Non-emergency Patient Transfer Scheduling and Assignment

Foster, Travis ; Vanberkel, Peter ; Venkatadri, Uday ; van Essen, Theresia

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Non-emergency Patient Transfer Scheduling and Assignment

Travis Foster, Peter VanBerkel, Uday Venkatadri and Theresia van Essen

Abstract Emergency Medical Services organizations are responsible for providing paramedic crews, vehicles and equipment to transfer patients from one location to another in emergency and non-emergency settings. They must solve difficult scheduling and assignment problems to ensure on-time arrival of patients and the efficient use of health care resources during non-emergency operations. Ambulances can serve both emergency and non-emergency requests but are rarely available to serve non-emergency requests. Therefore, non-emergency requests are the responsibility of Patient Transfer Units. The objective of this study is to develop a mathematical model that will assign Patient Transfer Units to non-emergency patient transfer requests, design a schedule that will minimize travel costs and balance workloads and apply it to a real-world case study. This paper also proposes a framework to utilize historical patient transfer data in the scheduling process. The mathematical model provides decision support for the non-emergency patient transfer scheduling process.

T. Foster (✉) · P. VanBerkel · U. Venkatadri
Dalhousie University, 6299 South St, Halifax, NS B3H 4R2, Canada
e-mail: tfos@dal.ca

P. VanBerkel
e-mail: peter.vanberkel@dal.ca

U. Venkatadri
e-mail: uday.venkatadri@dal.ca

T. van Essen
Delft University of Technology, Van Mourik Broekmanweg 6,
2628, XE Delft, Netherlands
e-mail: j.t.vanessen@tudelft.nl

1   Introduction

This paper examines the scheduling and assignment of non-emergency patient transfer requests. Such transfers can be between any of the following locations: a special care facility, a hospital or a personal residence. Patient transfers are an important part of public safety systems as they improve patient care by allowing access to proper and continuing medical care to patients. In some jurisdictions of Canada, transfers are conducted by Emergency Medical Service (EMS) providers with paramedic crews in ambulances or similar vehicles. Increasing transfer volumes add pressure to EMS providers [10].

From a scheduling and assignment perspective, patient transfers present a unique challenge to EMS and Operations Research (OR) scientists. Patient transfers are non-emergency requests and are often scheduled according to:

- Arrival time of request (advance notice or same day).
- The requested time of pickup.
- The availability of transport vehicles in the region.
- Logistical issues surrounding the transfer such as equipment required, current vehicle location and future pickups.

Many health care organizations, including EMS providers, collect and store large amounts of historical data. This paper presents a framework for using this historical data to help a model for patient transfer scheduling and assignment at the offline operational level [6].

In Sect. 2 we describe the scheduling problem faced by Nova Scotia EMS providers. In Sect. 3 we review related literature and position our research. In Sect. 4 we present the scheduling model and the framework for integrating their data within the model. In Sect. 5 we review our results on a real-world case study. In Sect. 6 we present our conclusions and future work.

2   Problem Description

This research is motivated by Emergency Health Services (EHS). EHS is the organization that provides EMS to the province of Nova Scotia in Canada. EHS uses vehicles called Patient Transfer Units (PTUs) for a significant portion of non-emergency patient transfers. For simplicity, the term "patient transfer" will refer specifically to non-emergency patient transfers.

Patient transfer requests are phoned into EHS when the sending facility has determined a patient requires extra care for transport to a medical appointment. After a pickup time has been agreed upon between EHS and the requesting party, EHS schedules the transfer in their system. A preliminary schedule for the following day is created in the evening with all requests that were submitted in advance. Same

day requests are submitted to EHS the following day during the operation period. The schedule is adjusted to account for these same day requests. EHS also assigns a paramedic crew (who are operating a PTU) during the day-of operation. This assignment depends on several factors including the location of the patient, the location and status of the crews, future patient transfers and operator knowledge. Delaying or arriving late to a patient pickup can result in a cancelled or rescheduled appointment for the patient and a deadhead trip (a completed trip without a patient) for a PTU to the next patient pickup.

This paper focuses on scheduling advance notice patient transfer requests to minimize total travel time while using historical data to inform the model. We are also interested in introducing workload balancing features to the model. We do not examine the same day scheduling portion of the problem at this time nor do we address crew assignment; in this phase of the research, only vehicle assignment and scheduling are considered. This model is then applied to a real-world case study in Nova Scotia.

3 Related Research

Patient transfer systems are a common part of health care systems and as such, scheduling patient transfers has been studied extensively. Patient transfer systems can be modelled as dial-a-ride problems (DARP), a class of Vehicle Routing Problems (VRP) and part of the Travelling Salesman group of problems. The dial-a-ride model develops vehicle routes and schedules for *n* requests divided among *k* vehicles. We refer readers to Cordeau and Laporte [3] and Ho et al. [7] for additional information regarding the DARP.

Detti et al. [4] studied a real-world health care example of the DARP with several constraints and multiple vehicle depots. They analyze the effectiveness of their heuristics and a Mixed Integer Programming (MIP) model from real and randomly generated data. Guerriero et al. [5] solved a multi-objective DARP considering travel and patient waiting time and demonstrated computational results from their two step heuristic approach. Workload balancing features such as cost-related objective functions and constraints are explored by Matl et al. [9] in VRPs, including the DARP. They also review types of VRP workload balancing measures. Berg and Essen [1] examined scheduling vehicles for patient transfer coverage while minimizing the impact on emergency vehicles. Marković et al. [8] developed prediction models using statistical and machine learning algorithms for capacity requirements of a new dial-a-ride system. Yalçındağ et al. [11] used data driven methods to estimate travel times in home health care.

Our paper applies a DARP scheduling model including workload balancing, and data driven statistical models to estimate travel and service times from historical data to act as an efficient scheduling tool and applied to a real-world case study of a non-emergency patient transfer system.

4 Methods

In this section, we first review the approach and assumptions made when modelling the scheduling process with a DARP model. This approach is used to mimic the actual scheduling process where a preliminary schedule is created the previous day for all patient transfer requests submitted in advance. Second, we discuss the historical data and how it is used as an input to the model.

4.1 Advance Request Model

The model creates a set of routes for k PTUs and n requests while minimizing the travel time across all routes. The model assumes a single depot to act as the start and end point for every vehicle. We also assume that we have a homogeneous fleet and patients. In reality, patients do have different needs but this has little impact on the time required for a request.

Our DARP model is based off of the three-index formulation found in Cordeau [2]. However, our model includes workload balancing constraints and variable shift times for the PTU crews. We use the time windows to ensure patients arrive at their destination in a timely manner. It is formulated on a directed graph $G = (V, A)$. All vertices on the graph are represented by $i, j = (0, ..., 2n + 1)$. The pickup nodes are represented by $P = (1, ..., n)$ and the drop-off nodes are represented by $D = (n + 1, ..., 2n)$. The depot is represented by nodes 0 and $2n + 1$. These three indices make up the vertex set $V = (0, 1, ..., n, n + 1, ..., 2n, 2n + 1)$. Each request is treated as a pair $(i, n + i)$ that must be visited in order and by the same vehicle.

K represents the set of vehicles. Each vehicle $k \in K$ has a capacity of Q_k, a minimum shift start time of $Tmin_k$ and a maximum shift end time of $Tmax_k$. Each node $i \in V$ has a service time d_i and a load q_i such that $q_{n+i} = -q_i$. These values for the depot are such that $d_0 = d_{2n+1} = q_0 = q_{2n+1} = 0$. Each node has a time window $[e_i, l_i]$ where e_i and l_i are the earliest and latest times that service may begin at node i. Each arc (i, j) has an associated travel time t_{ij}. We use the parameters wb^+ and wb^- as the maximum and minimum workload, respectively, for each PTU.

The model has three types of decision variables:

- x_{ij}^k is a binary decision variable and is 1 if vehicle k will traverse the route from node i to node j and is 0 otherwise.
- u_i^k decides the service start time at node i by vehicle k.
- l_i^k decides the number of patients in vehicle k after visiting node i.

The MIP formulation is (1)–(19):

$$Min \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} t_{ij} x_{ij}^k \tag{1}$$

$$\sum_{k \in K} \sum_{j \in V} x_{ij}^k = 1 \quad \forall \, (i \in P) \tag{2}$$

$$\sum_{i \in V} x_{0i}^k = \sum_{i \in V} x_{i,2n+1}^k = 1 \quad \forall \, (k \in K) \tag{3}$$

$$\sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{n+i,j}^k = 0 \quad \forall \, (i \in P, k \in K) \tag{4}$$

$$\sum_{j \in V} x_{ji}^k - \sum_{j \in V} x_{ij}^k = 0 \quad \forall \, (i \in P \cup D, k \in K) \tag{5}$$

Equation (1) is the objective function where we minimize the total travel time of the routes. Constraints (2) ensures each request is served once. Constraints (3) ensures that every vehicle route begins and ends at the depot. Constraints (4) ensures the pickup and delivery nodes of a request are served by the same vehicle. Constraints (5) certifies that the vehicle that enters a node will also depart from that node.

$$u_j^k \geq u_i^k + d_i + t_{ij} - M(1 - x_{ij}^k) \quad \forall \, (i, j \in V, k \in K) \tag{6}$$

$$u_{i+n}^k \geq u_i^k + d_i + t_{ij} - M(1 - x_{i,i+n}^k) \quad \forall \, (i \in P, k \in K) \tag{7}$$

$$u_{2n+1}^k \leq Tmax_k \quad \forall \, (k \in K) \tag{8}$$

$$u_0^k \geq Tmin_k \quad \forall \, (k \in K) \tag{9}$$

$$e_i \leq u_i^k \leq l_i \quad \forall \, (i \in V, k \in K) \tag{10}$$

Constraints (6) and (7) guarantee that service time is consistent among every node, and that the service time at node $i + n$ does not begin until after the service at node i is completed. Constraints (8) and (9) uphold vehicle shift start and end times. Constraints (10) ensures that a request is completed during it's time window.

$$l_j^k \geq l_i^k + q_j - M(1 - x_{ij}^k) \quad \forall \, (i, j \in V, k \in K) \tag{11}$$

$$l_j^k \leq l_i^k + q_j + M(1 - x_{ij}^k) \quad \forall \, (i, j \in V, k \in K) \tag{12}$$

$$l_i^k \leq Q_k \quad \forall \, (i \in V, k \in K) \tag{13}$$

$$l_i^k \leq Q_k + q_i \quad \forall \, (i \in V, k \in K) \tag{14}$$

$$l_i^k \geq 0 \qquad \forall \, (i \in V, k \in K) \tag{15}$$

$$l_i^k \geq q_i \qquad \forall \, (i \in V, k \in K) \tag{16}$$

$$\sum_{i \in K} \sum_{j \in V} (t_{ij} + d_i) x_{ij}^k \leq wb^+ \qquad \forall \, (k \in K) \tag{17}$$

$$\sum_{i \in K} \sum_{j \in V} (t_{ij} + d_i) x_{ij}^k \geq wb^- \qquad \forall \, (k \in K) \tag{18}$$

$$x_{ij}^k \in (0, 1) \qquad \forall \, (i \in V, j \in V, k \in K) \tag{19}$$

Constraints (11)–(16) ensures consistency for the passenger load in every vehicle. Inequalities (17) and (18) are the workload balancing constraints. We measure workload as the sum of the travel time and service time along a PTU's route.

4.2 Model and Data Framework

EHS captures data such as facility visited, length of time traveled, time spent with the patient and much more. We use historical patient transfer data to help estimate travel times, service times and generate samples to test the model. This process is illustrated in Fig. 1. The DARP model receives the service and travel time estimates from the patient transfer data in addition to the PTU shift schedules and q_i. The output from this model is the route set (x_{ij}^k), service time start times (u_i^k) and PTU load (l_i^k) for each node.

The patient transfer process can be broken into four stages as shown in Fig. 2. A PTU crew is sent to pick up a patient. They spend t_{ji} minutes travelling to pickup node i. They spend service time d_i at the pickup node. Once the patient is loaded into the PTU, the crew spends $t_{i,i+n}$ minutes travelling to the delivery node where they spend d_{i+n} minutes. When they have transferred care to the receiving facility, the patient transfer is complete. This is a special case where no additional patients are picked up once a patient transfer has begun, and $Q_k = 1$. We focus on this case as the majority of PTUs have a capacity of one. These model parameters are estimated from the patient transfer data. We first look at characteristics of patient transfer demand to determine how we will determine these parameters.

We check patient transfer requests by day of week and by time of day. The day of week analysis found the majority of patient transfer requests occur Monday to Friday. Focusing only on those days, Fig. 3 shows requests by time of day. Requests begin at 6 a.m. We see that requests taper off by 6 p.m.; therefore, that will be the end of the model's time horizon. We also note that advance notice requests are more likely to take place in the morning while same day requests take place in the afternoon.

With these patient transfer trips captured in the data, we use the historical average to estimate the expected time to travel from node i to node j. These estimates are used in the model as parameter t_{ij}.
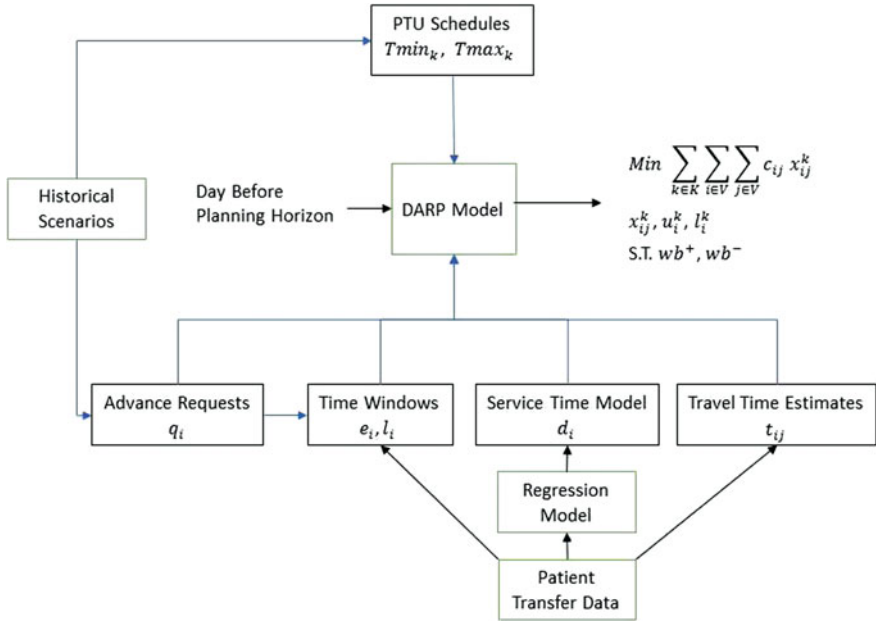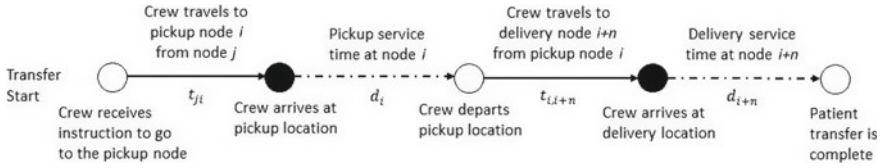
Fig. 1 Big data and OR framework

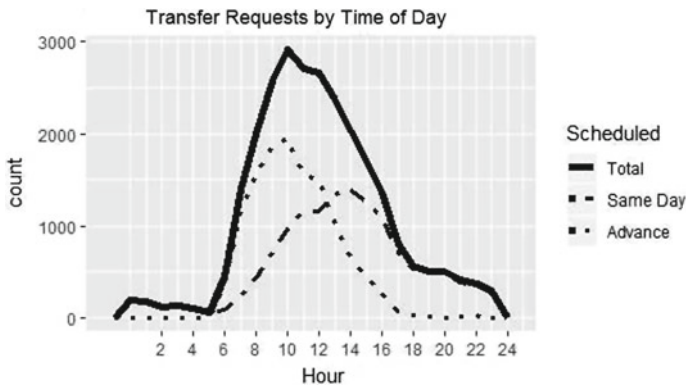**Fig. 2** Time stages in the patient transfer process

Fig. 3 Patient transfer requests by time of day

In this paper, we define two types of service time. The pickup service time is the time that paramedics spend with the patient at the pickup location before transit begins. Delivery service time is the time spent at the delivery location before completing the transfer. These are shown in Fig. 2 as d_i and d_{i+n} respectively. Service times are captured in the patient transfer data. However, numerous factors can influence the service time, including the facility type, equipment required and whether the patient is ready when the paramedics arrive. Using the patient transfer data, multi-linear regression models were created to estimate the expected pickup and delivery service times for requests. These models use common information that are included in every patient transfer record. The output of the regression model is used as input to the DARP model as d_i. We use day of week, time of day, facility, whether the patient is bariatric, whether special equipment is required and whether the request was advance notice or same day.

5   Results

EHS typically receives 15–45 advance requests per day in the city of Halifax. We test the model on randomly selected actual problems from the historical data. We can use the actual service times or generate predicted service times from the regression model. The computational times are plotted on a logarithmic scale in Fig. 4 shows that computation time (plotted on a logarithmic scale) increases with the number of requests. However, every problem tested solved in under one hour of run time using the Gurobi Optimizer. Since EHS would run the model overnight, we have shown that the model can be solved in a short time frame where the results will be of use to EHS. Table 1 shows a summary of our tests including the maximum, minimum and mean number of requests and time (measured in seconds) required to solve the DARP.

We compare our total travel time against the actual recorded travel time from the data for advance requests. However, we do not have the original schedules with only the advance requests. Instead, we have the final schedules with advance and same day requests. Therefore, we select the 9 days where same day requests make up the
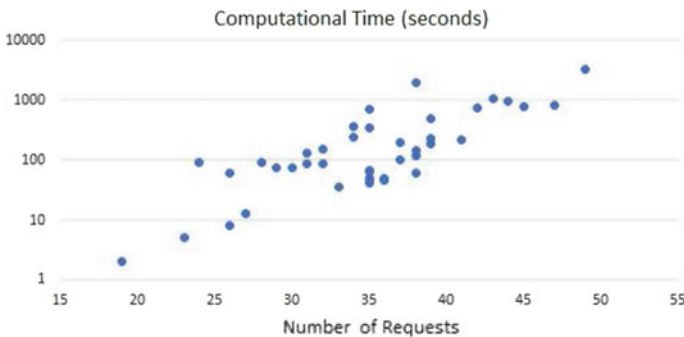

Fig. 4   DARP model computational time

Table 1 Computational results summary

	Requests	Computational time (s)
Max	49	3281
Min	19	2
Mean	33	319

Table 2 Advance requests study

Advance requests	Model travel time	Actual travel time	Actual completed requests	Travel time difference	Same day requests
24	645	747	22	102	8
26	771	833	26	62	8
32	847	1244	32	397	9
34	744	1048	34	304	11
35	768	993	35	225	10
37	890	933	36	43	8
39	920	1211	37	291	10
39	740	1191	38	451	11
43	1289	1271	41	−18	12

smallest percentage of all requests. While this is not a completely fair comparison, focusing on the days with the least impact from the same day requests is as close as we can get. The results are found in Table 2.

For these DARP instances, we see a total reduction in travel time of 1857 min, or 206 min per day. This is a 19.6% improvement on the actual travel time spent on the advance requests. The PTUs also only completed 301 advance requests versus the 309 that were planned for as 8 requests were cancelled. Cancellations can happen for a number of reasons, but it is most likely that this happens when the patient is not ready for transport. The average travel time spent per request as per the model is 24.6 min versus 31.5 min in the data, an improvement of 21.7%.

6　Conclusions and Future Work

In this paper, we studied non-emergency patient transfers and applicable operations research models. We modelled the advance request system using a Dial-A-Ride Problem approach, and developed a model that will schedule and route PTUs for a set of known requests. We also develop statistical models for delivery service times from historical patient transfer data. The historical data is used to inform model parameters as described in Fig. 1. We present this framework as a method to

incorporate Big Data and Analytics into non-emergency patient transfer scheduling. We test our model on historical data and find the model is computationally feasible for problem sizes we are interested in. Finally, we compare our model's output to the actual travel times for days where impact from same day requests is minimal and find travel time improvements of approximately 20%.

Future work involving this problem could include incorporating the same day requests into a model to dynamically update the routes as new demand arrives in real time and investigating the stochastic nature of travel and service times.

References

1. Berg, P., Essen, T.: Scheduling non-urgent patient transportation while maximizing emergency coverage. Transp. Sci., Accepted (2018)
2. Cordeau, J.F.: A branch-and-cut algorithm for the dial-a-ride problem. Oper. Res. **54**, 573–586 (2006)
3. Cordeau, J.F., Laporte, G.: The dial-a-ride problem: models and algorithms. Ann. Oper. Res. **153**, 29–46 (2007)
4. Detti, P., Papalini, F., Manrique de Lara, G.Z.: A multi-depot dial-a-ride problem with heterogeneous vehicles and compatibility constraints in healthcare. Omega **70**, 1–14 (2016)
5. Guerriero, F., Pezzella, F., Pisacane, O., et al.: Multi-objective optimization in dial-a-ride public transportation. Transp. Res. Procedia **3**, 299–308 (2014)
6. Hans, E., van Houdenhoven, M., Hulshof, P.J. H.: A framework for healthcare planning and control. In: Hall, R. (ed.) Handbook of Healthcare System Scheduling. International Series in Operations Research & Management Science, vol. 168. Springer, Boston (2012)
7. Ho, S., Szeto, W.Y., Kuo, Y., et al.: A survey of dial-a-ride problems: literature review and recent developments. Transp. Res. Part B **111**, 395–421 (2018)
8. Marković, N., Myungseob, K., Schonfeld, P.: Statistical and machine learning approach for planning dial-a-ride systems. Transp. Res. Part A **89**, 41–55 (2016)
9. Matl, P., Hartl, F., Vidal, T.: Workload equity in vehicle routing problems: a survey and analysis. Transp. Sci. (2017). https://doi.org/10.1287/trsc.2017.0744
10. Robinson, V., Goel, V., MacDonald, R.D., Manuel, D.: Inter-facility patient transfers in ontario: do you know what your local ambulance is being used for? Healthc. Policy (2009). https://doi.org/10.12927/hcpol.2009.20478
11. Yalçındağ, S., Matta, A., Şahin, E., Shanthikumar, J.G.: The patient assignment problem in home health care: using a data-driven method to estimate the travel times of care givers. Flex. Serv. Manuf. J. (2016). https://doi.org/10.1007/s10696-015-9222-6