STRATEGIES TO SUPPORT AMBULANCE SCHEDULING WITH EFFICIENT ROUTING SERVICES

Günter Kiechle
TU Wien

Karl Dörner
Universität Wien

Stefan Biffl
TU Wien

Follow this and additional works at: http://aisel.aisnet.org/wi2009

Recommended Citation
Kiechle, Günter; Dörner, Karl; and Biffl, Stefan, "STRATEGIES TO SUPPORT AMBULANCE SCHEDULING WITH EFFICIENT ROUTING SERVICES" (2009). Wirtschaftsinformatik Proceedings 2009. 83.
http://aisel.aisnet.org/wi2009/83

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2009 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
STRAATEGIES TO SUPPORT AMBULANCE SCHEDULING WITH EFFICIENT ROUTING SERVICES

Günter Kiechle\textsuperscript{1}, Karl Dörner\textsuperscript{2}, Stefan Biffl\textsuperscript{1}

Abstract
For regular patient transportation and emergency transportation with ambulance vehicles a dynamic dial-a-ride (DARP) problem has to be solved. We analyze two scheduling strategies for the regular dial-a-ride orders in order to minimize routing costs and to maximize transportation quality. Furthermore, we review practical requirements and suggest a conceptual architecture for a decision support system based on a customary control center system as basis for future business services.

Major results of this contribution are two promising and efficient solution procedures suitable for a simplified bi-objective version of the dynamic DARP and tested with real-world problem instances.

1. Introduction

Many emergency service providers, especially ambulance departments and companies which provide non-public maintenance services, face the problem to provide different types of services with one fleet of vehicles:
(1) Emergency coverage for a certain region to provide immediate emergency service;
(2) Efficient regular service: scheduled pick-up and delivery of patients, predetermined service tasks, periodic pick-ups, etc.

This is also the current situation for the largest Austrian regional emergency service providers (e.g., the Austrian Red Cross), where the same fleet is used to provide both emergency and regular transport services. Dynamic emergency aspects thus directly influence the schedule for the regular service. When an emergency occurs and an ambulance is required, the vehicle with the shortest distance to the emergency is assigned to serve the emergency patient. Therefore, it often happens that an ambulance vehicle that has been scheduled for a transport order of a patient, but has not yet started, serves the emergency request. Thus, another vehicle has to be reassigned to the regular patient and the overall regular service schedule has to be re-optimized. Ambulances that carry out emergency transports become available at the hospital after the emergency service and can then be used to carry out regular transportation orders. Again, the schedule for regular services has to be re-optimized [10].

\textsuperscript{1} TU Wien, Austria
\textsuperscript{2} Universität Wien, Austria
Regular transportation services are offered for handicapped persons or patients with minor diseases, who could not use taxi services. Thinking of optimization for ambulance scheduling, we have to consider at least two perspectives. From the perspective of a transportation provider, the objective is to minimize cost of operations. On the other hand, to maximize quality of service is the objective from a patient’s point of view. Although both objectives comprise a multitude of factors, a simplified model of reality is subject to our investigations. Basically, we use the length of a tour in terms of driving time to model costs and waiting time of patients to model transportation quality. A tour or route is defined as the overall movements of a vehicle over a day of operation.

This paper reviews some of the results and experiences from the Ambulance Routing research project, which deals with different perspectives on the ambulance scheduling, e.g. consideration of expected transportation requests in vehicle scheduling determined from experiences in the past or waiting strategies for maximizing coverage to reduce response times for emergency services. In this contribution we will concentrate on regular transportation services, where a minor part of transportation requests arises dynamically and most of the requests are known beforehand. Emergency requests disturb regular operations, but may be modeled as dynamic requests with high priority.

Some related work has been published where pick-up and delivery requests occur dynamically (see [1], [5] and [11]). Likewise, a range of optimization algorithms has been developed and evaluated for variations of this problem (see [1], [4], [5], [11] and [12]). To our best knowledge, a dynamic changing fleet size and this type of disruption caused by emergency requests have not been considered so far but seem particularly desirable to improve the capability of decision support systems to provide future business services in real-world contexts.

A major goal of the Ambulance Routing project is to demonstrate potential advantages of optimization algorithms in a decision support pilot system for ambulance scheduling. Related work in the field of decision support systems has demonstrated the practical use of optimization for real-world vehicle routing problems (see [2] and [9]). Providing dynamic routing services requires a certain information system infrastructure that integrates Positioning Systems, Wireless Communication, and Geographic Information Systems to process necessary inputs for optimization and decision support (see [6][7]), which in turn can provide business services useful for dispatchers.

The contributions of this paper are twofold. On the one hand, we describe two promising and efficient solution procedures suitable for a simplified bi-objective version of the dynamic “dial-a-ride problem” (DARP) and evaluate the efficiency of these procedures with real-world problem instances in section 4.

On the other hand, we review restrictions and constraints for the development of a decision support extension for ambulance scheduling in section 3 and suggest a system architecture for integration of our solution procedures into an existing control center system to provide interfaces for future business services in section 5.

2. Problem description

The regular patient transportation problem can be considered as a variation of the DARP with additional real-world constraints regarding customer preferences or requirements. A comprehensive description of the DARP is given in [3] and partly repeated here to set a common ground for specific variations later on.
Let $G = (N, A)$ denote a graph consisting of a set of nodes $N$ and a set of arcs $A$. The set of nodes $N$ holds all pick-up locations $P = \{1, \ldots, n\}$, all delivery locations $D = \{n+1, \ldots, 2n\}$ and two copies of the depot 0 and 2$n$+1.

$$N = P \cup D \cup \{0, 2n + 1\}$$

For each arc in set $A$, a weight $t_{ij}$ is given and represents the driving time between two locations. The DARP consists of designing vehicle routes and schedules for $n$ customers or patients who specify pick-up and drop-off requests between origins and destinations. Note, that the number of customers $n$ equals the number of pick-up locations and also the number of delivery locations. A typical situation is that the same patient will have two requests during the same day or within a certain period – an outbound request, usually from home to a hospital, and an inbound request for the return trip.

Each transportation order (or request) incorporates a pick-up location $i$ out of set $P$ and a delivery location $i+n$ out of set $D$. Furthermore, let $g_i$ be a binary value indicating whether a request is inbound or outbound. We set $g_i$ to 0 for inbound and $g_i$ to 1 for outbound requests. In addition we define time windows $[e_i, l_i]$ and $[e_{i+n}, l_{i+n}]$ as well as service (or loading) times $s_i$ and $s_{i+n}$ for each pick-up and delivery location.

Regarding time windows, we have two situations – on the one hand, patients should be picked-up as late as possible from their home when they are being transported to hospitals; on the other hand, patients should be picked-up as early as possible when they are transported from the hospital back home. Deviations from the desired pick-up and drop-off times within the specified time window are considered in the objective function as “waiting time”. Time windows for each request are defined either by a desired delivery time $l_{i+n}^*$ for outbound requests or a desired pick-up time $e_i^*$ for inbound requests while remaining start and end values of each related time window are computed as follows:

$$l_i = e_i + W_i \quad \forall i \in P \cup D$$
$$l_{i+n} = l_i + t_{i+j+n} + s_i \quad \forall i \in P$$
$$e_{i+n} = e_i + t_{i+j+n} + s_i \quad \forall i \in P$$

In the standard case discussed in the literature, a homogeneous fleet of vehicles is considered. The objective is to plan a set of minimum-cost vehicle routes while serving as many customers as possible under a set of constraints. The main difference between the DARP and most classical routing problems is the fact that in the DARP human beings are transported instead of goods as in the other problems. Thus, additional constraints, e. g., maximum ride times for the patients, no waiting times with a patient on board, preferred pick-up and drop-off times are considered (see [3]). Our real-world DARP is extended in two ways: patients have an individual maximum waiting time $W_i$ given and vehicles may carry up to an individual number of $C_k$ patients concurrently, assuming a set $K$ of vehicles with a fixed number of vehicles in operation.

To find a solution for the described problem two decisions have to be made. On the one hand, the relative order of pick-up and delivery locations on a tour has to be set. We use the binary decision variable $x_{ij}^k$ to denote whether the locations $i$ and $j$ are visited from vehicle $k$ directly one after another.
\( x_{ij}^k = \{0,1\} \quad \forall i, j \in N, \forall k \in K \)

On the other hand, the actual pick-up or delivery times have to be fixed. Here we use the decision variable \( B_{ij}^k \) to denote the starting time of the loading/unloading process of vehicle \( k \) at a pick-up or delivery location \( i \).

\[ B_{ij}^k \quad \forall i \in N \setminus \{0\}, \forall k \in K \]

Once the starting time \( B_{ij}^k \) is set, also the end time \( D_{ij}^k \) of a loading/unloading process which equals the departure time of vehicle \( k \) at a location \( i \) may be determined. Above defined relations for time windows are shown in Fig. 1 which depicts a single segment of a tour.

\[ L_i = t_{i,i+n} + W_i \quad \forall i \in P \]

Second, a vehicle \( k \) is not allowed to wait idle while a patient is on board. Below, we use \( Q_k^i \) to denote the load of a vehicle \( k \) after its visit at location \( i \): we force to serve the pick-up at location \( j \) and the delivery at location \( i \) without detour or waiting in between if \( Q_k^i \) is positive after visiting location \( i \).

\[ B_{ij}^k - D_{ij}^k = t_{ij} \quad \forall x_{ij}^k = 1 \land |Q_k^i| \neq 0 \]

The overall objective in our problem is to optimize two criteria: minimize transportation costs and maximize quality of service for the regular transportation orders. Each criterion is normalized to a value in the interval \([0,1]\) and weighted with a constant factor. Emergency transportation orders are ignored since they have to be served as soon as possible.

Total costs of transportation usually include vehicle and personnel costs like driver wages, fuel costs, and other costs. Since a detailed cost structure is not available, we consider the overall tour length stated in driving time to be a sufficient proxy measure for transportation costs. The objective of minimizing transportation costs is thus stated as

\[ \min \sum_{k \in K} \sum_{i \in N} \sum_{j \in N} x_{ij}^k t_{ij} \]

Quality of service for regular transportation orders depends on customer waiting times, relative customer ride time (i.e., total time spent in vehicles compared with minimum possible time), and difference between actual and desired pick-up and drop-off times. These criteria are interrelated:
some are treated as constraints, some as part of the objective function. In our interpretation, the
objective of maximizing quality of service for regular transportation orders is thus stated as mini-
mizing the sum of waiting times, where waiting times are defined as deviations from desired pick-
up or delivery time.

\[
\min \sum_{k \in K} \left( \sum_{i \in P | g_i = 0} (B_i^k - e_i^k) + \sum_{i \in P | g_i = 1} (l_{i+n}^* - B_{i+n}^k) \right)
\]

The weighting factors for the different objective criteria are not subject to our observations. We
believe that fixing their values remains as an open political and managerial challenge.

3. Decision support for ambulance dispatching

Patient transportation and emergency services in Austria are organized in regional units. Although
a number of control center solutions are used, scheduling of vehicles is done manually by human
dispatchers with all solutions. The solutions are supported by customized information systems pro-
viding immediate access to all relevant data about transportation requests, vehicle states and
equipments, etc.

To enhance the current way of ambulance scheduling we suggest extending existing control center
systems with components of a model-driven decision support system (see [13]). Due to practical
issues two side constraints have to be considered:

(1) Existing workflows, service- and administrative processes, roles of staff, and infrastructure are
not subject to change.
(2) The existing control center systems may not provide all interfaces to smoothly integrate optimi-
   zation components, scheduling recommendations, or even semi-automatic vehicle scheduling.
   Thus, decision support should work with a minimum set of interface requirements to existing sys-
   tems.

For a practical case study about using decision support elements in ambulance scheduling we colla-
brorate with the Red Cross in Salzburg (RCS) to use their current processes and infrastructure. RCS
uses a modern system that features wirelessly connected mobile information systems on the ve-
hicles. The system supports integrated order management and regularly sends vehicle positions and
operation states to the control center, which already helped improving efficiency of transportation
operations (see [7]).

In order to elicit ways to support dispatchers in their everyday work and to explore corresponding
requirements from an end-user perspective, two workshops with dispatchers were organized. Main
results of those discussions were:

(1) Dispatchers face difficulties to overlook the entire quantity of transportation requests to find
   synergies. Thus, the most important feature of a decision support system is a filter (or recommenda-
   tion), which transportation orders could (and also should) be serviced as a group. We call this strat-
   egy to “combine transportation requests” and provide a more detailed examination of this issue.
(2) The final assignment of a transportation order to a vehicle is done on short notice, because
   emergency requests and new regular transportation orders may come in any time. A rough schedule
   might be planned some time in advance, but usually this draft schedule is altered frequently. Thus,
   the response time of a real-world decision support system has to be clearly less than one minute
   while the planning horizon is typically a few hours.
(3) Human dispatchers consider a broad range of hard and soft constraints in their decisions, which could not exhaustively be incorporated in a model-based approach. Nevertheless, an intuitive and fast way of processing recommendations for assignments of orders to vehicles is required. (4) A decision support system should make no decisions without a human dispatcher – this clearly states that a semi-automatic ambulance scheduling approach is not up for discussion.

As a result of these findings we propose two features of the decision support system. First, a list of currently possible combinations of transportation requests should be given, regardless of vehicle availability. The list should also value the combinations in terms of tour length and waiting time as well as state the latest time that the implementation of the combination has to be decided. Second, a visual representation of a suggested schedule for all vehicles should be shown. With this illustration human dispatchers can view the suggested assignment of requests to vehicles.

3.1. Combination of transportation requests

If two or more transportation requests meet vehicle capacity and time restrictions, they may be combined, i.e., more than one patient is on board of a vehicle at the same time. Two basic schemes of combinations apply to our problem structure and are depicted in Fig. 2: The common sequence of requests is one after another like shown in sub-picture (1) where pick-up and delivery for each request are scheduled consecutively. In the first scheme of combination shown in sub-picture (2) two pick-ups are scheduled consecutively and the request that was picked up later is delivered first. The second scheme shown in sub-picture (3) also features two consecutive pick-ups first, but here the request picked up later is delivered last.

![Fig. 2: Transportation request combination schemes.](image)

Using these options for combinations of requests may lead to shorter tour lengths or waiting times or both, but only requests in temporal and spatial proximity are beneficial and fulfill the maximum transportation time restriction. Certainly, these schemes of combinations are also applied to a higher number of combined requests.

3.2. Visualization of a suggested schedule

In order to provide an intuitive visualization of calculated scheduling recommendations we propose to use a bar chart approach for communication with human dispatchers. Fig. 3 illustrates a prototype visualization module that shows each vehicle and its suggested requests on a separate line. The picture shows request types denoted as “D” for deliveries (and “P” for pick-ups), beginning times of pick-ups and ending times of deliveries as well as combinations of requests.
In order to study different dispatch strategies for the problem at hand, we developed a simple and effective solution procedure: We implemented a constructive heuristic approach. In the construction phase, we exploit the temporal structure of requests and use a nearest-neighbor measure inspired by [8]. Additionally, a more sophisticated solution approach corresponding to the pilot method of [14] was implemented and compared to the heuristic procedure.

Each solution procedure is capable to deal with the static and the dynamic version of the DARP. In the dynamic case, the list of transportation requests changes over time. Furthermore, disruptions are considered, which are caused by emergency requests that are not known beforehand and must be served with high priority. In case of an emergency, the empty vehicle which is closest to the emergency location will be redirected and is not available for regular operation any more. In this situation, a re-calculation of the schedule is necessary.

The heuristic solution procedure follows a greedy approach and constructs a valid solution in a step-wise process: in a pre-processing step a sorted list of requests is computed and processed starting from the earliest to the latest request until the list is empty. In each construction step a further request is added to one of the tours in the current solution. To determine the request to be added a set of most promising tour-request pairs is computed in each step. The decision criterion is the distance from the delivery location of the last request to the pick-up location of the current request.

The pilot method may be characterized as general approach to enhance a constructive heuristic procedure while providing a sophisticated decision rule in each construction step. The basic idea is to avoid unfavorable decisions by looking ahead for each choice to be made. In our case the heuristic solution procedure described above is used as initial heuristic that is extended with the pilot concept. Our pilot method procedure evaluates a number of promising options in each construction step while computing a pilot solution for each option. All pilot options are then reviewed using the overall objective function. The best solution determines the option that is finally fixed in each construction step. Note that in contrast to the classic approach reported by [14] we evaluate not all but only a limited number of promising solutions in each step.

To evaluate the solution methods a set of 15 problem instances taken from a larger real-world dataset of the Austrian Red Cross with a number of regular transportation requests ranging from 152 to 286 was used. Table 1 shows the average results for all 15 instances, where the objectives of tour length and waiting time were weighted 4:1. The results for the pilot method are average values – a number of 2 to 8 pilot solutions were evaluated in each stop of the procedure – and results for 2 and 8 pilot solutions. Table 2 shows run times for the solution methods calculating complete solutions for the smallest, a medium and the biggest problem instance on an Intel XEON CPU with 2.8 GHz and 4 GB RAM. Note that the heuristic procedure and even the pilot method with 2 pilot solutions are fast enough to provide solutions in most cases within the requested response time of one minute. Due to the dynamic nature of the problem, a rolling horizon may be applied in practice and
substantially shorten run times. This means that even a higher number of pilot solutions may be calculated and the solution procedure still performs in time as long as the number of considered requests in the rolling planning horizon is reduced properly.

Table 1: Optimization results of 15 instances with 150+ to 280+ transportation requests.

<table>
<thead>
<tr>
<th>Solution method</th>
<th>Objective value</th>
<th>Tour length</th>
<th>Waiting time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic procedure</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Pilot method (avg.)</td>
<td>91.7%</td>
<td>86.6%</td>
<td>120.6%</td>
</tr>
<tr>
<td>Pilot method (2 pilots)</td>
<td>93.1%</td>
<td>87.0%</td>
<td>127.8%</td>
</tr>
<tr>
<td>Pilot method (8 pilots)</td>
<td>91.1%</td>
<td>86.6%</td>
<td>116.0%</td>
</tr>
</tbody>
</table>

Table 2: Run times in seconds.

<table>
<thead>
<tr>
<th>Number of transportation requests per instance</th>
<th>Heuristic procedure (2 pilot solutions)</th>
<th>Pilot method (2 pilot solutions)</th>
<th>Pilot method (8 pilot solutions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smallest (152 reqs.)</td>
<td>0.09s</td>
<td>9.27s</td>
<td>32.8s</td>
</tr>
<tr>
<td>Medium (218 reqs.)</td>
<td>0.25s</td>
<td>35.9s</td>
<td>129s</td>
</tr>
<tr>
<td>Biggest (286 reqs.)</td>
<td>0.38s</td>
<td>69.2s</td>
<td>254s</td>
</tr>
</tbody>
</table>

Investigating parameter settings for the pilot method, we found that waiting time declines by far more than the tour length with an ascending number of pilot solutions. The results given in Fig. 4 are average values over 15 problem instances.

Fig. 4: Results of the pilot method depending on number of pilot solutions used

5. Integration of decision support services into a control center system

In Austria, several control center solutions for ambulance scheduling are currently run by regional branches of the Austrian Red Cross. The described solution procedure has the potential to enhance efficiency in any existing control center system. Therefore, decision support functionality to enable more efficient ambulance scheduling could and should be integrated into an existing system. It is the aim of the Ambulance Routing project to demonstrate potential advantages of optimization algorithms in a pilot setting together with the Red Cross Salzburg. This case study aims at illustrating the use of optimization algorithms to provide business services in a practical environment.

Since one could not expect a control center system to feature all the necessary interfaces to smoothly integrate scheduling recommendations we propose a loosely coupled system architecture that minimizes dependencies between decision support components and the rest of the system.
Fig. 5 shows an overview of the proposed decision support system, where on the left side an existing control center solution is depicted. Extensions for decision support are shown on the right side and color-coded differently. To provide scheduling recommendations for human dispatchers the system proceeds as follows:

1. A list of current and forthcoming requests and current vehicle states is obtained from the Control Center System and pre-processed in the Integration System
2. Solution procedures are applied to the transportation request data and a recommendation of a promising vehicle schedule is calculated
3. The recommended schedule is visualized to the human dispatcher, who can decide whether to incorporate parts of the suggested tour plans or not. The actual decisions of the dispatcher are fed back to the schedule recommender via (1).

![Ambulance routing decision support system overview.](image)

All input data (1) for the decision support sub-system is gathered via a single web-service interface that provides a complete snapshot of dispatching status. The Integration System {1} and the Schedule Visualizer {3} are implemented with Java technology, while the Solver {2} is implemented in C++ for performance reasons.

6. Conclusion and Outlook

In this paper we introduced a vehicle routing problem relevant to ambulance scheduling. The problem is based on the well-known dial-a-ride problem (DARP), but features a special extension that causes disruptions because of emergency service requests. We reported restrictions and constraints for the development of a decision support extension for an existing control center system. The findings were gathered from workshops with human dispatchers. We developed two solution approaches, a greedy heuristic and a pilot method procedure suitable for a simplified bi-objective version of the dynamic DARP. We evaluated the solution procedures with real-world test data from an ambulance service provider in Austria and compared the results. We found that both algorithms perform well enough in terms of runtime. Finally, we suggested a system architecture with a web-service interface for integration of our solution procedures into an existing system.

Next steps in the Ambulance Routing project will be the integration of all components of the architecture into a running prototype system, which will be used for evaluation of our approach under real-world conditions. Expected findings of these investigations will be actual improvements enabled through optimization as well as a measure for the quality of our schedule recommendations.
7. Acknowledgements

We would like to thank Stefan Achleitner, UweKirchengast, and Michael Sprinzl for their contributions to the bar chart visualization of vehicle schedules. Furthermore, financial support from the Austrian Science Fund (Fonds für wissenschaftliche Forschung; FWF) under grant #L286-N04 is gratefully acknowledged.

8. References


