases for 3-OR instances but it decreases for a larger OT. The myopic nature makes it unable to take advantage of the dynamic scheduling in a highly uncertain environment.

Impact of the workload: Both VOS^I and VOS^{II} decrease as ρ increases, i.e., the workload of the OT increases. For high workload case, the highest cost is the overtime cost. As a result, potential improvement by dynamic scheduling becomes smaller.

TABLE IV
OPTIMALITY GAP OF DYNAMIC SCHEDULING

$(\eta,\rho\%)$	GAP^{I}(%)			GAP^{II}(%)		
	Ave.	Min.	Max.	Ave.	Min.	Max.
(0.3,0.75)	7.4	0.1	14.7	6.3	0.1	12.8
(0.7,0.75)	8.5	5.1	14.8	7.7	3.8	18.4
(0.3,1.25)	5.6	1.3	11.2	4.1	1.0	8.3
(0.7,1.25)	7.8	1.9	17.3	6.0	1.6	9.6

F. Optimality of the dynamic assignment strategies

This subsection assesses the degree of optimality of the dynamic assignment strategies by comparing them with the lower bound cost *LB* of Section III. More specifically, the following optimality gap is evaluated:

$$GAP^{X} = \frac{cost^{X} - LB}{LB} \times 100\%$$

For the sake of the computational complexity of the LB model, we tested 80 instances of 3-OR problem. The results are presented on TABLE IV. The average gaps are less than 8% and 9% for strategy II and strategy I, respectively. As *LB* is obtained with perfect information, this optimality gap is relatively small. Further, high surgery duration variance degrades the optimality gap while high workload reduces the optimality gap. The former is obvious as *LB* is obtained with perfect information.

G. Impact of surgeon waiting time cost

This subsection investigates the impact of surgeon waiting cost on surgeon waiting time, OR idling time and OR overtime. For this purpose, the unit surgeon waiting cost is multiplied by a parameter κ, i.e., $\beta_i \in \{2.5\kappa, 1.5\kappa, 0.5\kappa\}$. Larger value of κ implies increased importance of the surgeons with respect to the OR utilization. Similar results are observed but only results of 3-OR cases are presented hereafter.

We tested 80 instances for 3-OR problem with 20 instances for each combination of (η, ρ). TABLE V presents the results. For each instance, we determine the total expected surgeon waiting time $\bar{W}$, OR idle time $\bar{I}$ and OR overtime $\bar{O}$. Table V gives their average values over 20 instances for each (κ, η, ρ). The coefficients of variation of $\bar{W}$, $\bar{I}$ and $\bar{O}$ are less than 0.05 which implies very small confidence intervals that are not given for the sake of readability.

As κ increases, or surgeons become more important, surgeon waiting times decrease, and the OR idle times and overtimes increase, which means the OT management changes from OR-oriented to surgeon-oriented. For example, when κ changes from 0.5 to 1, with strategy II, $\bar{W}$ decreases by 33%, $\bar{I}$ increases by 105% and $\bar{O}$ increases by 6.4%. The reduction of the surgeon waiting time is achieved by planning larger surgeon inter-arrival times in the proactive problem, leading to the increase of OR idle time. The higher increase of OR idle time than that of OR overtime is due to the higher cost of OR overtime.

The wasted times of surgeons and the OT are also impacted by the uncertainty of surgery durations and the workload of the OT.

TABLE V
PERFORMANCE MEASURES VS SURGEON WAITING COST

κ	$(\eta,\rho\%)$	Strategy I			Strategy II		
		$\bar{W}$	$\bar{I}$	$\bar{O}$	$\bar{W}$	$\bar{I}$	$\bar{O}$
0.5	(0.3,75)	79.3	20.7	42.7	80.3	20.9	41.4
	(0.7,75)	105.3	22.7	64.3	107.2	24.2	63.3
	(0.3,125)	310.2	22.7	480.1	306.7	22.0	479.3
	(0.7,125)	374.0	29.6	508.5	372.4	29.5	508.5
	Ave.	217.2	23.9	273.9	216.6	24.2	273.1
1.0	(0.3,75)	50.8	36.7	46.8	51.2	36.5	46.2
	(0.7,75)	69.9	44.7	68.3	72.0	45.6	68.0
	(0.3,125)	206.7	50.1	506.3	206.2	49.5	505.7
	(0.7,125)	255.8	66.6	542.6	253.0	66.3	542.4
	Ave.	145.8	49.5	291.0	145.6	49.5	290.6
2.0	(0.3,75)	30.6	63.4	51.9	30.8	63.5	51.6
	(0.7,75)	40.8	80.6	76.6	41.5	81.0	76.4
	(0.3,125)	123.7	102.1	557.1	122.3	101.2	556.2
	(0.7,125)	154.0	134.4	608.3	152.0	134.5	608.3
	Ave.	87.3	95.1	323.5	86.7	95.0	323.1

TABLE VI
VPS OF SMALL SIZE INSTANCES

$(\eta,\rho\%)$	VPS^{I}(%)			VPS^{II}(%)		
	Ave.	Min.	Max.	Ave.	Min.	Max.
(0.3,0.75)	7.2	-15.2	23.3	7.0	-20.9	22.6
(0.7,0.75)	6.8	-11.1	20.4	6.4	-14.4	20.4
(0.3,1.25)	9.8	1.1	23.1	10.0	0.9	21.6
(0.7,1.25)	10.1	1.1	19.2	10.1	3.2	17.9

H. Value of proactive solution

This subsection evaluates the contribution of the stochastic optimization of surgeon arrival times. We introduce the value of proactive solution defined as follows:

$$VPS^{X} = \frac{\overline{cost}^{X} - cost^{X}}{\overline{cost}^{X}} \times 100\%$$

where $\overline{cost}^{X}$ denotes the average cost of the strategy *X* but with surgeon arrival times determined with mean surgery duration. To obtain the surgeon arrival times of $\overline{cost}^{X}$, as in ASB, we first determine the surgery assignment in order to best balance OR workload, and the surgeries assigned to each OR are then sequenced in decreasing order of unit surgeon waiting cost β_i. The surgeon arrival time A_i is set to be the total mean durations of surgeries scheduled before *i* in the same OR.

We tested 80 instances for 3-OR with 20 instances for each combination of (η, ρ). The results are presented on TABLE VI. *VPS* can be significant for both strategies I and II, which means that the proactive solution is very important to dynamic assignment scheduling. The negative *VPS* in some instances shows that the arrival times that optimize the proactive model may not be adjustable to the dynamic assignment scheduling.

Joint optimization of surgeon arrival times and dynamic assignment policies is an open research issue.

VII. CONCLUSION

This paper addressed the dynamic assignment problem with given surgeon arrival times. We proposed a multi-stage stochastic programming model in which each stage corresponds to a surgery completion event. As the problem is hard to solve, we proposed a two-stage stochastic programming approximation approach. The second stage cost is evaluated by two heuristics: a one-period look ahead (OPLA) method and a multi-period look ahead (MPLA) method. A lower bound is provided based on perfect information. An efficient approach is also proposed to solve the proactive problem to determine surgeon arrival times. Numerical results showed that dynamic scheduling significantly improves static surgery scheduling, and the optimization of dynamic scheduling further improves the performance. The MPLA heuristic outperforms the myopic OPLA heuristic and is able to efficiently cope with surgery duration uncertainties and take advantage of large size of the operating theater. The results also showed that the quality of the proactive solution is very important.

Future work will focus on the study of stochastic programming algorithms for the joint optimization of the proactive and reactive problems. Another research direction is to take a starting time reliability approach in order to ensure some given probability for surgeries to start at their planned time. Structural properties of the scheduling policies are highly needed for this purpose.

REFERENCES

[1] S. Erdogan and B. Denton, "Surgery planning and scheduling," Wiley Encyclopedia of Operations Research and Management Science. New York: John Wiley & Sons, 2011.

[2] M. Lamiri, X. Xie and S. Zhang, "Column generation approach to operating theater planning with elective and emergency patients," IIE Transactions, vol. 40, pp. 838-852, 2008.

[3] S. Batun, "Scheduling multiple operating rooms under uncertainty", Ph.D. Dissertation, University of Pittsburgh, United States, 2011.

[4] N. Geng, X. Xie, Z. Jiang, "Capacity Reservation and Cancellation of Critical Resources," IEEE Trans. Automation Science and Engineering, Vol. 8, No. 3, pp. 470-481, July 2011.

[5] K.W. Tan, C. Wang and H.C. Lau, "Improving patient flow in emergency department through dynamic priority queue," Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE'12), Seoul, Korea, 2012.

[6] Z. Zhang, X. Xie, N. Geng, "Promise Surgery Start Times and Implementation Strategies," Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE'12), Seoul, Korea, 2012.

[7] S. Batun, B.T. Denton, T.R. Huschka and A.J. Schaefer, "Operating Room Pooling and Parallel Surgery Processing Under Uncertainty," Informs Journal on Computing, vol. 23, pp. 220-237, Spr 2011

[8] M. Haji and H. Darabi. "A Simulation Case Study: Reducing Outpatient Waiting Time of Otolaryngology Care Services using VBA," Proceedings

of the IEEE International Conference on Automation Science and Engineering (CASE'11), Trieste, Italy, Aug. 2011.

[9] A. Sadki, X. Xie and F. Chauvin, "Appointment Scheduling of Oncology Outpatients," Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE'11), Trieste, Italy, Aug. 2011.

[10] B. Cardoen, E. Demeulemeester and J. Beliën, "Operating room planning and scheduling: A literature review," European Journal of Operational Research, vol. 201, pp. 921-932, 2010.

[11] Cayirli, T., and Veral, E, "Outpatient Scheduling in Health Care: A Review of Literature," Production and Operations Management, vol. 12(4), pp. 519-549, 2003

[12] D. Gupta & B. Denton, "Appointment scheduling in health care: Challenges and opportunities," IIE Transactions, vol. 40, pp. 800-819, 2008

[13] E.N.Weiss, "Models for determining the estimated start times and case orderings," IIE Transactions, vol. 22, pp. 143-150, 1990.

[14] B. Denton and D. Gupta, "A Sequential Bounding Approach for Optimal Appointment Scheduling," IIE Transactions, vol. 35(11), pp. 1003-1016, 2003

[15] B. Denton, J. Viapiano and A. Vogl, "Optimization of surgery sequencing and scheduling decisions under uncertainty," Health Care Manag Sci, vol. 10, pp. 13-24, 2007.

[16] C. Mancilla and R. Storer, "A sample average approximation approach to stochastic appointment sequencing and scheduling," IIE Transactions, vol. 44(8), pp. 655-670, 2012

[17] Q. Kong, C. Lee, C. Teo and Z. Zheng. "Appointment Sequencing: Moving Beyond The Smallest-Variance-First Rule", working paper, National University of Singapore, 2012

[18] S. Erdogan and B. Denton, "Dynamic Appointment Scheduling of a Stochastic Server with Uncertain Demand," INFORMS Journal on Computing, pp. 1-17, 2011

[19] S. Erdogan, A. Gose and B. Denton, "On-line Appointment Sequencing and Scheduling," working paper, Stanford University, 2012

[20] Begen, M.A., and Queyranne,M., "Appointment scheduling with discrete random durations," Mathematics of Operations Research, vol. 36(2), pp. 240-257, 2011

[21] Begen, M.A., Levi, R. and Queyranne,M., "Technical Note—A Sampling-Based Approach to Appointment Scheduling," Operations Research, vol. 60, No. 3, pp. 675–681, May–June 2012.

[22] S. Gul, B. Denton, J. Fowler and T. Huschka, "Bi-Criteria Scheduling of Surgical Services for an Outpatient Procedure Center," Production and Operations Management, vol. 20(3), pp. 406-417, 2011

[23] Q. Kong, C. Lee, C. Teo and Z. Zheng, "Scheduling Arrivals to a Stochastic Service Delivery System using Copositive Cones," working paper, National Universtiy of Singapore, 2012

[24] B. Denton, A. Miller, H. Balasubramanian, and T. Huschka, "Optimal allocation of surgery blocks to operating rooms under uncertainty," Operations Research, vol. 58 (4 PART 1), pp. 802-816, 2010.

[25] J.R. Birge, & F. Louveaux, "Introduction to stochastic programming". Springer, New York, 1997.
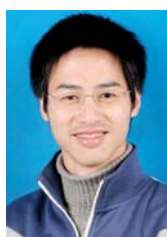
Zheng Zhang received the B.S. degree in Industrial Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2010.

Currently, he is a Ph.D. student in the Department of Industrial Engineering and Logistics Management, Shanghai Jiao Tong University, Shanghai, China. His research interests include appointment scheduling of healthcare systems.

Xiaolan Xie (SM'10) received his Ph.D degree from the University of Nancy I, Nancy, France, in 1989, and the Habilitation à Diriger des Recherches degree from the University of Metz, France, in 1995.

Currently, he is a distinguished professor of industrial engineering, the head of the department of Healthcare Engineering of the Center for Health Engineering and the head of IEOR team of CNRS UMR 6158 LIMOS, Ecole Nationale Supérieure des Mines (ENSMSE), Saint Etienne, France. He is also a chair professor and director of the Center for healthcare engineering at the Shanghai Jiao Tong University, China. Before Joining ENSMSE, he was a

Research Director at the Institut National de Recherche en Informatique et en Automatique (INRIA) from 2002 to 2005, a Full Professor at Ecole Nationale d'Ingénieurs de Metz from 1999 to 2002, and a Senior Research Scientist at INRIA from 1990 to 1999. His research interests include design, planning and scheduling, supply chain optimization, and performance evaluation, of healthcare and manufacturing systems. He is author/coauthor of over 250 publications including over 90 journal articles and five books. He has rich industrial application experiences with European industries. He is PI for various national and international projects including ANR-TECSAN HOST on management of winter epidemics, NSF China key project on planning and optimization of health care resources, French Labex IMOBS3 project on home health cares, FP6-IST6 IWARD on swarm robots for health services, FP6-NoE I*PROMS on intelligent machines and production systems, the FP5-GROWTH-ONE project for the strategic design of supply chain networks, the FP5- GRWOTH thematic network TNEE on extended enterprises.

He has been an associate editor for IEEE Transactions on Automation Science & Engineering, IEEE Transaction on Automatic Control, IEEE Transactions on Robotics & Automation and has been on the (conference) editorial board of the IEEE Robotics and Automation Society, Health Systems, and Int. J. Simulation & Process Modelling. He has a Guest Editor of various special issues on healthcare engineering and manufacturing systems. He is general chair of ORAHS'2007, IPC chair of the IEEE Workshop on Health Care Management WHCM'2010 and IPC member for many other conferences.

Na Geng (M'12) received her Ph.D. in Industrial Engineering from both Ecole Nationale Superieure des Mines de Saint-Etienne (ENSM.SE), France and Shanghai Jiao Tong University (SJTU), China, in 2010. Currently, she is an assistant professor in the Department of Industrial Engineering and Logistics Management, SJTU, Shanghai, China. Her research interests include capacity planning and allocation, advance scheduling, and surgery planning.