INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Surgical Patient Routing Problem: A Central Planner Approach

Sepehr Nemati, Oleg V. Shylo, Oleg A. Prokopyev, Andrew J. Schaefer

INFORMS Journal on Computing
Vol. 28, No. 4, Fall 2016, pp. 657–673
ISSN 1091-9856 (print) | ISSN 1526-5528 (online)

http://dx.doi.org/10.1287/ijoc.2016.0706

The Surgical Patient Routing Problem: A Central Planner Approach

Sepehr Nemati
Ivey Business School, University of Western Ontario, Ontario N6G 0N1, Canada, sproon@ivey.uwo.ca

Oleg V. Shylo
Department of Industrial and Systems Engineering, University of Tennessee, Knoxville, Tennessee 37996, oshylo@utk.edu

Oleg A. Prokopyev
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, prokopyev@engr.pitt.edu

Andrew J. Schaefer
Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, andrew.schaefer@rice.edu

Many patients face difficulties when accessing medical facilities, particularly in rural areas. To alleviate these concerns, medical centers may offer transportation to eligible patients. However, the operation of such services is typically not tightly coordinated with the scheduling of medical appointments. Motivated by our collaborations with the U.S. Veterans Health Administration, we propose an integrated approach that simultaneously considers patient routing and operating room scheduling decisions. We model this problem as a mixed-integer program. Unfortunately, realistically sized instances of this problem are intractable, so we focus on a special case of the problem that captures the needs of low-volume (e.g., rural) hospitals. We establish structural properties that are exploited to develop a branch-and-price algorithm, which greatly outperforms a commercial solver on the original formulation. We discuss several algorithmic strategies to improve the overall solution efficiency. We evaluate the performance of the proposed approach through an extensive computational study calibrated with clinical data. Our results demonstrate that there exist opportunities for healthcare providers to significantly improve the quality of their services by integrating scheduling and routing decisions.

Keywords: operation room scheduling; outpatient elective surgeries; mixed-integer programming; branch and price; vehicle routing
History: Accepted by Allen Holder, Area Editor for Applications in Biology, Medicine, and Healthcare; received November 2013; revised December 2014, August 2015, February 2016; accepted March 2016. Published online August 30, 2016.

1. Introduction

Operating room (OR) scheduling is a complex process that depends on many factors, including resource availability (Erdogan and Denton 2011). This is an active research area (Gupta and Denton 2008; Dexter et al. 1999, Dexter et al. 2003), and various issues arising in operating room management have been extensively studied (see, e.g., Blake and Carter 1997, Gupta 2007, Gupta and Denton 2008, Cardoen et al. 2010, Shylo et al. 2013). We refer the reader to the comprehensive operating room scheduling bibliography maintained by Dexter (2016).

One of the major problems arising in this context is due to patients having difficulties accessing healthcare, which is of particular importance for rural patients or those who require special medical care (Rask et al. 1994, Okoro et al. 2005, Fitzpatrick et al. 2004). Recognizing the growing need to improve access to medical care, several organizations in the United States offer transportation services for patients. A prime example is the Beneficiary Travel program offered by the U.S. Veterans Health Administration (VHA), which also includes the program on Automobiles and Adaptive Equipment for Certain Disabled Veterans and Members of the Armed Forces. The combined budget of these two programs was $820 million in FY 2010 alone (Burkhardt et al. 2011) and this number is projected to increase further.

There exist multiple alternatives for patients traveling to VHA facilities. For example, some hospitals directly operate their own fleet of transportation vehicles, while others use either contracted vendors or nonprofit service organizations, e.g., the Disabled American Veterans (DAV). We refer to a detailed survey by Burkhardt et al. (2011) that describes a number of related examples in place throughout the United States.

When a medical center provides transportation services, the OR scheduling problem becomes even more complex. Although the volume of such services is increasing, their operation has attracted relatively little attention in the literature (Burkhardt et al. 2011, Burkhardt and Garrity 2012). While the academic literature and healthcare practice contains a number of successful examples that involve applications of methodologically sound methods to OR scheduling, see, e.g., Blake and Donald (2002), Day et al. (2012), Denton (2009), van Oostrum et al. (2009), planners typically ignore transportation considerations. As (Burkhardt et al. 2011, p. 65) noted, the decision makers "who schedule medical appointments do not necessarily perceive transportation problems when they set up appointments." In this paper, motivated by our collaborations with a large VHA hospital, we propose to address this problem by developing an approach that simultaneously considers patient scheduling and vehicle routing decisions.

To the best of our knowledge no prior work integrates these two problems into a single framework. We call this problem the surgical patient routing problem (SPRP). Specifically, we consider the problem of optimizing scheduling and transportation decisions given a set of *outpatient* surgery requests that do not require an overnight stay at the hospital, i.e., patients arrive on a day of surgery and leave after the completion of the surgical and postoperative procedures. Furthermore, in our work we focus on *elective* surgeries, i.e., deferrable surgical procedures, which are planned and scheduled well in advance. An elective surgery may be either medically required, e.g., cataract surgery, or optional, e.g., cosmetic surgery. Finally, in the current work we assume an *open-booking* scheduling framework, where surgical resources are shared among specialty teams, individual surgeons, or surgical departments (Gupta 2007).

The planning horizon in SPRP consists of a finite number of time stages, which might span multiple days. This model integrates the decisions in a typical three-step process: (1) assign surgeries to available time stages, (2) schedule surgeries across ORs at each time stage, and (3) determine the transportation plan using a fleet of available vehicles. The objective is to minimize the total service cost of the patients, defined as the weighted sum of patients' total travel time and the time spent at the hospital. Since long travel and waiting times cause dissatisfaction among patients (Gupta and Denton 2008), this cost captures both goals. While maximizing patient satisfaction is in line with the VHA goals, our model can consider other objectives as well.

In our paper we make two relatively mild assumptions. First, we consider a batch scheduling process, where the scheduler is given a set (batch) of surgery requests that need to be allocated into a set of available ORs. This assumption implies that there exists a certain delay in assignment of surgeries to specific ORs and time stages. Subsequently, the base model can be used within a rolling-horizon scheduling framework, where it is resolved regularly as soon as a sufficiently large batch of new surgery requests is accommodated. Such an assumption is not restrictive and practical as we focus on elective surgeries. The batch scheduling approach is typical in the related OR scheduling literature; see, e.g., Batun et al. (2011), Hans et al. (2008), Min and Yih (2010), and Shylo et al. (2013). We also refer to a survey by Gupta and Denton (2008), where it is pointed out that the batch-based approach is common for scheduling elective surgeries, which are the main focus of this paper.

Our second assumption is that both scheduling and routing decisions are performed by the same scheduler, i.e., a *central planner*. Therefore, the approach proposed in our paper can be suitable for hospitals that operate their own fleets of transportation vehicles. VHA hospitals are, perhaps, prime examples of such healthcare facilities; recall that our collaboration with one such hospital served as the main motivation of this work.

The models presented in this paper integrate the surgical patient scheduling problem with the patient routing problem. There are some studies in the literature that consider home-healthcare scheduling, where the assignment of nurses or medical equipment to patients, who are visited outside of the hospitals, is integrated with the routing decisions; see, e.g., Begur et al. (1997) and Trautsamwieser and Hirsch (2014). However, we should note the models in these studies are mainly focused on staff or medical equipment routing for serving patients outside of the hospitals. They do not address the scheduling and routing considerations when the patients have to be treated at the hospitals. Furthermore, there is a vast body of literature on the joint scheduling and routing models focused on production scheduling and distribution applications. Sarmiento and Nagi (1999) and Chen (2010) provide a comprehensive review of such models in the operations management domain. However, these models typically have a different structure or objective that limits their applicability in the health service delivery domain.

Our model generalizes the delivery-man problem or traveling repairman problem (TRP) and the K-traveling repairman problem (KTRP), which are known to be *NP*-hard (Lucena 1990, Fischetti et al. 1993, Heilporn et al. 2010, Ngueveu et al. 2010, Ribeiro and Laporte 2012). In the TRP, the objective is to minimize the total arrival times to customers' locations by a single vehicle, rather than minimizing the length of the tour. The KTRP is a generalization of the TRP for multiple vehicles, where the routing decisions are not limited to a single vehicle, but involve a homogenous fleet of vehicles. The KTRP is a special case of SPRP in which all the surgery durations are identical. Thus, SPRP is also *NP*-hard. To the best of our knowledge, there is no exact algorithm for the KTRP in the literature. We also refer the reader to Park and Kim (2010), Laporte (1992), Solomon (1987), Ribeiro and Soumis (1994), Dumas et al. (1991), and Toth and Vigo (2002) for comprehensive reviews of the models and solution approaches for the vehicle routing problem (VRP) and its variations.

The contributions of this paper are as follows:

• We propose a novel class of mathematical programming problems that integrate operating room scheduling and transportation planning. Specifically, we model the simultaneous optimization of surgery scheduling and vehicle routing decisions (see Section 2.1).

• As even small instances of SPRP appear to be intractable, we focus on a special case of SPRP, the batch surgical patient routing problem (BSPRP), that captures the needs of low-volume (e.g., rural) hospitals with a relatively small number of surgeries at each time stage (see Section 2.2). This allows us to establish structural properties of the problem that are exploited in Section 3 to develop a branch-and-price algorithm to a set-partitioning reformulation of BSPRP. The performance of the developed method is further enhanced through several algorithmic strategies.

• We demonstrate the effectiveness of our solution approach and the value of integrating surgery scheduling with transportation decisions through an extensive computational study using clinical data (see Section 4). Admittedly, our models are somewhat stylized and simplified versions of the decisions faced in practice. However, they capture the main trade-offs with respect to the interactions between the patient routing and scheduling decisions. More importantly, our computational results demonstrate that there exist opportunities for healthcare providers to substantially improve the quality of their services by considering integrated approaches.

The remainder of this paper is organized as follows. In Section 2 we present mixed-integer programming (MIP) formulations for the SPRP and BSPRP and discuss some structural properties of the BSPRP. In Section 3 we exploit these structural properties and provide a set-partitioning formulation of BSPRP. We further present a branch-and-price algorithm to solve the set-partitioning formulation and discuss several algorithmic strategies that enhance the performance of the proposed algorithm. Extensive computational experiments using VHA data are presented in Section 4 to evaluate the effectiveness of our approach and to estimate the value of integrating surgery scheduling with routing decisions. We summarize general insights of our analysis and conclude in Section 5.

2. Mathematical Programming Formulations

First, we introduce the general SPRP model, which simultaneously considers transportation and surgery scheduling decisions for a set of elective outpatient surgery requests. We assume a finite planning horizon, a set of available ORs, a set of surgery requests, and a fleet of homogeneous transportation vehicles. However, our proposed approach is general enough to incorporate a fleet of heterogeneous vehicles at each time stage.

2.1. General Surgical Patient Routing Problem

The planning horizon in the SPRP model consists of a fixed number of time stages, which may differ in duration. Figure 1 illustrates a simple example that consists of two days and four stages. At each stage (e.g., a half day), round-trip shared-type rides using a fleet of identical vehicles are provided for the patients.

The objective function is to minimize the total service time cost of the patients. As mentioned earlier, each patient's service cost is defined as the weighted sum of his/her total travel cost (home-to-hospital and hospital-to-home) and the cost associated with his/her time spent at the hospital. The home-to-hospital and hospital-to-home routes are referred to as the *pick-up* and the *drop-off* routes, respectively. Specifically, each instance of the SPRP is associated with the following set of parameters:

• $\mathcal{N}$: a set of geographically dispersed patients, where $|\mathcal{N}| = n$;

• $\mathcal{K}$: a set of time stages, where $|\mathcal{K}| = K$;

• $\mathcal{B}$: a set of all available ORs in the hospital;

• $\mathcal{B}^k \subseteq \mathcal{B}$: a set of ORs available at time stage $k \in \mathcal{K}$;

• $\mathcal{Q}^k$: a set of available vehicles at stage $k \in \mathcal{K}$;

• L_b^k: the session length of OR $b \in \mathcal{B}^k$ at stage $k \in \mathcal{K}$;

• l_{ij}: the travel time between patients $i, j \in \mathcal{N}$;

• d_i: the surgery duration of patient $i \in \mathcal{N}$ (including pre- and post-incision periods);

• κ: the capacity of each vehicle;

• τ_b^k: the opening time of OR $b \in \mathcal{B}^k$ at stage $k \in \mathcal{K}$ (we assume that $\min_{b,k}\{\tau_b^k\} = 0$);

• c^r: the travel time cost, per patient per unit of travel time;

• c^h: the hospital time cost, per patient per unit of time spent at the hospital.

We define the following set of decision variables:

• $x_{ij}^{qk} \in \{0, 1\}$: 1 if vehicle q at stage k picks up patient i immediately before patient j, 0 otherwise;

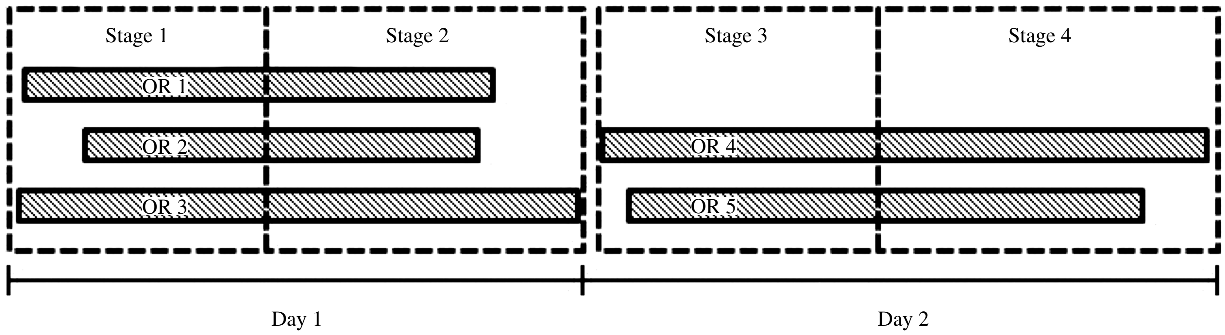

Figure 1 Planning Horizon Consisting of Two Days and Four Stages

- $\bar{x}_{ij}^{qk} \in \{0,1\}$: 1 if vehicle q at stage k drops off patient i immediately before patient j, and 0 otherwise;
- $z_{ib}^{k} \in \{0,1\}$: 1 if patient i's surgery is scheduled in OR b in stage k, and 0 otherwise;
- $u_i^{qk} \in \{0,1\}$: 1 if vehicle q picks up patient i assigned to stage k, and 0 otherwise;
- $\bar{u}_i^{qk} \in \{0,1\}$: 1 if vehicle q drops off patient i assigned to stage k, and 0 otherwise;
- $\eta_{ijb}^{k} \in \{0,1\}$: 1 if patients i and j are both assigned to OR b in stage k, and the surgery for patient i immediately precedes the surgery of patient j, and 0 otherwise;
- t_i^{qk}: pick-up time of patient i by vehicle q assigned to stage k;
- $\bar{t}_i^{qk}$: drop-off time of patient i by vehicle q assigned to stage k;
- s_{ib}^{k}: start time of patient i's surgery in OR b at stage k.

Note that t_i^{qk} is defined with respect to the earliest OR's opening time (which is assumed to be time 0); thus, it can possibly take negative values if patients arrive before the earliest OR's opening time.

Let $G=(\tilde{\mathcal{N}},\mathcal{A})$ be a directed graph, where $\tilde{\mathcal{N}}$ and $\mathcal{A}$ are its node and arc sets, respectively. Each node in $\tilde{\mathcal{N}}=\mathcal{N}\cup\{0,n+1\}$ either corresponds to the location of patient $i\in\mathcal{N}$ or the hospital, i.e., $\{0,n+1\}$. Let b_0 be a dummy OR assigned to node 0. The arc set $\mathcal{A}$ defines possible routes between patients' locations. For simplicity of exposition, we assume that $l_{ij}=l_{ji}$ for all $i,j\in\mathcal{N}$, $i\neq j$, and $l_{0i}=l_{i(n+1)}$ for all $i\in\mathcal{N}$. Note that for each stage $k\in\mathcal{K}$, variables $t_{(n+1)}^{qk}$ and $\bar{t}_0^{qk}$ define the arrival and departure times of vehicle $q\in\mathcal{Q}^k$ to and from the hospital, respectively. Without loss of generality, we do not consider multiple patient pick-ups or drop-offs; i.e., there is exactly one patient for each node in $\mathcal{N}$. Furthermore, $\mathcal{A}$ does not contain arcs of the type $(n+1,i)$ and $(i,0)$ for all $i\in\tilde{\mathcal{N}}$ as well as $(0,n+1)$.

The following set of constraints determines the pick-up schedule of the patients at each time stage while simultaneously considering vehicle capacity restrictions and forbidding overtime in ORs:

$$\sum_{b\in\mathcal{B}}\sum_{k\in\mathcal{K}} z_{ib}^{k}=1, \quad \forall i\in\mathcal{N}, \tag{1}$$

$$\sum_{q\in\mathcal{Q}^k} u_i^{qk}=\sum_{b\in\mathcal{B}^k} z_{ib}^{k}, \quad \forall i\in\mathcal{N}, \forall k\in\mathcal{K}, \tag{2}$$

$$\sum_{i\in\mathcal{N}} u_i^{qk}\le \kappa, \quad \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{3}$$

$$\sum_{j\in\mathcal{N}} x_{0j}^{qk}\le 1, \quad \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{4}$$

$$\sum_{j\in\mathcal{N}} x_{j(n+1)}^{qk}\le 1, \quad \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{5}$$

$$\sum_{j\in\tilde{\mathcal{N}}\setminus\{0\}} x_{ij}^{qk}=\sum_{j\in\tilde{\mathcal{N}}\setminus\{n+1\}} x_{ji}^{qk}, \quad \forall i\in\mathcal{N}, \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{6}$$

$$\sum_{j\in\tilde{\mathcal{N}}\setminus\{0\}} x_{ij}^{qk}=u_i^{qk}, \quad \forall i\in\mathcal{N}, \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{7}$$

$$t_i^{qk}+l_{ij}-M(1-x_{ij}^{qk})\le t_j^{qk}, \quad \forall i,j\in\tilde{\mathcal{N}}, \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{8}$$

$$-Mu_i^{qk}\le t_i^{qk}\le Mu_i^{qk}, \quad \forall i\in\tilde{\mathcal{N}}, \forall q\in\mathcal{Q}^k, \forall k\in\mathcal{K}, \tag{9}$$

$$\sum_{i\in\mathcal{N}} d_i z_{ib}^{k}\le L_b^k, \quad \forall b\in\mathcal{B}^k, \forall k\in\mathcal{K}. \tag{10}$$

Constraints (1) ensure that each patient is assigned to exactly one OR. Constraints (2) guarantee that each patient is picked up by a single vehicle, and only the vehicles available at the corresponding time stage are used for transportation. Furthermore, vehicle capacity is enforced by constraints (3). The pick-up routes should satisfy certain properties: constraints (4) and (5) ensure that each vehicle leaves and arrives to the hospital at most once, constraints (6) model the flow conservation for the pick-up routes, while constraints (7) guarantee that vehicles visit each patient's location in the pick-up routes at most once. The specific pick-up times are defined in constraints (8) and (9), where the positive constant M is large enough (e.g., $M\ge n(\max_{i,j\in\mathcal{N}} l_{ij}+\max_{i\in\mathcal{N}} d_i)$). Note that the former also act as sub-tour elimination constraints (Miller et al. 1960). Finally, overtimes in ORs are forbidden by (10).

The following constraints determine the sequence of surgeries in the ORs:

$$\tau_b^k - M(1-z_{ib}^k) \le s_{ib}^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (11)$$

$$t_{n+1}^{qk} - M(1-u_i^{qk}) \le s_{ib}^k, \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (12)$$

$$s_{ib}^k + d_i - M(1-\eta_{ijb}^k) \le s_{jb}^k, \quad \forall i,j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (13)$$

$$\eta_{ijb}^k + \eta_{jib}^k \le z_{ib}^k, \quad \forall i,j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (14)$$

$$\eta_{ijb}^k + \eta_{jib}^k \le z_{jb}^k, \quad \forall i,j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (15)$$

$$\eta_{ijb}^k + \eta_{jib}^k \ge z_{ib}^k + z_{jb}^k - 1, \quad \forall i,j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k. \quad (16)$$

Constraints (11) and (12) specify that the surgeries start after the OR start times and the arrival of vehicles to the hospital. Constraints (13)–(16) sequence the surgeries in the ORs and determine their start times.

Finally, the following set of constraints determines the drop-off schedule of the patients after the completion of their surgeries at each time stage:

$$\sum_{q \in \mathcal{Q}^k} \bar{u}_i^{qk} = \sum_{b \in B^k} z_{ib}^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \quad (17)$$

$$\sum_{i \in \mathcal{N}} \bar{u}_i^{qk} \le \kappa, \quad \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (18)$$

$$\sum_{j \in \mathcal{N}} \bar{x}_{0j}^{qk} \le 1, \quad \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (19)$$

$$\sum_{j \in \mathcal{N}} \bar{x}_{j(n+1)}^{qk} \le 1, \quad \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (20)$$

$$\sum_{j \in \tilde{\mathcal{N}} \setminus \{0\}} \bar{x}_{ij}^{qk} = \sum_{j \in \tilde{\mathcal{N}} \setminus \{n+1\}} \bar{x}_{ji}^{qk}, \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (21)$$

$$\sum_{j \in \tilde{\mathcal{N}} \setminus \{0\}} \bar{x}_{ij}^{qk} = \bar{u}_i^{qk}, \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (22)$$

$$\bar{t}_i^{qk} + l_{ij} - M(1-\bar{x}_{ij}^{qk}) \le \bar{t}_j^{qk}, \quad \forall i,j \in \tilde{\mathcal{N}}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (23)$$

$$-M\bar{u}_i^{qk} \le \bar{t}_i^{qk} \le M\bar{u}_i^{qk}, \quad \forall i \in \tilde{\mathcal{N}}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \quad (24)$$

$$\bar{t}_0^{qk} \ge s_{ib}^k + d_i - M(1-\bar{u}_i^{qk}), \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (25)$$

$$t_{(n+1)}^{qk} \le \bar{t}_0^{qk}, \quad \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}. \quad (26)$$

Constraints (17)–(24) are similar to those defining the pick-up routes given above. Observe that in (17) we assume that the vehicles used to drop off the patients at each stage are the same as those used to determine patients' pick-up schedule. Constraints (25) and (26) ensure that each vehicle leaves the hospital once the surgeries of all patients assigned to the vehicle are completed and a vehicle cannot depart from the hospital before it arrives to the hospital.

Note that in our model formulation of SPRP we assume that $\mathcal{Q}^k \cap \mathcal{Q}^{k'} = \varnothing$; i.e., vehicles at each time stage are distinct. Thus, there is no need for constraints linking the vehicles' arrival and departure times across the time stages. If it is not the case, then for each vehicle in $\mathcal{Q}^k \cap \mathcal{Q}^{k+1}$ we need extra constraints ensuring that the vehicle drops off all the patients from stage k before it starts picking up patients in stage $k+1$. Alternatively, our formulation remains valid without such constraints as long as there exists sufficient time between consecutive stages. In the remainder of the paper we assume that one of these simplifying conditions holds.

The objective function is the weighted sum of the patients' total travel time and the time spent at the hospital (or, simply, hospital time) denoted by Φ^r and Φ^h, respectively:

$$\Phi^r := \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}^k} \{u_i^{qk}(t_{(n+1)}^{qk} - t_i^{qk}) + \bar{u}_i^{qk}(\bar{t}_i^{qk} - \bar{t}_0^{qk})\}, \quad (27)$$

$$\Phi^h := \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{q \in \mathcal{Q}^k} \{\bar{u}_i^{qk} \bar{t}_0^{qk} - u_i^{qk} t_{(n+1)}^{qk}\}, \quad (28)$$

which results in the following MIP formulation:

[SPRP]

$$\min_{\mathcal{X}} \; \{c^r \Phi^r + c^h \Phi^h\}$$

subject to (1)–(25),

$$x_{ij}^{qk}, \bar{x}_{ij}^{qk} \in \{0,1\}, \quad \forall i,j \in \tilde{\mathcal{N}}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (29)$$

$$z_{ib}^k, u_i^{qk}, \bar{u}_i^{qk}, \eta_{ijb}^k \in \{0,1\}, \quad \forall i,j \in \mathcal{N}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (30)$$

$$s_{ib}^k \ge 0, \quad \forall i \in \tilde{\mathcal{N}}, \forall q \in \mathcal{Q}^k, \forall k \in \mathcal{K}, \forall b \in \mathcal{B}^k, \quad (31)$$

where $\mathcal{X}$ defines a joint vector of decision variables, i.e., $\mathcal{X} = (x, \bar{x}, z, u, \bar{u}, \eta, s, t, \bar{t})$. Note that the nonlinear terms in the objective of SPRP can be easily linearized. However, we observe that a large number of decision variables and constraints in the obtained MIP formulation makes its solution rather challenging. For example, even a special case of SPRP described in the next section and referred to as BSPRP is not solvable for reasonably sized instances by off-the-shelf MIP solvers (see Tables 1 and 2 along with the respective discussion in Section 4).

2.2. Batch Surgical Patient Routing Problem (BSPRP)

In the considered special case of SPRP referred to as BSPRP, we make the following two assumptions:

A1. There is exactly one vehicle available at each stage; i.e., $|\mathcal{Q}^k| = 1$ for all $k \in \mathcal{K}$. Thus, each patient arrives at and leaves from the hospital using the same vehicle.

A2. Instead of considering a separate time limit constraint for each OR as in constraint (10) in the SPRP formulation, we only require that the total length of surgeries scheduled for each time stage $k \in \mathcal{K}$ should not exceed $L^k = \sum_{b\in\mathcal{B}^k} L_b^k$.

Assumption **A1** captures the needs of low-volume hospitals, where a rather small number of surgeries is usually scheduled at each time stage. Thus, a single vehicle for each time stage is typically sufficient to satisfy all transportation requirements with respect to the total available OR time in the hospital.

Assumption **A2** implies that BSPRP does not require the assignment (and sequencing) of each patient's surgery to a specific OR, see constraints (11)–(16), which could be performed by a separate (optimization) procedure. Admittedly, the optimal solution of BSPRP may violate constraints (10). However, we assume that this either does not occur (which should be the case in most scenarios as long as the OR utilization is not too high) or the issue can be handled by the hospital management separately on a case-by-case basis introducing overtime in particular ORs. Note that as a result of these assumptions, we assume that the hospital time for each patient scheduled in stage k in the objective function of BSPRP is equal to the total surgery time for all the patients scheduled in the same stage.

While Assumptions **A1** and **A2** substantially simplify the original model, BSPRP remains *NP*-hard as it generalizes TRP. Moreover, the problem is still intractable using state-of-the-art commercial MIP solvers (see Section 4). However, under **A1** and **A2** we establish some structural properties that we subsequently exploit to develop an efficient branch-and-price solution approach (see Sections 3 and 4).

The BSPRP formulation uses the following notation:

- $x_{ij}^k \in \{0,1\}$: 1 if at stage k patient i is picked up immediately before patient j, and 0 otherwise;
- $\bar{x}_{ij}^k \in \{0,1\}$: 1 if at stage k patient i is dropped off immediately before patient j, and 0 otherwise;
- $z_i^k \in \{0,1\}$: 1 if patient i is assigned to stage k, and 0 otherwise;
- t_i^k: pick-up time of patient i if assigned to time stage k;
- $\bar{t}_i^k$: drop-off time of patient i if assigned to time stage k;
- t_i: pick-up time of patient i;
- $\bar{t}_i$: drop-off time of patient i,

where time variables are defined with respect to 0 (the earliest possible departure time of a vehicle for the patients' pick-up). A feasible pick-up/drop-off route of the vehicle at each time stage should only visit each patient at most once, and the sum of surgery durations of all visited patients should not exceed the session length of the stage. This is enforced by the following constraints, which are simplified versions of (1)–(10):

$$\sum_{j\in\mathcal{N}} x_{0j}^k \le 1, \quad \forall k \in \mathcal{K}, \tag{32}$$

$$\sum_{j\in\mathcal{N}} x_{j(n+1)}^k \le 1, \quad \forall k \in \mathcal{K}, \tag{33}$$

$$\sum_{j\in\tilde{\mathcal{N}}\setminus\{0\}} x_{ij}^k = \sum_{j\in\tilde{\mathcal{N}}\setminus\{n+1\}} x_{ji}^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{34}$$

$$t_i^k + l_{ij} - M(1 - x_{ij}^k) \le t_j^k, \quad \forall i, j \in \tilde{\mathcal{N}}, \forall k \in \mathcal{K}, \tag{35}$$

$$\sum_{i\in\tilde{\mathcal{N}}\setminus\{n+1\}} x_{ij}^k = z_j^k, \quad \forall j \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{36}$$

$$\sum_{k\in\mathcal{K}} z_j^k = 1, \quad \forall j \in \mathcal{N}, \tag{37}$$

$$\sum_{i\in\mathcal{N}} d_i z_i^k \le L^k, \quad \forall k \in \mathcal{K}, \tag{38}$$

recalling that $L^k = \sum_{b\in\mathcal{B}^k} L_b^k$. Similarly, using Assumptions **A1** and **A2** and simplifying the corresponding constraints from Section 2.1, we specify the drop-off routes:

$$\sum_{j\in\mathcal{N}} \bar{x}_{0j}^k \le 1, \quad \forall k \in \mathcal{K}, \tag{39}$$

$$\sum_{j\in\mathcal{N}} \bar{x}_{j(n+1)}^k \le 1, \quad \forall k \in \mathcal{K}, \tag{40}$$

$$\sum_{j\in\tilde{\mathcal{N}}\setminus\{0\}} \bar{x}_{ij}^k = \sum_{j\in\tilde{\mathcal{N}}\setminus\{n+1\}} \bar{x}_{ji}^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{41}$$

$$\sum_{i\in\tilde{\mathcal{N}}\setminus\{n+1\}} \bar{x}_{ij}^k = z_j^k, \quad \forall j \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{42}$$

$$\bar{t}_i^k + l_{ij} - M(1 - \bar{x}_{ij}^k) \le \bar{t}_j^k, \quad \forall i, j \in \tilde{\mathcal{N}}, \forall k \in \mathcal{K}, \tag{43}$$

$$t_{(n+1)}^k + \sum_{i\in\mathcal{N}} d_i z_i^k \le \bar{t}_0^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{44}$$

and the patients' pick-up and drop-off times:

$$t_i - M(1 - z_i^k) \le t_i^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \tag{45}$$

$$\bar{t}_i + M(1 - z_i^k) \ge \bar{t}_i^k, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}. \tag{46}$$

The objective function is determined using the patients' travel and hospital times given by Φ^r and Φ^h, respectively:

$$\Phi^r = \sum_{i\in\mathcal{N}} \left\{ \bar{t}_i - t_i - \sum_{k\in\mathcal{K}} (\bar{t}_0^k - t_{(n+1)}^k) z_i^k \right\}, \tag{47}$$

$$\Phi^h = \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{K}} \{ (\bar{t}_0^k - t_{(n+1)}^k) z_i^k \}, \tag{48}$$

which are simplified versions of (27)–(28). Using these constraints, model BSPRP is given by

[BSPRP]

$$\min_{x,\bar{x},z,t,\bar{t}} \quad \{c^r\Phi^r + c^h\Phi^h\}$$

subject to (32)–(46),

$$x_{ij}^k, \bar{x}_{ij}^k, z_i^k \in \{0,1\}, \quad \forall\, i,j \in \tilde{\mathcal{N}}, \forall\, k \in \mathcal{K}. \quad (49)$$

$$t_i^k, \bar{t}_i^k, t_i, \bar{t}_i \geq 0, \quad \forall\, i \in \tilde{\mathcal{N}}, \forall\, k \in \mathcal{K}. \quad (50)$$

Proposition 1 demonstrates that there always exists an optimal solution to BSPRP such that the pick-up and the drop-off routes are in the reverse order of each other. Formally,

Proposition 1. *There exists an optimal solution $\mathcal{X}^* = (x,\bar{x},z,t,\bar{t})$ for BSPRP such that if $x_{ij}^{*k}=1$, then $\bar{x}_{ji}^{*k}=1$ for all $i,j\in\tilde{\mathcal{N}}$ and $k\in\mathcal{K}$.*

Proof. For $k\in\mathcal{K}$, let $\mathcal{N}_k=\{u\in\mathcal{N}\mid z_u^k=1\}$. Let $I_{\mathcal{N}_k}=\langle 0,i_1,i_2,\ldots,i_{|\mathcal{N}_k|},n+1\rangle$ and $J_{\mathcal{N}_k}=\langle 0,j_1,j_2,\ldots,j_{|\mathcal{N}_k|},n+1\rangle$ be any arbitrary pick-up and drop-off orderings of patients, respectively, where 0 and $n+1$ denote the hospital. Next, let $Q=\langle 0,q_1,q_2,\ldots,q_{|\mathcal{N}_k|},n+1\rangle$ be a pick-up ordering of patients such that

$$l_{q_1q_2}+2l_{q_2q_3}+\cdots+|\mathcal{N}_k|\cdot l_{q_{|\mathcal{N}_k|}(n+1)} = \min_I\{l_{i_1i_2}+2l_{i_2i_3}+\cdots+|\mathcal{N}_k|\cdot l_{i_{|\mathcal{N}_k|}(n+1)}\}; \quad (51)$$

i.e., Q corresponds to a pick-up ordering of patients that results in the smallest value of the total travel time.

Let patient $p\in\mathcal{N}_k$ be the vth patient that is dropped off (in drop-off ordering J) and the uth patient who is picked up (in pick-up ordering I). Then the service time cost of patient p, denoted by $C_p^{(u,v)}$, is

$$C_p^{(u,v)} = c^r(l_{0j_1}+l_{j_1j_2}+\cdots+l_{j_{v-1}j_v}+l_{i_ui_{u+1}} + l_{i_{u+1}i_{u+2}}+\cdots+l_{i_{(u+1)}i_{|\mathcal{N}_k|}}) + c^h\cdot\sum_{u\in\mathcal{N}_k} d_u,$$

and the total service time cost of the patients in $\mathcal{N}_k$ is

$$\begin{aligned}
&c^r\Phi^r+c^h\Phi^h\\
&= c^r\big(|\mathcal{N}_k|\cdot l_{0j_1}+(|\mathcal{N}_k|-1)\cdot l_{j_1j_2}+\cdots+l_{j_{|\mathcal{N}_k|-1}j_{|\mathcal{N}_k|}} + |\mathcal{N}_k|\cdot l_{0i_{|\mathcal{N}_k|}}+(|\mathcal{N}_k|-1)\cdot l_{i_{|\mathcal{N}_k|-1}i_{|\mathcal{N}_k|}}+\cdots+l_{i_1i_2}\big)\\
&\quad + c^h\cdot|\mathcal{N}_k|\cdot\sum_{u\in\mathcal{N}_k} d_u\\
&= c^r\big(|\mathcal{N}_k|\cdot(l_{0j_1}+l_{0i_{|\mathcal{N}_k|}})+(|\mathcal{N}_k|-1)\cdot(l_{j_1j_2}+l_{i_{|\mathcal{N}_k|-1}i_{|\mathcal{N}_k|}}) +\cdots+(l_{j_{|\mathcal{N}_k|-1}j_{|\mathcal{N}_k|}}+l_{i_1i_2})\big)+c^h\cdot|\mathcal{N}_k|\cdot\sum_{u\in\mathcal{N}_k} d_u\\
&\geq c^r\big(|\mathcal{N}_k|\cdot(l_{0j_1}+l_{0q_{|\mathcal{N}_k|}})+(|\mathcal{N}_k|-1)\cdot(l_{j_1j_2}+l_{q_{|\mathcal{N}_k|-1}q_{|\mathcal{N}_k|}}) +\cdots+(l_{j_{|\mathcal{N}_k|-1}j_{|\mathcal{N}_k|}}+l_{q_1q_2})\big)+c^h\cdot|\mathcal{N}_k|\cdot\sum_{u\in\mathcal{N}_k} d_u\\
&\geq 2c^r\big(l_{q_1q_2}+\cdots+(|\mathcal{N}_k|-1)\cdot l_{q_{|\mathcal{N}_k|-1}q_{|\mathcal{N}_k|}}+|\mathcal{N}_k|\cdot l_{0q_{|\mathcal{N}_k|}}\big) + c^h\cdot|\mathcal{N}_k|\cdot\sum_{u\in|\mathcal{N}_k|} d_u,
\end{aligned}$$

where the last two inequalities follow by the definition of Q in (51). Thus,

$$\sum_{p\in\mathcal{N}_k} C_p^{(u,v)} \geq c^r\Phi^r+c^h\Phi^h \geq 2c^r(l_{q_1q_2}+\cdots+|\mathcal{N}_k|\cdot l_{0q_{|\mathcal{N}_k|}})+c^h\cdot|\mathcal{N}_k|\cdot\sum_{u\in\mathcal{N}_k} d_u, \quad (52)$$

which implies that pick-up ordering Q and its reverse for drop-off ordering is at least as good as using $I_{\mathcal{N}_k}$ and $J_{\mathcal{N}_k}$. □

The linear programming (LP) relaxation of BSPRP is usually weak because of the presence of the big-M constraints. However, using Proposition 1 we reformulate BSPRP to obtain an equivalent MIP model that provides better LP relaxation bounds. Special cases of this formulation with $c^h=0$, or, equivalently, $d_i=d_j$ for all $i,j\in\mathcal{N}$, are considered by Ngueveu et al. (2010) and Heilporn et al. (2010) for the K-traveling repairman problem.

Let binary variable $\gamma_{ij}^k=1$ if both patients i and j are assigned to the same time stage k. Then using (52) we obtain the following equivalent MIP model:

[BSPRP]

$$\min_{x,z,\gamma,t} \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{N}}\left\{2c^r t_i^k + c^h\left(d_i z_i^k + \sum_{j\in\mathcal{N}\setminus\{i\}} d_j\gamma_{ij}^k\right)\right\} \quad (53)$$

subject to (32)–(38),

$$\gamma_{ij}^k = z_i^k z_j^k, \quad \forall\, i,j\in\mathcal{N}, \forall\, k\in\mathcal{K}, \quad (54)$$

$$t_i^k \geq 0, \quad \forall\, i\in\tilde{\mathcal{N}}, \forall\, k\in\mathcal{K}, \quad (55)$$

$$x_{ij}^k, z_i^k, \gamma_{ij}^k \in\{0,1\}, \quad \forall\, i,j\in\tilde{\mathcal{N}}, \forall\, k\in\mathcal{K}. \quad (56)$$

Note that nonlinear constraints (54) link the patient assignment decisions to the time stages. They can be easily linearized using a standard approach with an additional set of constraints. While MIP model (53)–(56) provides tighter LP relaxation bounds compared to the previous formulation, its solution using commercial MIP solvers is still computationally expensive; e.g., after three hours CPLEX 12.4 yields optimality gaps of at least 29% in solving instances of BSPRP with 20 patients and four stages. To overcome this concern, in the next section we present a set-partitioning reformulation of BSPRP that we consequently exploit to develop a branch-and-price algorithm.

3. A Branch-and-Price Approach

Because of the special structure of the objective function in BSPRP, Proposition 1 indicates that it is always optimal to pick up the patients in the reverse order of their drop-off sequence and vice versa. We refer to an assignment of patients to a time stage and their drop-off sequence as a transportation *route*, which is denoted by $r = (i_1, i_2, \ldots, i_k)$ with $|r| = k$ throughout the rest of this section. Furthermore, we define the position of patient j in the drop-off ordering as $r[j] = i_j$ with an additional convention that $r[0] = 0$. Denote by ϕ_r the total cost of travel and hospital times for all patients visited in route r:

$$\phi_r = 2c^r \sum_{j=0}^{|r|-1} (|r| - j) l_{r[j],\, r[j+1]} + c^h |r| \sum_{j=1}^{|r|} d_{r[j]},$$

which is an equivalent reformulation of the terms in (53).

Next, we consider a sequence of OR times available across all time stages, i.e., L^k, $k \in \mathcal{K}$. In particular, let $L_{(m)}$ be the mth smallest value in $\{L^1, \ldots, L^K\}$, where $m \in \mathcal{M} = \{1, \ldots, m_{\max}\}$ with $m_{\max} \leq K$, and $K(m)$ be the number of stages with total OR session length of $L_{(m)}$. Furthermore, let R be the set of all potential feasible transportation routes and binary decision variable θ_r, $r \in R$, be equal to one if route r is chosen in the corresponding solution. Additionally, let parameter $a_{ir} \in \{0, 1\}$ indicate whether route r includes patient i, and let $b_{mr} \in \{0, 1\}$ indicate whether route r satisfies $L_{(m-1)} < \sum_{j=1}^{|r|} d_{r[j]} \leq L_{(m_{\max})}$ with $L_{(0)} = 0$. BSPRP can then be reformulated as

$$[\text{SP}] \qquad \min \sum_{r \in R} \phi_r \theta_r \tag{57}$$

$$\text{subject to} \sum_{r \in R} a_{ir} \theta_r = 1, \quad \forall\, i \in \mathcal{N}, \tag{58}$$

$$\sum_{r \in R} b_{mr} \theta_r \leq \sum_{j=m}^{m_{\max}} K(j), \quad \forall\, m \in \mathcal{M}, \tag{59}$$

$$\theta_r \in \{0, 1\}, \quad \forall\, r \in R, \tag{60}$$

where constraints (58) ensure that each patient is assigned to exactly one route and constraint (59) limits the total number of possible route selections with respect to the OR availability times.

Note that model (57)–(60) assumes that there always exists a feasible schedule. This restriction requires only a slight modification of the proposed approach when adjusting for more general problem settings, where not all patients can be scheduled because of either the OR session length restrictions or the vehicle capacity limitations. Specifically, to capture this generalization, we can replace the objective function of SP by

$$\min \alpha \left\{ |\mathcal{N}| + \sum_{r \in R} \left(\phi_r - \alpha \sum_{i \in \mathcal{N}} a_{ir} \right) \theta_r \right\}, \tag{61}$$

where $\alpha > 0$ is a fixed penalty cost of not providing service to a patient. Furthermore, this modification requires constraints (58) to be relaxed as $\sum_{r \in R} a_{ir} \theta_r \leq 1$ for all $i \in \mathcal{N}$. Observe that if all patients are served, then (61) reduces to (57). Nevertheless, in the remainder of this section, for simplicity of exposition we focus on the model (57)–(60).

In general, the LP relaxation of formulation (57)–(60) provides tight bounds. However, the size of all potential routes $|R|$ is $O(n^\kappa)$, which is quite large in practice even for fixed values of κ. This results in an excessive number of decision variables and model parameters in (57)–(60). Therefore, we propose to employ a branch-and-price framework to generate promising routes on an "as needed" basis. In the remainder of the paper, we refer to the LP relaxation of model (57)–(60) as the master problem (MP).

3.1. Route Generation

Branch and price is a technique that incorporates column generation within a branch-and-bound procedure (Barnhart et al. 1998). Define the restricted master problem, denoted by RMP(R'), as the LP relaxation of SP that consists of a restricted set of columns generated so far, denoted by $R' \subseteq R$. At each node of the search tree, the column generation method iteratively solves the RMP(R') and a pricing problem. The purpose of the pricing problem is either to produce columns with the most negative reduced costs based on the dual solution of the current RMP(R') or to prove that none exists. At each iteration, newly generated columns are introduced to RMP(R'), and the process terminates and a lower bound for the corresponding node is obtained whenever no additional column prices out favorably. We refer the reader to Barnhart et al. (1998) for a more detailed description of the branch-and-price framework.

Consider the following dual variables of RMP(R'):

- π_i: dual variable corresponding to constraint (58) for patient i;
- μ_m: dual variables corresponding to constraint (59) for session index m.

The reduced cost $\bar{\phi}_r$ of a potential route, $r \in R$, is given by

$$\bar{\phi}_r = \phi_r - \sum_{i \in \mathcal{N}} a_{ir} \pi_i - \sum_{m \in \mathcal{M}} b_{mr} \mu_m. \tag{62}$$

Note that any optimal solution of RMP(R') is a feasible solution to MP. However, it is not (necessarily) an optimal solution unless there is no column left in $R \backslash R'$ that prices out favorably. This dynamic process of generating columns is called pricing, and the problem itself is referred to as the pricing problem. Given a dual solution π to RMP(R'), an integer λ, and the OR length $L_{(m)}$ for $m \in \mathcal{M}$, we propose an integer-programming formulation for the pricing problem, referred to as $\text{RPP}^{\lambda, L_{(m)}}(\pi)$. Similar formulations are considered in Picard and Queyranne (1978), Fox et al. (1980), Bigras et al. (2008), Queyranne and Schulz (1994), Keha et al. (2009), and Heilporn et al. (2010) when $d_i = 0$ for all $i \in \mathcal{N}$ or $c^h = 0$.

For $1 \leq \lambda \leq \kappa$ and $m \in \mathcal{M}$, let $R(\lambda, L_{(m)})$ be the set of all potential transportation routes r that satisfy $|r| = \lambda$ and $b_{mr} = 1$. Our pricing problem determines a column (i.e., drop-off route) with the smallest reduced cost among those in $R(\lambda, L_{(m)})$ for each $\lambda \leq \kappa$ and $m \in \mathcal{M}$. Specifically, we model this problem as a problem of finding a path with the smallest cost on a multilayer network, denoted by $\mathcal{F}^\lambda$. Each network $\mathcal{F}^\lambda$ consists of λ layers, where each layer encompasses n nodes associated with the patients. Moreover, $\mathcal{F}^\lambda$ includes nodes 0 and $n+1$ that represent the start and the end location of each route. Starting from node 0 and ending at $n+1$, each feasible path on $\mathcal{F}^\lambda$ is composed of distinct nodes selected at each layer, defining the drop-off sequence of the selected patients, which also provides us with the pick-up sequence according to Proposition 1.

The integer programming formulation of the problem $\text{RPP}^{\lambda, L_{(m)}}(\pi)$ uses the following decision variables:

- $\zeta_i^t \in \{0, 1\}$ if node i is selected in layer t, for $i \in \mathcal{N}$ and $t = 1, \ldots, \lambda$;
- $w_{ij}^t \in \{0, 1\}$ if nodes i and j are selected in layers t and $t+1$, for $i, j \in \mathcal{N} (i \neq j)$ and $t = 1, \ldots, \lambda - 1$.

Following Heilporn et al. (2010), we refer to these variables as the *position* and the *transition* variables, respectively. Given λ as the possible number of patients in a route, the objective function for $\text{RPP}^{\lambda, L_{(m)}}(\pi)$ can be modeled as

$$\mathcal{Z}_{\lambda, L_{(m)}}(\zeta_i^t, w_{ij}^t) = 2c^r \left(\lambda \sum_{j \in \mathcal{N}} l_{0j} \zeta_j^1 + \sum_{t=1}^{\lambda-1} \sum_{i, j \in \mathcal{N}, i \neq j} (\lambda - t) l_{ij} w_{ij}^t \right) + \sum_{t=1}^{\lambda} \sum_{i \in \mathcal{N}} (c^h \lambda d_i - \pi_i) \zeta_i^t.$$

As a result, for a given $\lambda \leq \kappa$ and $m \in \mathcal{M}$ a column, $\hat{r}$ in $R \backslash R'$, for which $\mathcal{Z}_{\lambda, L_{(m)}}(\zeta_i^t, w_{ij}^t) - \sum_{m \in \mathcal{M}} b_{m\hat{r}} \mu_m \leq 0$, is priced out and introduced to $\text{RMP}(R')$. Therefore, using the dual variable π, the pricing problem can be described as

$[\text{RPP}^{\lambda, L_{(m)}}(\pi)]$

$$\text{minimize} \quad \mathcal{Z}_{\lambda, L(m)} \tag{63}$$

$$\text{subject to} \quad \sum_{t=1}^{\lambda} \zeta_i^t \leq 1, \quad \forall i \in \mathcal{N}, \tag{64}$$

$$\sum_{i \in \mathcal{N}} \zeta_i^t = 1, \quad t = 1, \ldots, \lambda, \tag{65}$$

$$\sum_{j \in \mathcal{N} \backslash \{i\}} w_{ij}^t = \zeta_i^t, \quad t = 1, \ldots, \lambda - 1, \forall i \in \mathcal{N}, \tag{66}$$

$$\sum_{j \in \mathcal{N} \backslash \{i\}} w_{ji}^t = \zeta_i^{(t+1)}, \quad t = 1, \ldots, \lambda - 1, \forall i \in \mathcal{N}, \tag{67}$$

$$\sum_{j \in \mathcal{N}} \sum_{t=1}^{\lambda} d_j \zeta_j^t \leq L_{(m_{\max})}, \tag{68}$$

$$\sum_{j \in \mathcal{N}} \sum_{t=1}^{\lambda} d_j \zeta_j^t \geq L_{(m-1)} + \epsilon, \tag{69}$$

$$\zeta_i^t \in \{0, 1\}, \quad t = 1, \ldots, \lambda, \forall i \in \mathcal{N}, \tag{70}$$

$$w_{ij}^t \in \{0, 1\}, \quad t = 1, \ldots, \lambda - 1, \forall i, j \in \mathcal{N}. \tag{71}$$

where ϵ is a sufficiently small constant parameter such that $0 < \epsilon < \min_{m \in \mathcal{M}} \{L_{(m)} - L_{(m-1)}\}$.

The formulation ensures that each node is visited at most once (64) and exactly one node is selected in each layer (65), while constraints (66) and (67) link the position and transition variables. Finally, constraints (68) and (69) impose session availability restrictions. It can be easily shown that $w_{ij}^t = \zeta_i^t \cdot \zeta_j^{t+1}$ for all $i, j \in \mathcal{N}$ in an optimal solution. Hence, the binary restrictions for variables w_{ij}^t can be relaxed and replaced by nonnegativity requirements.

Our computational tests in Section 4 show that the proposed pricing problem is solvable using commercial solvers. However, in the next section we introduce a heuristic approach to generate multiple potential columns to enhance the performance of the method at the first iterations of the column generation method.

3.2. Algorithmic Enhancements

Next, we discuss several computational considerations that are important in implementing our branch-and-price algorithm. In particular, we describe our branching strategy in Section 3.2.1 and introduce a greedy column generation heuristic in Section 3.2.2.

3.2.1. Patient-Pair Branching. Using the traditional branching on variables for solving large-scale set-partitioning problems often results in an unbalanced search tree and may destroy the structure of the pricing problem (Mehrotra et al. 1998). To overcome this difficulty, we adapt an alternative branching scheme from Ryan and Foster (1981) that is known to be a very effective strategy for set-partitioning applications including vehicle routing and crew scheduling problems (Barnhart et al. 1998).

The branching scheme used in Ryan and Foster (1981) is based on the following proposition. Although the authors do not consider column generation, their branching scheme is effective in this context as well (Barnhart et al. 1998).

Algorithm 1 (Greedy heuristic algorithm for initial transportation route generation)

Input: κ
Output: Δ_R

set $\Delta_R = \varnothing$
foreach $m \in \mathcal{M}$ **do**
  foreach $\lambda = \kappa, \ldots, 2$ **do**
    foreach $i \in \mathcal{N}$ **do**
      set $\mathcal{S} = \mathcal{N} \backslash \{i\}$, $r = \varnothing$, and let $r[1] = i$, $\mathcal{L} = \lambda(2c^r l_{0i} + c^h d_i)$, $D = d_i$, and $j = 1$
      while $j \le \lambda$ *and* $\mathcal{S} \neq \varnothing$ **do**
        for $k \in \mathcal{S}$ **do**
          if $D + d_k > L_{(m_{\max})}$ **then**
            $\mathcal{S} = \mathcal{S} \backslash \{k\}$
        if $\mathcal{S} \neq \varnothing$ **then**
          $r[j] \in \arg\min_{k \in \mathcal{S}} \{2(\lambda - j)c^r l_{kr[j]} + \lambda c^h d_k\}$
          $\mathcal{L} = \mathcal{L} + 2(\lambda - j)c^r l_{r[j+1]r[j]} + \lambda c^h d_{r[j+1]}$
          $D = D + d_{r[j+1]}$
          $\mathcal{S} = \mathcal{S} \backslash \{r[j+1]\}$, and $j = j + 1$
      if $|r| = \lambda$ *and* $D > L_{(m-1)}$ **then**
        $\Delta_R = \Delta_R \bigcup \{r\}$

Proposition 2 (Barnhart et al. 1998). *If A is a 0–1 matrix, and a basic solution to $Ax=1$ is fractional, i.e., at least one of the components of x is fractional, then there exist two rows s and t of the master problem such that*

$$0 < \sum_{k:\, a_{sk}=1,\ a_{tk}=1} x_k < 1.$$

Note that the matrix associated with the set-partitioning constraints (58) in any restricted master problem is a 0–1 matrix. Hence, using the result of Proposition 2, the pair of branching constraints are

$$\sum_{k:\, a_{sk}=1,\, a_{tk}=1} \theta_k = 0 \quad \text{and} \quad \sum_{k:\, a_{sk}=1,\, a_{tk}=1} \theta_k = 1;$$

i.e., rows s and t are covered by different columns in the first (left) case and by the same column in the second (right) case. In our application each row corresponds to a patient, so in the first branch we force patients s and t to be assigned to different stages, by introducing the following constraint into the pricing problem:

$$\sum_{k=1}^{\lambda} \zeta_s^k + \sum_{k=1}^{\lambda} \zeta_t^k \le 1. \tag{72}$$

We refer to constraint (72) as the *decoupling* branching constraint. Similarly, we introduce a *coupling* branching constraint as

$$\sum_{k=1}^{\lambda} \zeta_s^k = \sum_{k=1}^{\lambda} \zeta_t^k, \tag{73}$$

into the pricing problem on the other branch, where (73) indicates that patients t and s should be assigned to the same stage. We refer to this specialized branching rule as *branching on patient-pairs*. The number of patients is finite, so the branch-and-price algorithm terminates finitely under this branching scheme.

3.2.2. Improving Column Generation. Often during initial column generation iterations, large values of dual variables can negatively impact solution times of the subproblems. To avoid this, we initialize the restricted master problem with a set of columns that is sufficient to obtain a feasible solution. The convergence rate of the branch-and-price algorithm can be further improved, if the set of the initial columns consists of those that are likely to be in the final optimal set. Therefore, we propose to generate this set heuristically using a greedy approach.

For each $L_{(m)}$, the proposed heuristic (see Algorithm 1) identifies a set of stage assignments consisting of at most κ patients and their transportation routes (the drop-off sequence), denoted by $\Delta_R \subseteq \bigcup_{\lambda \le \kappa} R(\lambda, L_{(m)})$. Specifically, the algorithm iteratively constructs a drop-off sequence of $\lambda \le \kappa$ patients, r, such that $|r| = \lambda$ and $r[1] = i$ for all $i \in \mathcal{N}$, where $|r|$ is the number of patients in the route and $r[j]$ denotes the jth patient in the sequence. When constructing each r, the set of patients who have not been included so far is denoted by $\mathcal{S}$.

The heuristic selects a patient $r[j+1]$ to be visited next from the set of unvisited patients $\mathcal{S}$ in order to minimize the term $2(\lambda - j)c^r l_{r[j+1]r[j]} + \lambda c^h d_{r[j+1]}$, which is consistent with the objective function $\mathcal{Z}_{\lambda, L(m)}(\cdot)$ in the pricing problem. Furthermore, condition $D + d_{r[j+1]} \le L_{(m_{\max})}$ imposes the overtime restriction on the sum of surgery durations assigned to the stage when selecting the next patient to visit at iteration j. This process continues until the number of visited patients equals a desired length of the route λ, providing the estimated cost of the route $\mathcal{L}$. These routes are used to populate the initial master problem at the root node.

Algorithm 2 (Greedy heuristic algorithm for the pricing problem)

Input: $\kappa, \pi, \mu, \mathcal{N}_p^c = \{(u_1, v_1), \ldots, (u_{n_p^c}, v_{n_p^c})\}, \mathcal{N}_p^d = \{(\bar{u}_1, \bar{v}_1), \ldots, (\bar{u}_{n_p^d}, \bar{v}_{n_p^d})\}$
Output: Δ_R

set $\Delta_R = \varnothing$
foreach $m \in \mathcal{M}$ **do**
  foreach $\lambda = \kappa, \ldots, 2$ **do**
    foreach $i \in \mathcal{N}$ **do**
      set $\mathcal{S} = \mathcal{N} \backslash \{i\}$, $r = \varnothing$, and let $r[1] = i$, $\mathcal{L} = \lambda(2c^r l_{0i} + c^h d_i) - \pi_i$, $D = d_i$, and $j = 1$
      while $j \leq \lambda$ *and* $\mathcal{S} \neq \varnothing$ **do**
        for $h = 1, \ldots, n_p^d$ **do**
          if $r[j] = \bar{u}_h$ **then**
            $\mathcal{S} = \mathcal{S} \backslash \{\bar{v}_h\}$
          if $r[j] = \bar{v}_h$ **then**
            $\mathcal{S} = \mathcal{S} \backslash \{\bar{u}_h\}$
        for $k \in \mathcal{S}$ **do**
          if $D + d_k > L_{(m_{\max})}$ **then**
            $\mathcal{S} = \mathcal{S} \backslash \{k\}$
        if $\mathcal{S} \neq \varnothing$ **then**
          let $r[j+1] \in \arg\min_{k \in \mathcal{S}} \{2(\lambda - j)c^r l_{kr[j]} + \lambda c^h d_k - \pi_k\}$, and $\mathcal{S} = \mathcal{S} \backslash \{r[j+1]\}$
          set $\mathcal{L} = \mathcal{L} + 2(\lambda - j)c^r l_{r[j+1]r[j]} + \lambda c^h d_{r[j+1]} - \pi_{r[j+1]}$, $D = D + d_{r[j+1]}$, and $j = j + 1$
      if $|r| = \lambda$ **then**
        let *feasible* $=$ *True*
        for $h = 1, \ldots, n_p^c$ **do**
          if $u_h \in r$ *and* $v_h \notin r$ **then**
            feasible $=$ *False*
          if $v_h \in r$ *and* $u_h \notin r$ **then**
            feasible $=$ *False*
        $\hat{m} = \arg\max_{m \in \mathcal{M}} \{m\colon L_{(m-1)} < \mathcal{L} \leq L_{(m_{\max})}\}$
        if $D > L_{(m-1)}$, $\mathcal{L} - \sum_{h=1}^{\hat{m}} \mu_h < 0$ *and feasible* $=$ *True* **then**
          set $\Delta_R = \Delta_R \bigcup \{r\}$

When generating a feasible column at the column generation stage, one might consider generating multiple columns simultaneously (Barnhart et al. 1998). Sol (1994) shows that multiple column generation can be efficient for set-partitioning problems. Moreover, usually there is more than one column with negative reduced cost at the column generation stage (especially in initial stages), so it may be advantageous to generate some subset of columns heuristically without solving the exact pricing problem; see Desaulniers (2010).

To do so, we employ a greedy procedure (see Algorithm 2) by modifying Algorithm 1 to accommodate the dual variables and the branching information available at each node of the branch-and-price tree. However, if the heuristic approach (namely, Algorithm 2) fails to identify appropriate columns with negative reduced cost (i.e., $\Delta_R = \varnothing$), we resort to solving the exact pricing problem to ensure the correctness of the overall solution method.

In Algorithm 2, define $\mathcal{N}_p^c = \{(u_1, v_1), \ldots, (u_{n_p^c}, v_{n_p^c})\}$ and $\mathcal{N}_p^d = \{(\bar{u}_1, \bar{v}_1), \ldots, (\bar{u}_{n_p^d}, \bar{v}_{n_p^d})\}$ to be the coupling and decoupling constraints at node p of the branch and bound tree, respectively. In these sets each coupling pair (u_i, v_i), $i = 1, \ldots, n_p^c$ corresponds to patients u_i and v_i that are enforced to be at the same stage, while each decoupling pair $(\bar{u}_j, \bar{v}_j)$, $j = 1, \ldots, n_p^d$ indicates that patients $\bar{u}_j$ and $\bar{v}_j$ should be scheduled at different stages. Algorithm 2 is a greedy algorithm that iteratively constructs a feasible transportation route, consisting of a desired number of patients λ that (i) retains the branching rule restrictions, (ii) satisfies the no-overtime constraint (38) for the given stage, and (iii) calculates the proper modified cost using the dual information. To achieve (i), we dynamically modify the set of feasible unvisited patients $\mathcal{S}$ with respect to decoupling branches whenever a patient is selected. The information about the active coupling branches is exploited when a complete route is constructed at the last step of the inner loop in the algorithm. Similar to our strategy in Algorithm 1, (ii) is verified by checking the corresponding constraint whenever a new patient is considered. Finally, (iii) is achieved by considering the dual prices as penalty costs when selecting any particular patient.

4. Computational Study

In this section, we first describe test instances considered in our study, followed by the results of our computational experiments. The focus is on comparing our proposed approach to a commercial solver (CPLEX 12.4), evaluating the performance of our solution techniques, and highlighting the value of integrated surgery scheduling and vehicle routing decisions. We implement our branch-and-price algorithm using BCP (version 1.3.4), a framework for branch, cut, and price, using its default settings (Ralphs and Ladanyi 2003). All computational experiments are conducted on an Intel Xeon PC with 3 GHz CPU and 3 GB RAM. Finally, we note that Tables 5–7 that contain details of our computational experiments in Sections 4.3 and 4.4 are provided in the online supplement (available as supplemental material at http://dx.doi.org/10.1287/ijoc.2016.0706) of the paper.

4.1. Test Instances

We use data provided by a VHA hospital in Pittsburgh that include durations and turnover times for all surgery cases performed between 2006 to 2009. However, this data set does not include patients' location information for privacy reasons. To alleviate this issue, we generate synthetic patients by sampling a residence in Western Pennsylvania from publicly available data on the veteran population provided by the VHA (Department of Veterans Affairs 2013) and the operational data of surgery durations. The surgery durations are point estimates of each surgery request that include the pre- and post-incision times (see Shylo et al. 2013 for more details). We chose ophthalmology procedures, as these are typically outpatient procedures.

Table 1 provides the key characteristics of these problem instances. To be consistent with the historical records on the average number of surgeries in the surgery data set, we restrict our problem instances to those composed of 4 to 6 possible surgeries to be scheduled at each stage. Throughout our experiments, we consider equal routing and hospital time costs; i.e., $c^r = c^h$.

Table 1 Test Instances

Instance class	$\lvert\mathcal{N}\rvert$	K	κ
P1	16	4	4
P2	20	4	5
P3	24	4	6
P4	20	5	4
P5	25	5	5
P6	30	5	6
P7	24	6	4
P8	30	6	5
P9	36	6	6

4.2. Solution Method Performance: CPLEX vs. B&P Algorithm

In Table 2, we report the performance of our branch-and-price algorithm against CPLEX 12.4 (with default settings) for 18 problem instances. Some of these instances are smaller than those described in Table 1 as CPLEX cannot handle most of the instances in Table 1. The results of the branch-and-price algorithm reported in Tables 2 and 3 are obtained using a single column generation approach in which the pricing problem is solved exactly. The discussion of other possible strategies is presented in Section 4.3.

In Table 2 we consider two classes of test instances. The first, which we refer to as ESL, contains the problems with equal session lengths (i.e., OR time availability) in all times stages; i.e., $L^{k_1} = L^{k_2}$ for all k_1, $k_2 \in \mathcal{K}$. The other class, which is referred to as DSL, consists of instances with different values of L^k for each $k \in \mathcal{K}$. For example, in Table 2 DSL instances with three stages and 12 patients have $L^1 = 360$, $L^2 = 300$ and $L^3 = 210$, while ESL instances correspond to $L^1 = L^2 = L^3 = 360$; similarly, DSL instances with four stages and 24 patients have $L^1 = 300$, $L^2 = 465$ and $L^3 = 450$, while ESL instances correspond to $L^1 = L^2 = L^3 = 465$.

Computational results in Table 2 demonstrate that CPLEX 12.4 cannot solve the selected instances to optimality except for some small-sized ones within a given three-hour time limit, although it finds near optimal integer feasible solutions in some cases. However, our branch-and-price approach is able to obtain an optimal solution and establish its optimality for all instances within four minutes. The performance of CPLEX slightly improves for most of DSL problem instances, highlighting the effect of symmetry in the solution space of ESL instances. On the other hand, it is important to note that the performance of the proposed branch-and-price algorithm is not significantly affected as a result of considering unequal session lengths across the stages.

Our computational study in Table 2 is based on the assumption that $K \cdot \kappa = \lvert\mathcal{N}\rvert$; i.e., there is sufficient capacity to schedule all patients (recall our notation that $K = \lvert\mathcal{K}\rvert$). This assumption has the side effect of making hospital times equal across all solutions. In Table 3 we provide additional computational results using ESL instances with $K \cdot \kappa \le \lvert\mathcal{N}\rvert$, which makes hospital costs a somewhat more important factor. Specifically, each row of this table refers to a specific combination of the number of stages and the capacity of each vehicle available at each stage; e.g., the first row shows a problem setting with three stages, where at each stage a vehicle of capacity four is available. Next, for each stage and vehicle capacity combination we first generate a problem instance with $\kappa \cdot \lvert\mathcal{K}\rvert$ patients. We then increase the number

Table 2 **Performance of the Developed Branch-and-Price Approach Compared to CPLEX**

				B&P (sec.)		CPLEX (sec.)		CPLEX opt. gap (%)		CPLEX best vs. B&P. (%)	
K	κ	Patients	Instance	ESL	DSL	ESL	DSL	ESL	DSL	ESL	DSL
3	4	12	1	1.14	1.24	13.37	10.91	0.00	0.00	0.00	0.00
			2	1.60	1.49	12.37	13.16	0.00	0.00	0.00	0.00
			3	0.37	0.32	14.31	12.05	0.00	0.00	0.00	0.00
3	5	15	1	11.43	12.35	151.96	122.87	0.00	0.00	0.00	0.00
			2	37.65	38.57	143.36	130.00	0.00	0.00	0.00	0.00
			3	9.30	15.22	144.10	157.98	0.00	0.00	0.00	0.00
3	6	18	1	13.09	12.02	3,038.93	3,152.90	0.00	0.00	0.00	0.00
			2	11.32	11.70	1,682.19	1,597.68	0.00	0.00	0.00	0.00
			3	245.25	268.43	5,593.52	5,422.93	0.00	0.00	0.00	0.00
4	4	16	1	1.94	10.87	2,410.89	2,361.94	0.00	0.00	0.00	0.00
			2	25.72	24.57	10,645.76	9,976.88	0.00	0.00	0.00	0.00
			3	5.65	4.46	6,677.41	6,325.04	0.00	0.00	0.00	0.00
4	5	20	1	7.59	8.94	8,391.75	8,001.00	0.00	0.00	0.00	0.00
			2	7.82	10.06	>10,800	>10,800	29.27	27.55	14.32	13.39
			3	58.98	61.75	>10,800	>10,800	16.87	17.22	16.6	16.87
4	6	24	1	132.76	250.98	>10,800	>10,800	37.99	39.00	25.10	23.88
			2	51.59	79.43	>10,800	>10,800	38.6	38.14	36.78	35.80
			3	64.97	62.00	>10,800	>10,800	39.35	37.90	30.14	29.44

Notes. The branch-and-price approach solves all instances to optimality. For CPLEX we report its optimality gaps and the quality of its best feasible solutions.

of patients incrementally by adding three randomly generated patients at each step and proceed until we generate four problem instances. Table 3 reports (1) the optimal objective function value for each problem instance, (2) solution times using the branch-and-price algorithm, (3) solution times using CPLEX, (4) the optimality gap (%) for the CPLEX solution as reported by the solver, and (5) the optimality gap (%) with respect to the optimal solution found by B&P approach. Note that the objective function is of the form (61) and equality constraints (58) are relaxed to be inequalities.

Table 3 **Evaluating the Effect of the Number of Patients on Performance**

K	κ	Patients	Objective	B&P (sec.)	CPLEX (sec.)	CPLEX opt. gap (%)	CPLEX best vs. opt. (%)
3	4	12	4,502	1.14	13.37	0.00	0.00
		15	4,224	1.65	157.68	0.00	0.00
		18	4,092	6.42	2,136.04	0.00	0.00
		21	4,086	5.78	8,979.23	0.00	0.00
3	5	15	5,785	11.43	151.96	0.00	0.00
		18	5,785	23.65	3,145.67	0.00	0.00
		21	5,640	30.36	9,736.52	0.00	0.00
		24	5,622	29.42	>10,800	18.95	9.23
3	6	18	8,518	13.09	3,038.93	0.00	0.00
		21	8,304	17.12	6,492.81	0.00	0.00
		24	8,304	16.65	>10,800	21.89	13.87
		27	7,798	159.78	>10,800	29.90	19.72
4	4	16	4,924	1.94	2,410.89	0.00	0.00
		19	4,776	5.32	8,120.33	0.00	0.00
		22	4,637	27.49	>10,800	32.50	29.79
		25	4,637	24.33	>10,800	37.00	35.44
4	5	20	7,675	7.59	8,391.75	0.00	0.00
		23	7,386	13.78	>10,800	39.71	35.66
		26	6,536	11.25	>10,800	40.12	31.25
		29	6,536	46.98	>10,800	39.88	30.02

Notes. The branch-and-price approach solves all instances to optimality. For CPLEX we report its optimality gaps and the quality of its best feasible solutions.

Similar to the previous set of experiments, the results indicate that the performance of the branch-and-price algorithm is not considerably affected as the number of patients increases. One explanation for this observation is that solving a problem instance under the assumption $K\cdot\kappa \le |\mathcal{N}|$ requires only slight changes at the pricing stage of the B&P algorithm (specifically, the reduced costs of generated columns need to be modified accordingly). However, increasing the number of patients significantly deteriorates the performance of CPLEX because of the increased number of variables in the corresponding MIP formulation. In the remainder of the paper (except one of the studies discussed in Section 4.4; see the results in Table 7 of the online supplement) we focus on the problem instances with $K\cdot\kappa = |\mathcal{N}|$.

4.3. Solution Method Performance: B&P Algorithmic Strategies

Next, we evaluate the effect of different computational strategies (discussed in Section 3.2.2) on the performance of our branch-and-price algorithm. We consider four algorithmic strategies summarized in Table 4. Each strategy is characterized by the number of columns introduced into the restricted master problem at each column generation iteration (No. of columns), and the solution technique used for solving the pricing problem. Under the "Direct" strategy,

Table 4 Description of Column Generation Strategy

Strategy	No. of columns	Pricing strategy
SCD	Single	Direct
MCD	Multiple	Direct
SCH	Single	Heuristic
MCH	Multiple	Heuristic

the pricing problem is solved exactly using CPLEX. Under the "Heuristic" strategy, we first employ a heuristic approach (Algorithm 2); however, we resort to using CPLEX whenever the heuristic method fails. We further consider either single or multiple column generation combined with "Direct" and "Heuristic" approaches. Under the "Single" method, a single column with the most negative reduced cost is introduced to the restricted master problem at each iteration, while for the "Multiple" approach, we add a set of columns with negative reduced costs.

Table 5 (in the online supplement) reports results for five randomly generated instances of each problem class (Table 1) under each possible strategy, with the best strategy shown in bold for each row. These algorithmic methods are compared based on the solution time (in seconds), the number of explored nodes, and the depth of the resulting search trees as shown in Table 5.

Our computational results in Table 5 show that the solution time is rather sensitive to the algorithmic strategies. As expected, the numerical results also indicate that embedding a heuristic approach for the column generation typically reduces the overall solution time (SCH and MCH columns) compared to solving the pricing problem exactly. For example, for class P6 strategies SCD and MCD require 535 and 928 seconds on average, respectively, as compared to 384 and 396 seconds for strategies SCH and MCH, respectively. Furthermore, this example also demonstrates that adding multiple columns at each column generation iteration does not necessarily reduce the solution time when compared to a single column approach. However, for the majority of the test instances (specifically, 30 of 45 in Table 5), the multiple column generation scheme outperforms the single column generation approach. Finally, we should also note that in all problem instances the heuristic algorithm fails to find a column with negative reduced cost in at least one node of the branch-and-price tree.

4.4. Value of Integrating Surgery Scheduling and Vehicle Routing Decisions

In this study we model the integration of surgery scheduling and transportation to minimize the total service time cost of the patients. As emphasized earlier, such decisions are typically made independently. In practice surgeries are often assigned to ORs based on some scheduling rule (e.g., first-fit) depending on the OR availability upon the arrival time of the surgery requests and disregarding patient transportation considerations. Subsequently, given the obtained surgery schedule the vehicle routing decisions are made separately, possibly, using another optimization approach. To evaluate the value of the integrated decision-making framework, we compare our framework to heuristic methods, where scheduling and routing decisions are performed sequentially.

In our first set of experiments (see Table 6 in the online supplement) we assume that $K \cdot \kappa = |\mathcal{N}|$. Given a batch of surgery requests we randomly assign the patients to time stages satisfying stage and vehicle capacity constraints. Having assigned all the surgeries to the stages, we exploit the following two routing methods to determine the pick-up and drop-off schedule for the patients at each stage. First, we consider a traveling salesman problem based (TSP-based) approach, where the objective is to determine a set of minimum length tours with respect to the residential locations of the patients at each stage. Second, we consider a traveling repairman problem based (TRP-based) approach that minimizes the sum of the times necessary to visit all patients locations. The pick-up routes in the TRP-based method are assumed to be in the reverse order of the drop-off routes. Note that the resulting routing problems (TSP-based and TRP-based) are solved exactly to determine the optimal patient visit order by vehicles.

In Table 6, we compare the objective function values in BSPRP obtained using solutions of these two heuristics to those obtained by the branch-and-price method. For each problem instance (five in each problem class), we generate 10 random surgery-to-time-stage assignments, solve the resulting routing problems using TSP- and TRP-based approaches, and then compute an average objective function value for each instance. The columns labeled as Gap (%) in the table report for both heuristic methods the gap between these average objective function values and the optimal solution obtained by the branch-and-price method. When compared to the heuristic approaches, the integrated scheduling framework provides a considerable improvement in service costs, ranging between 9% to 25%. These observations show the value of integrating the routing and scheduling decisions and highlight that healthcare providers can potentially improve the quality of their services by considering scheduling and transportation decisions simultaneously. Note that for the considered objective, the TRP-based formulation should dominate the TSP-based one (assuming both are solved to optimality) as the TSP-based method considers the vehicle travel time instead of the total patients travel time as shown in Table 6. Indeed, according to our results the

TRP-based formulation outperforms the TSP-based method with (i) the average gap from the optimal solution value of 13%–18% as compared to 15%–22%, (ii) the minimum gap from the optimal solution value of 9% as compared to 11%, and (iii) the maximum gap from the optimal solution value of 20% as compared to 25%. In a sense, these results numerically illustrate the relative importance of considering the total patient travel time instead of simply the vehicle travel time.

In our second set of experiments (see Table 7 in the online supplement) we assume that $K \cdot \kappa < |\mathcal{N}|$. Under this approach we first assign the patients to the stages in order to minimize their total waiting times at the hospital and then determine the exact transportation routes for the patients in each stage using either a TSP- or TRP-based method similar to the computational study discussed above. In all approaches we use the same penalty α for not providing service to a patient; see the objective function in (61).

Our results are provided in Table 7 (in the online supplement). Each row in this table is associated with a problem instance composed of a specific number of stages, capacity of vehicles at each stage, and the total number of patients. For each problem instance (five in each problem class), we randomly generate 10 settings of session lengths in each stage and report the average objective function values obtained for each solution method. Similar to Table 6 the columns labeled as Gap (%) report the gap between the average objective function values of heuristics and the average value of optimal solutions obtained by the branch-and-price method. As in the previous set of experiments, it is evident from the obtained results that the TRP-based method outperforms the TSP-based one with (i) the average gap from the optimal solution value of 11%–15% as compared to 14%–18% for TSP and (ii) the maximum gap from the optimal solution value of 21% as compared to 29% (note that the minimum gaps from the optimal solution value is about 8% for both methods). More importantly, the results support our earlier observations (see Table 6 in the online supplement) that integrating decisions under a unified framework improves the total cost of solutions compared to both TSP- and TRP-based approaches.

Finally, we should note that in addition to two natural heuristic methods discussed above, other heuristic approaches are possible for solving difficult combinatorial optimization problems arising in our integrated framework. For example, we could consider scheduling all patients in a certain geographical area to the same time stage, which should allow decomposing the original problem into a set of smaller but more manageable subproblems. Ultimately, applicability of such heuristic ideas typically depends on the structure of the input data and requires further tuning. We leave it as an interesting topic of future research.

5. Conclusions

In this paper, we propose an integrated approach that simultaneously considers surgery scheduling and vehicle routing decisions for a given set of elective outpatient surgery requests using available ORs in a hospital. The overall objective is to minimize the total service cost that incorporates transportation and hospital times for all patients. We focus on elective surgeries that allow for some delay between the surgery request and its actual assignment to a specific OR and start time. Furthermore, the developed model assumes that all decisions are performed by the same scheduler, i.e., the central planner. Therefore, the proposed approach can be suitable for healthcare facilities (such as VHA hospitals) that operate their own fleets of transportation vehicles.

Because of the difficulty of the overall integrated problem, our main focus is on the special case of the problem, where the key assumption is that there is exactly one transportation vehicle available for each time stage and each patient's hospital waiting time is equal to the total surgery times of all the patients in the same stage. By exploiting the structure of the problem, we develop a branch-and-price algorithm, which is further enhanced with several algorithmic strategies to improve the overall solution efficiency.

Although our models are somewhat stylized and simplified versions of the decisions faced in practice, they capture the main trade-offs with respect to the interactions between the patient routing decisions and the OR scheduling decisions. One could argue that if the decision makers optimize the routing and scheduling decision separately, then the obtained solutions should be close to optimal. However, our experiments with test instances calibrated using real historical and geographical data illustrate that an integrated optimization approach can improve the overall performance by 10%–20%. Thus, our computational results clearly demonstrate that there exist opportunities for healthcare providers to improve the quality of their services by considering integrated approaches.

We view this paper as the first step in this direction. Future research should relax our key assumptions. In particular, we believe that more general settings with uncertain surgery durations (one of the challenging issues in practical OR scheduling) and uncertain transportation times are of foremost importance. Nevertheless, we should note that our models can be applied in a stochastic environment if one adds a carefully defined slack time in every surgery and transportation request to compensate for random events. Such a simple method is typical in OR scheduling practice if the decisions are performed using a deterministic approach with point estimates of surgery durations.

Furthermore, future research should involve various types of practical considerations faced in everyday OR scheduling, including (i) special equipment availability, e.g., some types of surgeries can be scheduled only at specific ORs; (ii) downstream capacity constraints, e.g., the post-anesthesia care unit capacity; (iii) surgeon availability, e.g., some surgeries cannot be scheduled too close to each other, or simultaneously, as they are performed by the same surgeon in, possibly, different ORs; (iv) patient preferences, e.g., morning versus afternoon appointments; and (v) transportation time-windows requirements, e.g., there may exist constraints for the earliest possible patient pick-up times. Admittedly, more general settings of our proposed model leads to more realistic problems but fall beyond the scope of this paper and can be considered as separate future research.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/ijoc.2016.0706.

Acknowledgments

The authors thank the anonymous reviewers and the associate editor for their constructive and helpful comments. The authors thank Robert Monte and Scott Moore of the Veterans Affairs Pittsburgh Healthcare System and Gabriel L. Zenarosa of the University of Pittsburgh for their assistance and feedback. This paper does not necessarily reflect the views of the Veterans Health Administration or its employees. This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development [Grant VA224-13-C-0576] and is also partially supported by National Science Foundation [Grants CMMI1333758 and CMMI1400009].

References

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1998) Branch-and-price: Column generation for solving huge integer programs. *Oper. Res.* 46(3):316–329.

Batun S, Denton BT, Huschka TR, Schaefer AJ (2011) Operating room pooling and parallel surgery processing under uncertainty. *INFORMS J. Comput.* 23(2):220–237.

Begur SV, Miller DM, Weaver JR (1997) An integrated spatial DSS for scheduling and routing home-health-care nurses. *Interfaces* 27(4):35–48.

Bigras LP, Gamache M, Savard G (2008) The time-dependent traveling salesman problem and single machine scheduling problems with sequence dependent setup times. *Discrete Optim.* 5(4):685–699.

Blake JT, Carter MW (1997) Surgical process scheduling: A structured review. *J. Soc. Health Systems* 5(3):17–30.

Blake JT, Donald J (2002) Mount Sinai Hospital uses integer programming to allocate operating room time. *Interfaces* 32(2): 63–73.

Burkhardt JE, Garrity R (2012) Sharing the costs of human services transportation. Transportation research record. *J. Transportation Res. Board* 2277(1):57–64.

Burkhardt JE, Rubino JM, Yum J (2011) *TCRP Research Results Digest 99: Improving Mobility for Veterans* (Transportation Research Board, National Academies, Washington, DC).

Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. *Eur. J. Oper. Res.* 201(3):921–932.

Chen Z-L (2010) Integrated production and outbound distribution scheduling: Review and extensions. *Oper. Res.* 58(1):130–148.

Day R, Garfinkel R, Thompson S (2012) Integrated block sharing: A win–win strategy for hospitals and surgeons. *Manufacturing Service Oper. Management* 14(4):567–583.

Denton BT (2009) Practice abstracts. *Interfaces* 39(6):549–551.

Department of Veterans Affairs (2013) The Veteran Population Model (VetPop2007). Accessed June 2013, http://www.va.gov/vetdata.

Desaulniers G (2010) Branch-and-price-and-cut for split-delivery vehicle routing problem with time windows. *Oper. Res.* 58(1): 179–192.

Dexter F (2016) Bibliography of operating room management articles. Accessed February 2016, http://www.FranklinDexter.net/bibliography_TOC.htm.

Dexter F, Traub RD, Macario A (2003) How to release allocated operating room time to increase efficiency: Predicting which surgical service will have the most underutilized operating room time. *Anesthesia Analgesia* 96(2):507–512.

Dexter F, Macario A, Traub RD, Hopwood M, Lubarsky DA (1999) An operating room scheduling strategy to maximize the use of operating room block time: Computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesthesia Analgesia* 89(1):7–20.

Dumas Y, Desrosiers J, Soumis F (1991) The pickup and delivery problem with time windows. *Eur. J. Oper. Res.* 54(1):7–22.

Erdogan SA, Denton BT (2011) Surgery planning and scheduling. Cochran JJ, Cox LA, Keskinocak P, Kharoufeh JP, Smith JC, eds. *Wiley Encyclopedia of Operations Research and Management Science* (John Wiley & Sons, New York).

Fischetti M, Laporte G, Martello S (1993) The delivery man problem and cumulative matroids. *Oper. Res.* 41(6):1055–1064.

Fitzpatrick AL, Powe NR, Cooper LS, Ives DG, Robbins JA (2004) Barriers to health care access among the elderly and who perceives them. *Amer. J. Public Health* 94(10):1789–1794.

Fox KR, Gavish B, Graves SC (1980) An *n*-constraint formulation of the (time-dependent) traveling salesman problem. *Oper. Res.* 28(4):1018–1021.

Gupta D (2007) Surgical suites' operations management. *Production Oper. Management* 16(6):689–700.

Gupta D, Denton BT (2008) Appointment scheduling in health care: Challenges and opportunities. *IIE Trans.* 40(9):800–819.

Hans E, Wullink G, van Houdenhoven M, Kazemier G (2008) Robust surgery loading. *Eur. J. Oper. Res.* 185(3):1038–1050.

Heilporn G, Cordeau JF, Laporte G (2010) The delivery man problem with time windows. *Discrete Optim.* 7(4):269–282.

Keha AB, Khowala K, Fowler JW (2009) Mixed integer programming formulations for single machine scheduling problems. *Comput. Indust. Engrg.* 56(1):357–367.

Laporte G (1992) The vehicle routing problem: An overview of exact and approximate algorithms. *Eur. J. Oper. Res.* 59(3): 345–358.

Lucena A (1990) Time-dependent traveling salesman problem—the deliveryman case. *Networks* 20(6):753–763.

Mehrotra A, Johnson EL, Nemhauser GL (1998) An optimization based heuristic for political districting. *Management Sci.* 44(8):1100–1114.

Miller CE, Tucker AW, Zemlin RA (1960) Integer programming formulation of traveling salesman problems. *J. ACM* 7(4):326–329.

Min D, Yih Y (2010) Scheduling elective surgery under uncertainty and downstream capacity constraints. *Eur. J. Oper. Res.* 206(3):642–652.

Ngueveu SU, Prins C, Wolfler-Calvo R (2010) An efficient memetic algorithm for the cumulative capacitated vehicle routing problem. *Comput. Oper. Res.* 37(11):1877–1885.

Okoro CA, Strine TW, Young SL, Balluz LS, Mokdad AH (2005) Access to health care among older adults and receipt of preventive services. Results from the behavioral risk factor surveillance system, 2002. *Preventive Medicine* 40(3):337–343.

Park J, Kim B (2010) The school bus routing problem: A review. *Eur. J. Oper. Res.* 202(2):311–319.

Picard J-C, Queyranne M (1978) The time-dependent traveling salesman problem and its application to the tardiness problem in one-machine scheduling. *Oper. Res.* 26(1):86–110.

Queyranne M, Schulz AS (1994) Polyhedral approaches to machine scheduling. Technical Report 408, Department of Mathematics, Technical University of Berlin.

Ralphs TK, Ladanyi L (2003) COIN-OR: Software tools for implementing custom solvers. Accessed June 2012, http://www.coin-or.org/Presentations/INFORMSWorkshop03.pdf.

Rask KJ, Williams MV, Parker RM, McNagny SE (1994) Obstacles predicting lack of a regular provider and delays in seeking care for patients at an urban public hospital. *J. Amer. Medical Assoc.* 271(24):1931–1933.

Ribeiro CC, Soumis F (1994) A column generation approach to the multiple-depot vehicle scheduling problem. *Oper. Res.* 42(1): 41–52.

Ribeiro GM, Laporte G (2012) An adaptive large neighborhood search heuristic for the cumulative capacitated vehicle routing problem. *Comput. Oper. Res.* 39(3):728–735.

Ryan DM, Foster BA (1981) An integer programming approach to scheduling. Wren A, ed. *Computer Scheduling of Public Transport Urban Passenger Vehicle and Crew Scheduling* (North-Holland, Amsterdam), 269–280.

Sarmiento A, Nagi R (1999) A review of integrated analysis of production-distribution systems. *IIE Trans.* 31(11):1061–1074.

Shylo OV, Prokopyev OA, Schaefer AJ (2013) Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS J. Comput.* 25(4):682–692.

Sol M (1994) Column generation techniques for pickup and delivery problems. Unpublished doctoral dissertation, Eindhoven University of Technology, Netherlands.

Solomon MM (1987) Algorithms for the vehicle routing and scheduling problems with time window constraints. *Oper. Res.* 35(2):254–265.

Toth P, Vigo D (2002) *The Vehicle Routing Problem* (Society for Industrial and Applied Mathematics, Philadelphia).

Trautsamwieser A, Hirsch P (2014) A branch-price-and-cut approach for solving the medium-term home health care planning problem. *Networks* 64(3):143–159.

van Oostrum JM, Parlevliet T, Wagelmans APM, Kazemier G (2009) Method for clustering surgical cases to allow master surgical scheduling. *INFOR* 49(4):254–260.