Scheduling and routing ambulances that provide inter-facility patient transfers

Laleh Haerian Ardekani, Dan Haight, Armann Ingolfsson*,
Mo Salama, and Matt Stanton
University of Alberta School of Business

August 28, 2014

Abstract

In Edmonton and Calgary, Canada, a specialized ambulance fleet transfers patients between hospitals. We use a vehicle routing heuristic to schedule transfers that are known a day in advance, and a similar heuristic to accommodate emergent transfers in real time. We use simulation experiments to compare the performance of this approach to historical schedules and find that improvements appear to be possible on all performance metrics that we investigated, ranging from 60% for dead-heading to 3% for travel times.

Keywords: Inter-Facility Patient Transfer, The Vehicle Routing Problem, Ambulance Routing, Ambulance Scheduling

1 Introduction

The primary task of emergency medical services (EMS) systems is to provide out-of-hospital medical care and transport by ambulance to a healthcare facility for further care, for patients that experience medical emergencies. Ambulances are also used, however, for transporting patients from one healthcare facility to another—for example, to transport a patient from a long-term care facility to and from a hospital for dialysis treatment. Such interfacility patient transfers (IFT) can either be done using a shared EMS-IFT fleet or using a dedicated IFT fleet. From an operational point of view, an essential difference between EMS and

*Corresponding author: armann.ingolfsson@ualberta.ca

IFT is that IFT trips can be scheduled whereas EMS calls must be responded to as soon as possible. The planning and management of EMS systems has received increasing attention in recent years from operations researchers (see (Ingolfsson, 2013)) but planning methods for EMS systems do not apply directly to the operation of IFT systems. In this paper, we develop methods to schedule and route IFT trips and we use simulation, calibrated with data from Edmonton and Calgary, Canada, to test the effectiveness of our methods.

Alberta Health Services is responsible for both EMS and IFT in Calgary, Edmonton, and other jurisdictions in the province of Alberta. IFT uses a dedicated and specialized fleet in both Calgary and Edmonton, but some IFT requests are handled by EMS ambulances, when all IFT ambulances are busy. Some other provinces in Canada also use dedicated IFT fleets, for example, Ontario (MacDonald et al., 2004) and British Columbia (British Columbia Patient Transfer, 2014). In contrast, the SOS Alarm Company in Sweden (SOS Alarm, 2014) employs a joint EMS-IFT fleet (Andersson et al., 2004). A recent study on IFT services in Ontario (Robinson et al., 2009) shows that one inter-facility transfer takes place per year for every 37 people, on average. The corresponding frequency is 23% higher in Edmonton (one transfer per year per 30 people) and 29% lower in Calgary (one transfer per year per 52 people). IFT transfer volumes differ between regions for various reasons, the number of healthcare facilities in the region and their degree of specialization, for instance.

The IFT fleets in Edmonton and Calgary respond to both planned (advance) and emergent patient transfer requests. The planned requests are usually made the previous day or earlier, whereas the emergent ones are made closer to the requested pickup times. The percentage of emergent requests is 61% in Edmonton and 75% in Calgary. Based on the patient's health condition, each IFT request is assigned a type code of either red, yellow, green, or blue, in decreasing order of request priority. These type codes dictate the length of patient's pickup time window. IFT dispatchers try to schedule transfers such that patients are picked up within their specified time windows. However, the the arrivals of higher priority requests frequently require re-routing, which can cause pickup delays, especially for lower priority requests.

A patient transfer delay can have a cascading effect: The transfer patient could be occupying a bed that is needed for another patient currently in the emergency department, for example. Or, if the transfer patient is scheduled for a diagnostic imaging test, a transfer delay could lead to forced idleness of expensive imaging equipment and staff. At the time when we complete this project, the scheduling and routing of the IFT fleet in Edmonton and Calgary is done manually. Under the manual system, significant time window violations, tardiness, and also frequent deployment of EMS fleet for IFT requests were observed. Given the valuable

and limited ambulatory resources, and considering the sensitivity of the patient transfer operations, it is of great importance to schedule and handle these resources efficiently.

We make the following contributions: We develop (1) a heuristic for scheduling advance IFT requests, (2) a real-time heuristic for accommodating and scheduling emergent requests, and we (3) use simulation experiments to compare the performance of our heuristics to the existing IFT routing system. Our experiments predict that using our heuristics could greatly reduce tardiness, travel times, and deadheading, while simultaneously reducing the number units needed to provide the service.

In Section 2, we review related literature on ambulance scheduling and routing problems. Section 3 describes IFT operations in Edmonton and Calgary. Section 4 formally defines the scheduling problem. Section 5 outlines our scheduling and routing heuristic. Section 6 reports results from simulation experiments that we used to compare the performance of the heuristic to current approaches. Section 7 concludes.

2 Literature Review

The problem of scheduling and routing an IFT fleet can be viewed as a generalization of the vehicle routing problem (VRP) (Toth and Vigo, 2002), or as a generalization of the dial-a-ride problem (DARP) (Cordeau and Laporte, 2007). The IFT scheduling and routing problem has constraints on pickup and dropoff locations, and extra constraints on transportation time windows, vehicle station depots, and crew shift schedules. The VRP, and therefore its generalization the DARP, are NP-hard problems (Lenstra and Rinnooy Kan, 1981). Using integer programming technology to find feasible, let alone optimal, solutions to VRP and DARP has proven to be difficult even for moderate-sized problem instances. These difficulties have motivated researchers to develop heuristics and metaheuristics to find near-optimal solutions to the VRP, DARP, and related problems. In considering the available heuristics, we needed to keep in mind the requirements for a heuristic to be used to schedule IFT calls in Edmonton and Calgary. On average, in Edmonton and Calgary, there are 108 and 55 calls to be scheduled per day, respectively. There is one depot in Edmonton, for a fleet of 38 vehicles, and there are two depots in Calgary, for a fleet of size 34. Furthermore, we need heuristics for both advance and emergent scheduling.

Tabu search has proven to be one of the most successful solution methods for variations of VRP and DARP (Cordeau and Laporte, 2005). A promising strategy for solving IFT scheduling and routing problems is to start with a flexible VRP or DARP heuristic and extend it to incorporate IFT-specific constraints. We follow this strategy and therefore we focus on review of past literature on the following topics: (1) Heuristics

to solve the VRP, (2) Heuristics to solve the DARP, and (3) Heuristics for advance and emergent IFT patient transfers.

Potvin et al. (1996) developed a tabu search heuristic for the solution of VRP with time windows (VRPTW). They tested this heuristic on randomly generated instances of 100 customers, with Euclidean distances as travel times. Cordeau et al. (2001) designed and later improved (Cordeau et al., 2004) a VRP tabu search heuristic, that can be extended to VRP with time windows and multiple depots. They tested this heuristic on instances with up to 5 vehicles and 6 depots. The fleet is assumed to be homogeneous with limits on capacity and travel duration.

Coslovich et al. (2006) studied unexpected customer ride requests in a dynamic DARP, where unplanned customer requests take place at vehicle's previous dropoff location. Jørgensen et al. (2007) designed and tested a genetic algorithm for solving a DARP with time windows. They solved instances with 24 to 144 customers and a single depot. Parragh et al. (2009) developed a method to solve DARP with time windows, using variable neighborhood search followed by path re-linking. They modeled travel times using Euclidean distances, and tested their model on randomly generated instances with a single depot, with 16 to 96 requests, and 2 to 8 vehicles.

IFT operators typically respond to both advance and emergent demand. We require real-time dynamic routing algorithms to accommodate emergent transportation requests in the routing schedule. Madsen et al. (1995) developed an insertion algorithm to solve a dynamic DARP for transporting clients in the city of Copenhagen. This algorithm was designed to be used in an on-line scheduling system. It was tested on instances with 4 vehicles (with different types of seats) and 30 requests. Attanasio et al. (2004) developed parallel heuristics for the dynamic DARP, based on a tabu search designed for the static DARP. Melachrinoudis et al. (2007) solved a static DARP with soft time windows for transporting patients between health care organizations, using tabu search. They tested their heuristic on instances with up to 50 clients and 15 depots, with randomly generated locations. Beaudry et al. (2010) used tabu search for a dynamic DARP. They tested their heuristic on data from patient transfers to a German hospital, where transportation demand is not known in advance.

Ho and Haugland (2011) studied a version of DARP with probabilistic patient requests. They used tabu search on instances with up to 144 nodes and found optimal solution in all instances. Kergosien et al. (2011) used tabu search and adaptive memory procedures to schedule IFT transfers between care units of a French hospital complex. In their setting, 70% of transfer requests are emergent, and there are extra constraints on request priorities, vehicle types, and requirements on vehicle disinfection after transferring certain types

of patients. Häll et al. (2012) developed a modeling system for simulating DARP. This system schedules and simulates operations of a dial-a-ride service with dynamic demand, multiple depots, vehicle capacities, and vehicle schedules. Their DARP simulation model allows for service level evaluation based on pickup time-window violations and limits on the duration of patient transportation. Two heuristics are used to insert emerging demand into schedules, and to re-optimize the current schedule. Khouadjia et al. (2013) studied and compared state-of-the-art heuristics and metaheuristics for variations of the dynamic VRP. They compared the performance of various heuristics and metaheuristics on VRP benchmark data, in terms of accuracy, stability, and the ability of the algorithm to react quickly to changes.

3 Edmonton and Calgary IFT Operations

The cities of Edmonton and Calgary are located in the province of Alberta, Canada, with a population of around one million each. The IFT fleets in Edmonton and Calgary are composed of 38 and 34 dedicated ambulance units, respectively. The IFT crews work 8 to 12-hour shifts, each of which includes two 15-minute coffee breaks and a 30-minute meal break. The shifts are scheduled to match the supply of units to the demand of IFT requests.

Health care facilities submit requests to transport patients from their location to other facilities. A IFT request specifies the pickup and dropoff locations and the desired pickup and dropoff times. The IFT dispatcher categorizes the call and the category determines the pickup time window duration: 20-minutes for red, 60 minutes for yellow, 3 hours for green, and 24 hours for blue.

We use data from January to September 2011 as our sample to illustrate various aspects of the IFT operations in Edmonton and Calgary. Figure 1 shows the average hourly number of advance and emergent calls, for both cities. We see that the weekday and weekend patterns differ considerably. Each day, the advance requests are scheduled first, and then the emergent requests are scheduled as they occur throughout the day. When there are no available IFT units, emergent requests are passed to the EMS fleet. Table 1 shows the distribution of IFT requests by patient code. Table 2 shows the number of IFT requests handled by EMS fleet, by patient code. Overall, the proportion of IFT requests handled by EMS was 2.6% in Edmonton and 17.2% in Calgary. Figures 2-3 show the average number of scheduled and busy IFT units for the two cities.

An IFT call is composed of seven time intervals: 1) Call evaluation time, 2) Unit dispatch time, 3) Unit chute time, 4) Travel time to patient site, 5) Crew time at pickup location, 6) Transport time to destination,

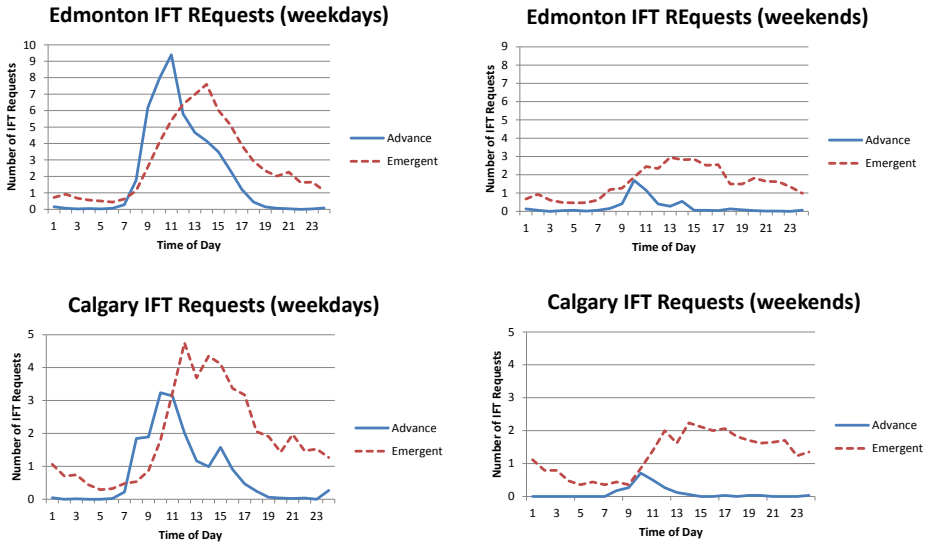

Figure 1: Average Number of Advance and Emergent Calls

	Edmonton		Calgary	
	Emergent	Advance	Emergent	Advance
Blue	67%	33%	74%	26%
Green	45%	55%	49%	51%
Yellow	96%	4%	93%	7%
Red	99%	1%	100%	0%
Total	61%	39%	75%	25%

Table 1: Distribution of IFT Requests Types

City	Requests	Blue	Green	Yellow	Red	Total
Edmonton	EMS-handled	48 (0.6%)	159 (1.2%)	243 (7.6%)	215 (45.9%)	2.6%
	Total	8,378	13,253	3,176	468	
Calgary	EMS-handled	33 (3.0%)	81 (4.9%)	577 (25.6%)	215 (66.6%)	17.2%
	Total	1,112	1,658	2,174	323	

Table 2: Number of IFT Requests Handled by the EMS Fleet from December 2010 to September 2011

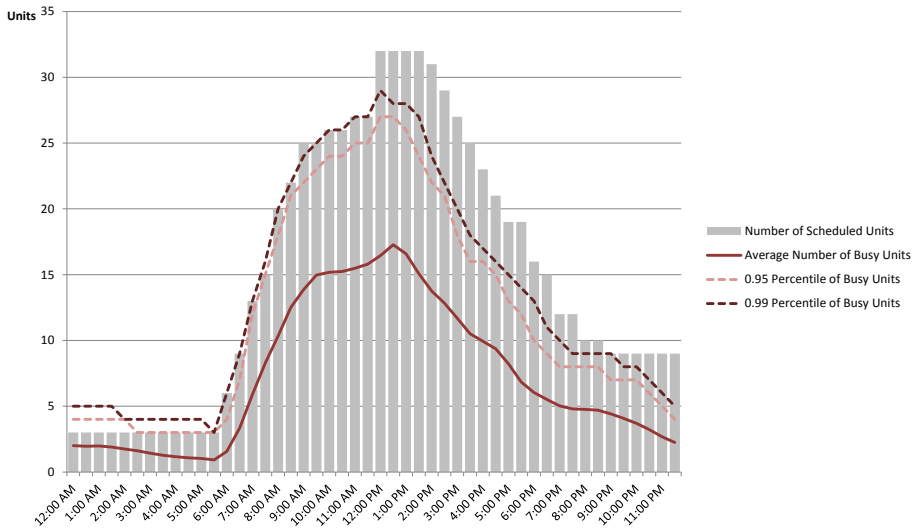

Figure 2: Average Number of Scheduled and Busy Units, Edmonton

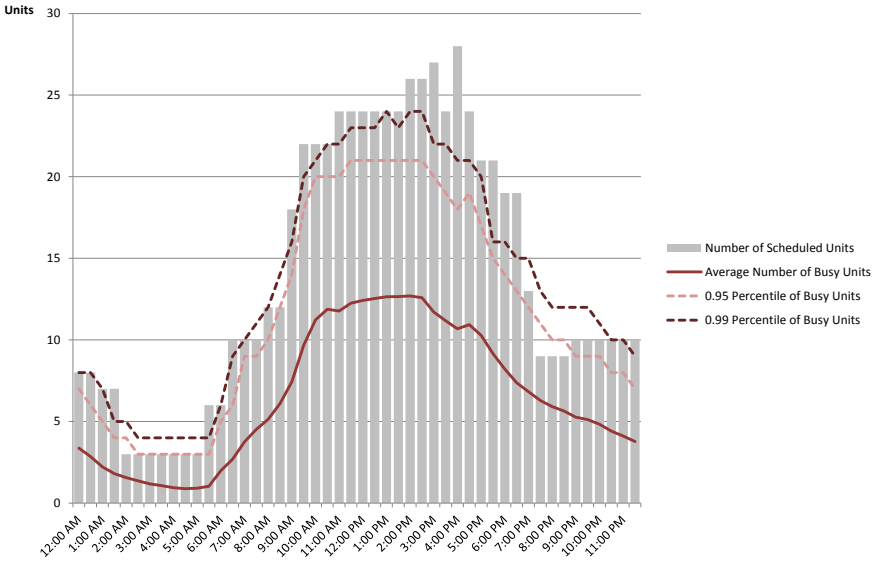

Figure 3: Average Number of Scheduled and Busy Units, Calgary

		Blue	Green	Yellow	Red
Edmonton	Weekday	0.46	0.43	0.53	0.83
	Weekend	0.42	0.41	0.49	0.83
Calgary	Weekday	0.54	0.51	0.40	0.46
	Weekend	0.55	0.41	0.35	0.42

Table 3: Average crew times at pickup locations (hours)

		Blue	Green	Yellow	Red
Edmonton	Weekday	0.51	0.53	1.00	1.12
	Weekend	0.49	0.77	1.02	1.14
Calgary	Weekday	0.34	0.64	0.56	0.95
	Weekend	0.44	0.65	0.60	0.97

Table 4: Average crew times at dropoff locations (hours)

and 7) Crew time at dropoff location. Advance calls have longer call evaluation and dispatch times than emergent calls. Average crew times at pickup and dropoff locations differ by patient code, as shown in Tables 3-4.

4 The IFT Scheduling and Routing Problem

The IFT scheduling and routing problem is to find an optimal routing schedule for the IFT units, given their availabilities and inter-facility transportation requests. This problem is a generalization of the multi-depot VRPTW, with pickups and deliveries, where vehicle availability is constrained by work shifts. Each patient has an *origin* node (pickup location) and a *destination* node (dropoff location). The origin node has a specified time window. We assume that if the patient is picked up within this time window, then the patient dropoff is guaranteed to be on time. We assume that the IFT vehicles have a capacity of one patient, and therefore a unit must drop off one patient before picking up the next patient.

We view the following as hard constraints:

- Each vehicle is stationed in a depot, where it starts and finishes its route.
- A crew takes two 20-minute coffee breaks and one 30-minute meal break during its shift.
- Each vehicle has a capacity of one patient.

We view the follow as soft constraints, which can be violated but with a penalty:

- A crew does not work past the end of its shift.
- A patient must be picked up within her pickup time window.

5 Solution Approach

The method we propose begins by generating a solution for the advance requests and subsequently inserts emergent requests as they occur. The method has three algorithmic components.

Algorithm 1 : A greedy heuristic based on the Clarke and Wright (1964) savings algorithm, to generate an initial routing schedule for the advance requests.

Algorithm 2 : A tabu search heuristic based on Cordeau et al. (2001)'s unified tabu search heuristic, to improve the initial solution.

Algorithm 3 : A real-time heuristic for inserting emergent requests in the existing solution where they best fit.

Assume that there are m_a advance requests and m_e emergent requests to be scheduled. Label the requests in ascending order of request arrival. Let $V = \{1, ..., n\}$ be the set of IFT units. We express the schedule S as an ordered set of vehicle routes $S = (s_1, ..., s_n)$, where s_v is the daily schedule for vehicle $v \in V$. We denote the total scheduled travel time for vehicle $v \in V$ as $t(s_v)$, its total tardiness as $w(s_v)$, and its total overtime work hours as $o(s_v)$. We evaluate the routing solutions using the function $f(S) = \sum_{v \in V} t(s_v)$ $+\alpha \sum_{v \in V} w(s_v)$ $+\beta \sum_{v \in V} o(s_v)$, with α and β being positive parameters.

We use Algorithm 1 to generate an initial solution. This algorithm generates an initial solutionby assigning an empty route to each vehicle, starting from the vehicle's depot at the beginning of the vehicle's shift, and end in the same place at the end of the shift.

We use the tabu search heuristic shown in Algorithm 2 to improve the initial solution from Algorithm 1. We define the following tabu search neighborhood $N(S)$ of a solution S. Select a request i that is part of route s_k in S and select another route, $s_{k'}V$, $k' \neq k$. Move request i from s_k to $s_{k'}$ by inserting i's destination and origin nodes between two of the requests already in route $s_{k'}$. The neighborhood $N(S)$ is the set of all solutions that can be generated in this way—by varying the request i, the index of the route k' that request i is moved to, and the position in $s_{k'}$ where i is inserted.

In each iteration, we generate the neighborhood $N(S)$ of the current solution S and we select the best non-tabu solution $\hat{S}$ in the neighborhood to become the new current solution. We add the selected request and the vehicle route to which the request is moved to the tabu list. We run the heuristic for η iterations and return the best solution found, S^*. We update the parameters α and β in the same fashion as in Cordeau et al. (2001), in order to make it likely that the soft constraints (time-window and shift-hours constraints)

Algorithm 1 Generate an Initial IFT Solution

```
Set A ← {1, ..., m_a}.
Initialize solution S with an empty route s_v, starting and ending at the depot for vehicle v, for each vehicle
v ∈ V.
Set S* ← S.
while A ≠ ∅ do
  Set i ← a randomly chosen request from A.
  Set f(S*) ← ∞.
  for each vehicle v ∈ V do
    Generate Ŝ by inserting request i in the route s_v that minimizes f(Ŝ).
    if f(Ŝ) < f(S*) then
      Set S* ← Ŝ, and f(S*) ← f(Ŝ).
    end if
  end for
  Remove request i from A.
  Set S ← S*.
end while
return S*
```

are satisfied by the final solution. Specifically, if $\hat{S}$ violates one or more time window constraint, then α is set to $\alpha(1+\delta)$; otherwise it is set to $\alpha/(1+\delta)$, where δ is a positive parameter. We use the same method to update the parameter β for shift violation constraints. Similar Cordeau et al. (2001)'s unified tabu search, we use the following search diversification strategy. Let ρ_{ik} be the number of times request i has been inserted into s_k so far. If $f(\hat{S}) \geq f(S^*)$, then add the penalty $p(\hat{S}) = \lambda t(\hat{S})\sqrt{nm}\sum_{i \in s_k, k \in V} \rho_{ik}$ to $f(\hat{S})$, where λ is a positive parameter and m is the number of requests provided to Algorithm 2 ($m_a \leq m \leq m_a + m_e$). If $f(\hat{S}) < f(S^*)$, then we set the penalty value $p(\hat{S})$ to zero. The pseudo code for Algorithm 2 is as follows.

Algorithm 2 IFT Scheduling Tabu Search

```
Set α ← 1, and β ← 1.
Set S* ← the solution from Algorithm 1, and set the current solution S ← S*, and set f(S*) ← f(S).
for i = 1, ..., η do
  Generate N(S).
  Find a non-tabu solution Ŝ ∈ N(S) with the minimal value for f(Ŝ) + p(Ŝ).
  if f(Ŝ) < f(S*) then
    S* ← Ŝ, and f(S*) ← f(Ŝ).
  end if
  Update α and β.
  Update ρ_ik
  Update the tabu list.
  Set S ← Ŝ.
end for
return S*
```

We use Algorithms 1 and 2 to generate a solution that includes only the advance requests. We insert the emergent requests as they occur, using Algorithm 3. We insert emergent requests so as to minimize the

total tardiness of the solution. After $r > 0$ emergent-request insertions, we rerun Algorithm 2 to improve the solution.

Algorithm 3 Insert Emergent Request e in the IFT Schedule

```
Set S ← current IFT schedule.
Set c(S*) = ∞.
for each vehicle v ∈ V do
   Find Ŝ by inserting request e in s_v ∈ S, with v ∈ V chosen to minimize w(Ŝ).
   if w(Ŝ) < w(S*) then
      S* ← Ŝ, c(S*) ← w(Ŝ).
   end if
end for
if solution has not improved in the last r calls to Algorithm 3 then
   Run Algorithm 2 to improve S*
end if
return S*
```

6 Simulation Results

We used 2011 data on IFT operations in Edmonton and Calgary for 40 days in August and September. The data for each request includes the call time, pickup time, origin and destination locations, patient type code, and travel log of the deployed unit. We used Algorithms 1-3 to generate routing schedules for the 40 days. We used a simulation model to compare the actual and proposed schedules.

6.1 The IFT Simulation Parameters and Assumptions

We aggregated demand in each city to 1,000 demand nodes. We used the travel log data and shortest road network distances to estimate travel time distributions, using the method from Budge et al. (2010). The parameter estimates for Edmonton(Calgary) were:

- Cruising speed $= v_c = 1.603(1.501)$ km/min.
- Acceleration $= a = 0.096(0.141)$ km/min^2.
- Coefficient of variation parameters: $b_0 = 8.366(9.617), b_1 = 0.000(0.538), b_2 = 0.023(0.011)$.

We also included a pre-travel interval of 1.5 minutes for each travel segment, to account for time used for evaluation, dispatch, and chute times.

	Blue		Green		Yellow		Red	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Edmonton	26.5	27.4	25.1	73.4	26.1	18.2	34.4	30.4
Calgary	35.3	16.4	29.0	15.8	25.7	17.9	29.1	31.9

Table 5: Lognormal distribution parameters for scene times at pickup locations (minutes)

	Blue		Green		Yellow		Red	
	Mean	stdev	Mean	stdev	Mean	stdev	Mean	stdev
Edmonton	26.4	21.5	27.2	25.0	45.3	40.9	49.8	35.4
Calgary	22.2	17.3	35.1	22.2	35.5	27.9	59.4	57.6

Table 6: Lognormal distribution parameters for scene times at dropoff locations (minutes)

In Algorithms 1 and 3, we estimated crew times at pickup and dropoff locations, using average values, shown in Tables 3 and 4. Similar to Ingolfsson et al. (2008), we found scene times to be well-approximated by lognormal distributions. Tables 5 and 6 show lognormal parameters for crew times during weekdays.

We made the following assumptions for the simulation:

- IFT units only perform IFT transfers (that is, they do not respond to EMS calls).
- The shift for each unit, including scheduled breaks, is given for each planning day. Extra unscheduled units that are deployed for a day, are included in the simulation as well, to mirror reality.
- Each crew along with its unit, is continuously available throughout the time that it is scheduled to be on duty. That is, we ignore unplanned breakdowns and crew absences.
- Each unit has a capacity of one patient.
- Every unit begins and ends its route at its depot.
- A unit remains at a dropoff location until it leaves for the next transfer, except that the unit returns to its depot after completing the last transfer of the day.
- When a unit is scheduled to take a break, it completes the request currently in progress, if any, and then takes the break at its current location.
- Break times and patient transfers are not preempted for the sake of higher priority emergent requests.

We set the tabu search parameters as follows: $\theta = [7.5 \log(n)]$ (resulting in tabu tenure values from 9 to 12 for Edmonton and from 7 to 11 for Calgary), $\delta = 0.5$, and $\lambda = 0.015$.

In the simulation model, travel times and service time intervals are randomly generated as each travel or service activity is about to start. We assume that when simulated units start an activity, they are not

allowed to preempt their tasks for the sake of higher priority requests. Also, we assume that rescheduling upcoming clients of tardy units is not allowed while they are still in the middle of an activity. We allow rescheduling of planned clients only after a previous client on the same route was picked up late.

The simulation model uses Algorithms 1-2 to generate an initial solution. As the simulation progresses and emergent requests arrive, the model calls Algorithm 3 to schedule these requests. The simulation model might reschedule some routes in case a unit arrives late on a client site and we anticipate late pickups for succeeding clients on the same route.

6.2 Simulation Experiments

We compared the performance of simulated units with that of deployed units based on real instances on the same days. Table 7 shows by how much each IFT metric has decreased for the simulation results compared to actual performance from historical data. Table 8 shows average tardiness and the probability of tardiness for various code types of both simulation experiments and actual values. Figures 4 and 5 show cumulative probability for wait times (i.e. actual pickup times minus requested pickup times), based on different code types for Edmonton and Calgary.

Simulation results suggest that for red and yellow code types, the probability and duration of wait times beyond time windows (i.e. tardiness), is smaller for simulation experiments than that of the actual IFT performance. For wait times within time windows, as we get closer to the end of time windows, the probability of facing longer wait times are smaller for simulation experiments than that of the actual results. For green code type calls, the probability of tardiness for around an hour after the time window is higher for the simulation experiment than actual results. Probabilities of waiting beyond time windows for calls of type blue, are equal to zero for both actual and simulated calls. However, probability of waiting within time windows for simulated calls are higher than that of the original.

Figure 5 shows a very similar pattern for the experiment results in Calgary, except that the probability of tardiness for calls of type green is close to zero, compared to Edmonton.

Simulation results suggest that significant improvements can be achieved through using the IFT scheduling and routing algorithm. The results show significant decrease in the frequency of patient time window violations as well as total tardiness, on the simulated days compared to actual. Total number of deployed simulated IFT units and their travel times are smaller than the actual values.

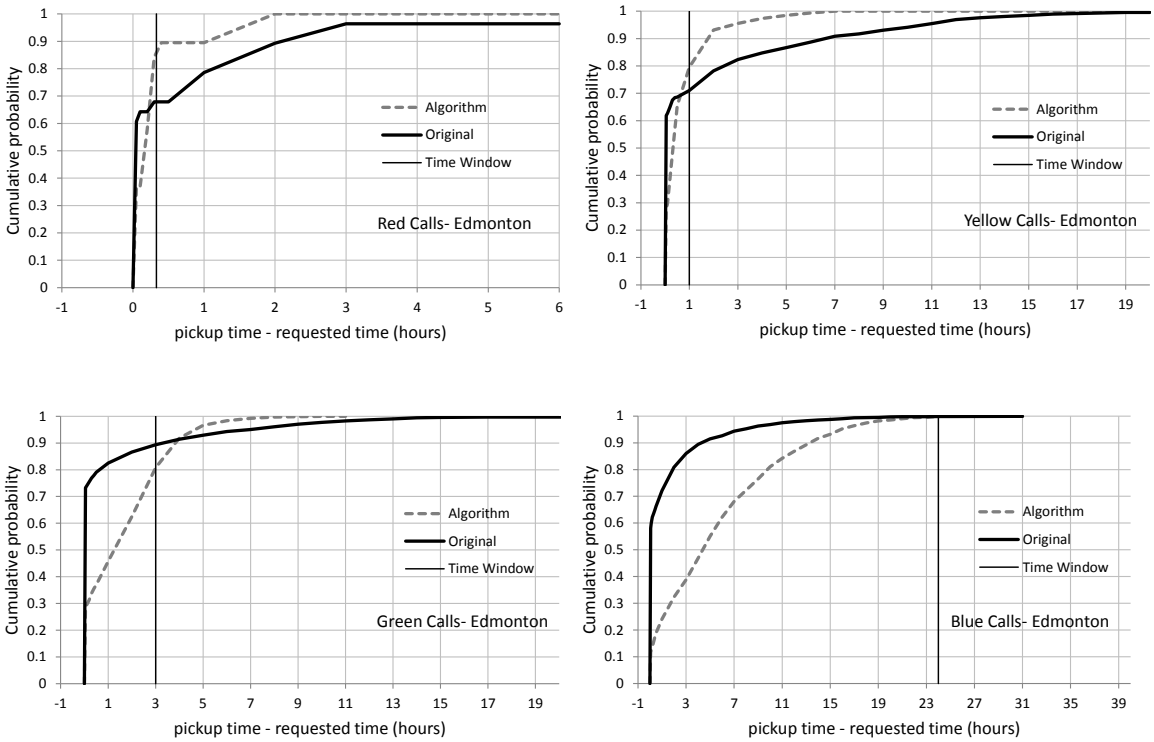

Figure 4: Edmonton Simulation Results

	Tardiness	Travel time	Deadheading	Used Units
Edmonton	67%	15%	3%	16%
Calgary	42%	22%	2%	34%

Table 7: Improvement in the IFT metrics in the algorithm solution compared to actual schedules

		Average Tardiness (hours)				Percentage of Tardy Calls			
		Blue	Green	Yellow	Red	Blue	Green	Yellow	Red
Edmonton	Algorithm	0.01	0.32	0.37	0.24	0.5%	21%	19%	42%
	Actual	0.06	0.51	1.33	0.26	0.2%	10%	30%	26%
Calgary	Algorithm	0.00	0.17	0.12	0.04	0%	17%	14%	16%
	Actual	0.00	0.32	0.19	0.17	0%	4%	11%	23%

Table 8: Algorithm solution compared to actual schedules in terms of tardiness

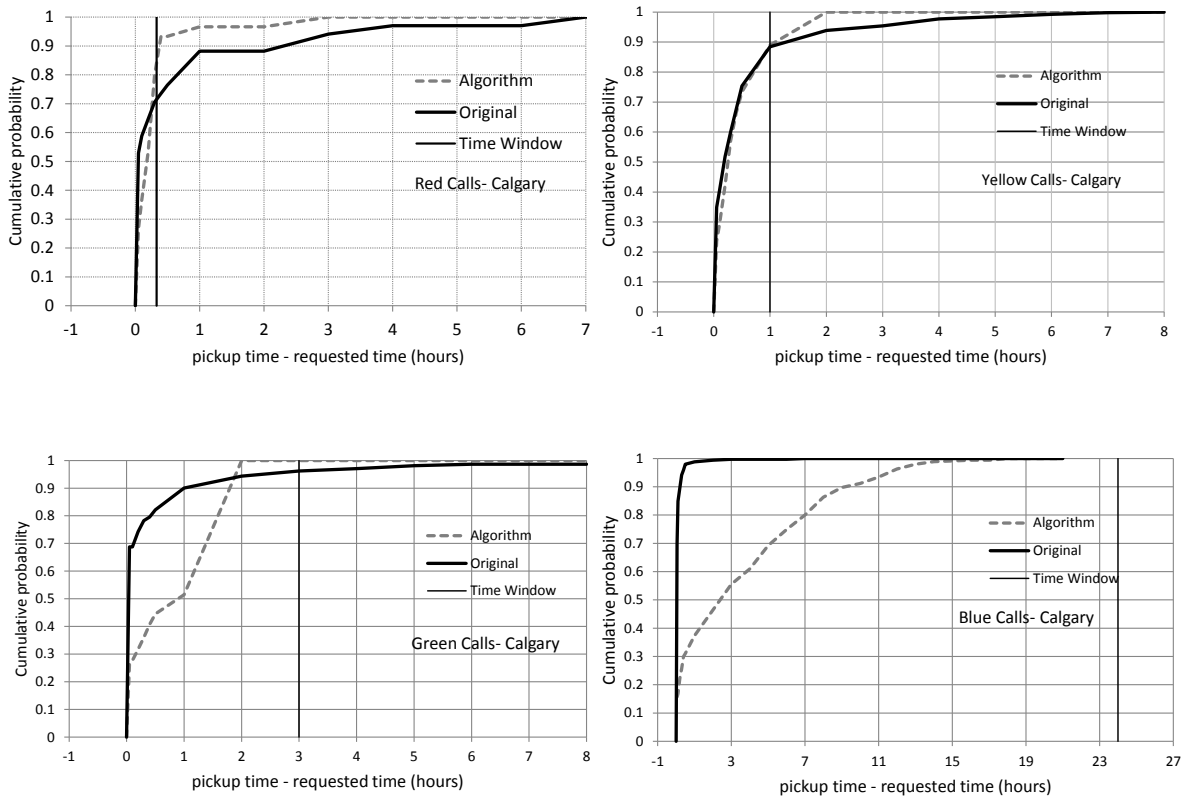

Figure 5: Calgary Simulation Results

	Actual	IFT-handled Only	IFT and EMS Handled
Number of units used	29.1	23.1	26.8
Travel time (hours)	47.0	35.8	44.2
Dead-heading (hours)	17.2	16.0	19.9
Number of missed time windows	7.0	2.8	4.1
Tardiness (hours)	35.7	10.3	11.4
Average EMS-handled calls covered	-	-	3.8
Total EMS-handled calls	119	119	5

Table 9: Simulation results on IFT and EMS-handled requests in Edmonton

	Actual	IFT-handled Only	IFT and EMS Handled
Number of units used	16.6	10.5	18.2
Travel time (hours)	14.1	6.5	18.5
Dead-heading (hours)	5.2	3.3	8.6
Number of missed time windows	1.2	0.0	0.3
Tardiness (hours)	5.5	2.2	5.9
Average EMS-handled calls covered	-	-	7.9
Total EMS-handled calls	253	253	10

Table 10: Simulation results on IFT and EMS-handled requests in Calgary

6.3 IFT Simulation Results on EMS-handled Requests

To test our method's ability to accommodate IFT requests that were handled by the EMS fleet, we generated schedules for all the IFT requests both handled by the IFT and the EMS fleet in the month of August. We used the simulation model to evaluate the solution. Tables 9 and 10 show the average values of these experiment results. We compare the actual performance of the IFT fleet with the algorithm solution for both cases of scheduling only IFT-handled requests, and IFT and EMS-handled requests. The simulation results suggest that out of all the EMS-handled calls, the algorithm is capable of accommodating 114 requests out of total 119 in Edmonton, and 243 out of 253 in Calgary.

7 Conclusion

We proposed an IFT ambulance scheduling and routing method. The method schedules both advance and emergent transfer requests, with the objective of minimizing travel times, patient time window violations, and fleet shift hours violations. We compared the performance of schedules generated with the the method to the performance of historical schedules by simulating 40 days in 2011 for the cities of Edmonton and Calgary. Our simulation results show significant potential for improvement, in terms of time window violations, tardiness, travel times, and unit utilization. Improvements in unit utilization have the potential to decrease the number

of uncovered requests and thereby reduce the amount of work that is transferred from IFT to EMS.

References

T Andersson, S Petersson, and P Värband. Opal–optimized ambulance logistics. In *Proceedings of TRISTAN V : The Fifth Triennial Symposium on Transportation Analysis*, 2004.

A Attanasio, J-F Cordeau, G Ghiani, and G Laporte. Parallel tabu search heuristics for the dynamic multi-vehicle dial-a-ride problem. *Parallel Computing*, 30(3):377–387, 2004.

A Beaudry, G Laporte, TerTesa Melo, and S Nickel. Dynamic transportation of patients in hospitals. *OR Spectrum*, 32:77–107, 2010.

British Columbia Patient Transfer, 2014. URL `www.bcas.ca/services/patient-transfer-services/`. (Online; retrived 4 July 2014).

S Budge, A Ingolfsson, and D Zerom. Empirical analysis of ambulance travel times: The case of Calgary Emergency Medical Services. *Management Science*, 56(4):716–723, 2010. ISSN 0025-1909.

G Clarke and JW Wright. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4):568–581, 1964.

J-F Cordeau and G Laporte. Tabu search heuristics for the vehicle routing problem. In R Sharda, S Voß, C Rego, and B Alidaee, editors, *Metaheuristic Optimization via Memory and Evolution*, volume 30 of *Operations Research/Computer Science Interfaces Series*, pages 145–163. Springer US, 2005.

J-F Cordeau and G Laporte. The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, 153:29–46, 2007.

J-F Cordeau, G Laporte, and A Mercier. A unified tabu search heuristic for vehicle routing problems with time windows. *The Journal of the Operational Research Society*, 52(8):928–936, 2001.

J-F Cordeau, G Laporte, and A Mercier. Improved tabu search algorithm for the handling of route duration constraints in vehicle routing problems with time windows. *The Journal of the Operational Research Society*, 55(5):542–546, 2004.

L Coslovich, R Pesenti, and W Ukovich. A two-phase insertion technique of unexpected customers for a dynamic dial-a-ride problem. *European Journal of Operational Research*, 175(3):1605–1615, 2006.

C Häll, M Högberg, and J Lundgren. A modeling system for simulation of dial-a-ride services. *Public Transport*, 4:17–37, 2012.

S Ho and D Haugland. Local search heuristics for the probabilistic dial-a-ride problem. *OR Spectrum*, 33: 961–988, 2011.

A Ingolfsson. *EMS Planning and Management*, pages 105–128. Springer, 2013.

A Ingolfsson, S Budge, and E Erkut. Optimal ambulance location with random delays and travel times. *Health Care Management Science*, 11:262–274, 2008.

RM Jørgensen, J Larsen, and KB Bergvinsdottir. Solving the dial-a-ride problem using genetic algorithms. *Journal of the Operational Research Society*, 58:1321–1331, 2007.

Y Kergosien, C Lente, D Piton, and J-C Billaut. A tabu search heuristic for the dynamic transportation of patients between care units. *European Journal of Operational Research*, 214(1):442–452, 2011.

M Khouadjia, B Sarasola, E Alba, E-G Talbi, and L Jourdan. Metaheuristics for dynamic vehicle routing. In E Alba, A Nakib, and P Siarry, editors, *Metaheuristics for Dynamic Optimization*, volume 433 of *Studies in Computational Intelligence*, pages 265–289. Springer Berlin / Heidelberg, 2013.

JK Lenstra and AHG Rinnooy Kan. Complexity of vehicle routing and scheduling problems. *Networks*, 11 (2):221–227, 1981.

RD MacDonald, B Farr, M Neill, J Loch, B Sawadsky, C Mazza, K Daya, C Olynyk, and S Chad. An emergency medical services transfer authorization center in response to the Toronto severe acute respiratory syndrome outbreak. *Prehospital Emergency Care*, 8(2):223–231, 2004.

O Madsen, H Ravn, and J Rygaard. A heuristic algorithm for a dial-a-ride problem with time windows, multiple capacities, and multiple objectives. *Annals of Operations Research*, 60:193–208, 1995.

E Melachrinoudis, AB Ilhan, and H Min. A dial-a-ride problem for client transportation in a health-care organization. *Computers and Operations Research*, 34(3):742–759, 2007.

SN Parragh, KF Doerner, RF Hartl, and X Gandibleux. A heuristic two-phase solution approach for the multi-objective dial-a-ride problem. *Networks*, 54:227–242, 2009. ISSN 0028-3045.

JY Potvin, T Kervahut, BL Garcia, and JM Rousseau. The vehicle routing problem with time windows Part I: Tabu search. *INFORMS Journal on Computing*, 8(2):158–164, 1996.

V Robinson, V Goel, RD Macdonald, and D Manuel. Inter-facility patient transfers in Ontario: Do you know what your local ambulance is being used for? *Healthcare Policy*, 4(3):53–66, 2009.

SOS Alarm, 2014. URL `www.sosalarm.se/`. (Online; retrived 4 July 2014).

P Toth and D Vigo. *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, 2002.