Trip Plan Generation using Optimization: A Benchmark of Freight Routing and Scheduling Policies within the Carload Service Segment

Lars Backåker* and Johanna Törnquist Krasemann

Linköping University, Department of Science and Technology, 601 74 Norrköping, Sweden

The rail freight carload service segment enables the distribution of freight volumes down to the unit of single rail cars, and stand as an important alternative to road transportation. However, this service segment is often associated with significant uncertainties and variations in daily freight volumes. Such uncertainties are challenging to manage since operating plans generally are established long in advance of operations. Flexibility can instead be found in the way trip plans are generated. Previous research has shown that a commonly used trip plan generation policy does not exploit the available flexibility to the full extent. In this paper, we therefore suggest an optimization-based freight routing and scheduling (OFRS) policy to address the rail freight trip plan generation problem. This OFRS-policy generates trip plans for rail cars while still restricted by the customer commitments. The policy involves a MIP formulation with a continuous time representation and is solved by commercial software. We apply the OFRS-policy on a case built on real data provided by the Swedish rail freight operator, Green Cargo, and assess the performance of the policy comparing the current industry practice. The results show that by using the OFRS policy, we can achieve a reduction in the total transportation times, number of shunting activities and potentially also a reduction in the service frequency given the considered transport demand.

*Corresponding author (lars.backaker@liu.se, +46 (0) 11363481)

Keywords: Rail freight planning, trip plan generation, routing and scheduling, dynamic assignment
1. Introduction

Rail freight operators manage a number of different service segments among which one is the carload service segment, also known as the single wagonload (SWL) service segment. Since the service segment enables the distribution of freight volumes down to the unit of single rail cars, rail freight carload transportation is an important alternative to road transportation and especially less-than-truckload transportation. However, at the same time as this service segment can be clearly motivated from a service-oriented perspective, it brings significant planning challenges related to the need for effective planning to reach economy-of-scale and profitability. That the carload service segment is subject to significant uncertainties and large variations in freight volumes is today widely recognized. At the same time, it is common to initiate the planning process up to one year in advance of operations, which limits the possibilities to change the operating plan when the transport demand fluctuates over time. Flexibility can instead be found in the way trip plans are generated, where each transport request (i.e. rail car) is assigned a route and a schedule through the network of different train services. Such flexibility appears in terms of routing and scheduling options, where rail cars can be routed on to different paths of the service network, postponed at yards and scheduled on a number of different train services. However, in previous case studies [1] we have observed that the type of routing and scheduling principles currently used by the main Swedish rail freight operator does not exploit the available flexibility to the full extent and do not select train services in an optimized way. We have also been given indications that the principle currently applied to manage capacity shortages on services, i.e. the First-Booked-First-Served (FBFS) principle, reduces the degree of flexibility even further (see Section 2.2).

In this paper, we therefore suggest an optimization-based freight routing and scheduling (OFRS) policy to deal with the rail freight trip plan generation problem. The OFRS-policy routes and schedules rail cars onto available train services freely while still restricted by the customer commitments (e.g. agreed delivery time frames) and service characteristics (e.g. departure times and capacity limits). The policy involves a Mixed Integer Linear Programming (MILP) formulation which is solved by commercial optimization software. In contrast to previous models, our MILP-formulation has a continuous time representation which enables a more detailed representation of the service network and reduces the number of required binary variables. We apply the OFRS-policy on a case built on real data provided by the Swedish rail freight operator, Green Cargo. We also assess the performance of the OFRS-policy in a benchmark with the industry practice currently used by Green Cargo for the trip plan generation.

The structure of the paper is as follows; in Section 2 we provide an introduction to the rail freight planning and distribution process, and existing rail car priority principles. In Section 3, we provide an overview and discussion of related work followed by Section 4, which presents our proposed optimization-based dynamic trip plan generation policy (OFRS). In Section 5, we outline our experiments and the results are presented and discussed in Section 6. In Section 7, we provide conclusions and some directions for future research.

2. Rail freight carload planning and distribution

This section presents the planning and distribution process for rail freight carload transportation and primarily from a European perspective, where services operate according to pre-defined timetables.
2.1. The rail freight planning process

The complexity of planning rail freight operations has lead to the development of a step-wise, partly iterative planning process [2]. The process has been well described in literature; see e.g. Ahuja et al. [3] for a comprehensive introduction, or Crainic and Laporte [4] for a review of related planning models. Fig. 1 provides an overview of the general planning process including the most essential planning activities illustrated by dashed boxes. Solid lines are used to represent the information flows between these activities and also highlight the iterative workflow between activities.

Fig. 1. Essential rail freight planning activities in the planning process (*external events).

The long-term planning process is primarily based on forecasts of future transport demand which provides an indication of how large flows that may occur on different transport relations. When trains operate according to pre-defined schedules, four main planning activities are carried out in a step-wise manner. The blocking plan (A) specifies how rail cars are to be grouped, also known as classified, into blocks. The assignment is highly dependent on the origin and destination of the rail car and possibly also on the service class. These blocks then need to be transported on different regional train services defined by the train make-up plan and the timetable (B). This activity is also referred to as train scheduling. In Europe, where the railway sector is becoming increasingly deregulated, rail freight operators are required to apply for train slots well in advance, and the final timetable is then created and published by the independent infrastructure manager. This activity is known to be time-consuming and complicates the planning process from the operator point of view. When the timetable is finalized, the locomotive schedule (C) is established to assign locomotives to each slot in the timetable. The crew schedule (D) is then created by assigning staff to the locomotive schedule. Iterative loops and cycles are commonly introduced in between the above mentioned planning activities. The joint result of these four planning activities is then the operating plan, also known as the production plan or master plan, and concludes the long-term planning process.

In the operational planning phase, the continuously incoming stream of transportation requests (E) is managed by the operator in what is referred to as the booking process and for each rail car, a trip plan (F) is created. This trip plan specifies when the rail car is to be picked-up for transport and how it is supposed to be 1) routed through the pre-defined network and 2) assigned to specific train services. Each transportation request consists of a single, alternatively multiple, rail car(s) with
certain characteristics; typically release time, origin, destination, weight, length, shipper (i.e. sender) and consignee (i.e. receiver). The general approach when generating trip plans is to 1) first let the classification scheme defined by the blocking plan determine which train services that are available for the individual rail cars and then 2) select among the available train services according to certain basic principles, e.g. first-available-departure. Capacity on services is in this phase roughly considered and in situations when services already are overbooked rail cars are simply scheduled onto the next available departure. The assignment principle is repeated at each intermediate terminal the rail car is planned to traverse between its origin and destination.

The transport demand may vary significantly and uncertainties in terms of daily freight volumes complicate the capacity requirement estimations during the long-term planning process. Consequently, the capacity of train services from time to time becomes insufficient. In Sweden, train capacities generally lie in the span of up to 630 meters and between 1100-1600 tons. Since the Swedish railway network is shared among different actors, sidings are required for passenger trains to be able to overtake e.g. slower freight trains, and the length dimension is thereby foremost restricted by the length of the available sidings.

The short-term planning also addresses the empty rail car distribution in the service network (G). The balancing of the flow of empty rail cars is crucial for enabling the distribution of loaded rail cars within the network.

2.2. The rail freight distribution process

Once a transportation request has been processed and the corresponding rail car(s) has been assigned a trip plan, information regarding preliminary pick-up and delivery dates is communicated to the customer. In such way the customer can prepare the rail car(s) for pick-up at its origin terminal. The origin terminal is often the terminal of the shipper. The first distribution activity then involves a local train which operates to collect these single rail cars from a set of terminals, and each terminal is visited a strictly limited number of times per day. Depending on daily freight volumes it is even possible that pick-up activities are cancelled and the collection of rail cars is postponed until the next day. The local trains then leave the collected rail cars at larger terminals to be shunted, if necessary, into blocks according to the blocking plan and connected to the scheduled regional train service. Later the blocks are separated at major terminals and local trains are used to transport the individual rail cars to their final destinations.

During the distribution process it is possible that the rail cars are re-classified for further distribution and the trip plan is then revised using simple priority principles. Examples of such priority principles are the First-in-First-Out (FIFO) principle alternatively the First-Booked-First-Served (FBFS) principle [5]. The FIFO-principle is frequently used by North American rail freight operators and implies that rail cars are prioritized based on the time of arrival at intermediate locations. In situations when the capacity is sufficient cut-off times of services are used to determine if rail cars are on time for specific departures. When capacity is insufficient, the priority principle is used to determine which rail cars that has to be postponed for later departures. The FBFS-principle is what can be considered an extension of the FIFO-principle and is commonly used by the Swedish rail freight operators. In addition to train cut-off times, the FBFS-principle also considers the time of booking as a priority parameter. When capacity shortages arise on specific train services, rail cars with early booking times are prioritized over rail cars with later booking times.
We choose to jointly refer to the currently used industry practice for trip plan generation, including the rail car scheduling principle (first-available-departure) and the priority principle (FBFS), as the base-line policy further on in this paper.

3. Related work

The challenges in the area of rail freight planning are well-known and many of them have gained a lot of attention – especially the different planning tasks in the long-term planning process (see Fig. 1 in the previous section). Traditionally, these tasks were solved as separated problems due to their complexity and lack of computational support to solve them together. Over the years, advanced models and methods have been developed and the use of integrated planning models has become more frequent. An examples is the use of Service Network Design (SND) formulations which tend to jointly address planning issues related to both the blocking plan problem (A), the assignment of blocks to trains (B), and implicitly also the trip plan generation (F). The SND-formulations generally serve to define the network of services. This is done by determining which train services to operate and at which service frequency, in order to accommodate certain transport demand and customer commitments while minimizing operating costs (alternately maximizing profit). Time-space networks are used to represent the movement of services and freight flows over time on and between a set of terminals. The main objective is often to determine operating plans for standard weeks of operations. Occasionally also balance constraints of vehicles entering and exiting the terminals are included as to represent fleet and asset management circulations. Comprehensive reviews of SND applications within the transportation sector have been published by Crainic [6] and also later by Weiberneit [7]. To the best of our knowledge, the first SND formulation was proposed by Crainic and Rousseau [8] out of which several contributions have focused on extensions and applications. For rail freight applications we can refer to Zhu et al. [9], Andersen and Christiansen [10], Andersen et al. [11], Lulli et. al [12], Campetella et. al [13] and Ceselli et. al. [14]. Kim et al. [15] address an airline application, while Meng and Wang [16] as well as Lai and Lo [17] focus on maritime applications. Jarrah et al. [18] and Dall’Orto et al. [19] propose approaches for Less-than-Truckload (LTL) problems.

The SND models are often deterministic and capacitated multi-commodity network design (CMND) formulations, though, also stochastic contributions have been provided by e.g. Lium et al. [20]. The flows of commodities and trains are typically represented in terms of a time-space network where time is discrete and divided into distinguishable time periods (often days). The formulations thereby contain a large number of binary alternatively integer variables which grow significantly with an increased planning horizon and when the level of time discretion increases. Consequently, if it is important to model the time dimension in more detail, the existing models are not applicable without larger modifications.

The blocking problem (A) has been addressed in a number of publications, see e.g. the early review of related work by Assad [21]. Other contributions have been published by Bodin et. al. [22] and Newton et. al. [23] who adopt a path-based network design formulation. Also Barnhart et. al. [24] have emphasized on generating feasible solutions for real-sized problem instances. More recently, Ahuja et. al. [25] have suggested an algorithm based on Very Large-scale Neighborhood (VLSN) search techniques that find near optimal solutions to the blocking problem in relatively short computational time. New solution strategies using Ant Colony optimization algorithms have also been proposed by Yaghini et al. [26]. The above mentioned publications provide insights on how to
route rail cars and how to classify rail cars into blocks, but all are intended for use as decision support in the long-term planning process. The train make-up problem concerns the assignment of already existing blocks to trains and can be viewed as an extension of the blocking problem. When considered jointly, such approaches have clear similarities with SND formulations. However, detailed train scheduling aspects still remain [27].

Train scheduling models (B) originate from long-term planning levels. Though, lately also pure models intended for operational train dispatching problems have been developed. Train scheduling formulations provide adequate time representations also for the scheduling of freight movements and also here formulations that address multiple planning aspects are common. Yano and Newman [28] proposed a model where train schedules are to be established based on a dynamic arrival of freight at origins. Each transportation request imposes a due date for final delivery and requires certain amount of capacity to be transported. The objective is to minimize distribution costs associated with differentiated item holding costs induced due to postponements and storage activities at terminal locations.

Research approaches which focus on the trip plan generation problem (F) are most similar to the approach proposed in this paper. Two different approaches are presented by Kwon et al. [29] and Anghinolfi et al. [30], respectively. Kwon et al. [29] focus on finding improvements of the current rail car scheduling practices in the U.S. and propose a Capacitated Multi-commodity Flow Problem (CMFP) formulation. The motivation lies in the ineffectiveness of current rail car scheduling practice, since it does not explicitly consider capacity limits of resources. Train services operate according to fixed schedules and their capacity is restricted by maximum length. A time-space network is adopted to represent the movements of groups of rail cars (referred to as commodities) as blocks and the movement of blocks on trains in the service network. Each commodity has a pre-defined origin, destination, length and time window which specifies acceptable delivery times to the final destination. Penalty costs are introduced for late arrivals with respect to the time windows and different time windows are adopted to represent differentiated service classes. The formulation is path-based and each commodity is assigned a path which describes the train sequence used to distribute the commodity from its origin to its destination. Each time a commodity arrives after the defined delivery time window a penalty is induced and the objective is to minimize penalty costs resulting from the selected commodity paths. Holding links are used to represent when a rail car is held at intermediate terminals and dummy links for the arrivals and departures. The problem is formulated as a Mixed Integer Problem, where binary variables are used to model if the commodities are assigned to a certain block and then to a certain train, while the quantity of each commodity that flows is a continuous variable. Experimental studies are conducted on a hypothetical network based on a major U.S. railroad using projected volume by day of week to represent transportation request variability. Potential commodity paths are pre-generated and column generation techniques along with the IBM optimization sub-routine library and C-programming language are used to solve large instances (12 terminals, 16 trains, 59 blocks and 9712 cars for a one week period) in relatively short time.

Anghinolfi et al. [30] address a routing and scheduling problem in a similar context and with automated freight terminals. The transport demand (referred to as orders) is modeled in terms of containers or swap-bodies (referred to as boxes) which are to be assigned to available train departures. The boxes are then assigned to rail cars on the selected train services. Each box has a
release time and due date for final delivery. Both the train services and the rail cars have a limited capacity in terms of weight and length. For each box, a set of all possible train service sequences is generated in a pre-analysis which excludes the capacity restrictions. Then a Binary Integer Programming formulation with the capacity constraints is applied to the sets of possible train service sequences for the boxes, where binary variables are used to select rail cars and train sequences. The objective is to minimize costs associated with each train sequence, additional train costs and potential penalty costs for orders not being served. The formulation is used on realistic-sized, though fictive, problem instances and it is reported that only a few problem instances are solved satisfactory when using commercial MIP-solvers. Two heuristics, one “ad hoc heuristic” and one “randomized neighborhood search heuristic” are therefore developed for being able to find good solutions to the problems. Kwon et al. [29] and Angholfini et al. [30] address what is to be considered off-line planning problems and focus on the assessment of service improvements related to the trip plan generation problem. None of the two mentioned applications include hard constraints on delivery requirements; instead penalties are used to capture costs of arrivals outside of the specified time windows and orders being un-served. The two approaches also require that the set of possible train sequences alternatively service paths, are pre-generated, resulting in a two-step solution process.

Another related problem area concerns equipment distribution (G), where empty rail cars are to be repositioned in the service network as to compensate for regional imbalances in freight flows. A recent review of the current state of the art in empty rail car distribution using optimization-based decision support systems is provided by Gorman et al. [31]. Optimization-based equipment distribution models intended for railway systems with pre-defined operating plans, as in Sweden, have been proposed by e.g. Joborn et al. [32]. Here both routing and scheduling decisions are addressed where empty rail cars are assigned to existing services as to be able to balance the highly dynamic demand for distribution equipment.

4. An optimization-based freight routing and scheduling approach

As mentioned earlier, we use the term base-line policy to denote both the trip plan generation principle (which selects the first available train departure) and the priority principle (FBFS) applied during capacity shortages which are currently used by Green Cargo. According to what has been observed in previous studies presented in Backåker et al. [1], the base-line policy does not enable the full use of existing system flexibility. We provide two examples to illustrate its weaknesses and the potential benefits in terms of increasing flexibility that would be possible to achieve by replacing the base-line policy. First, consider the fact that a rail car \( k \) with a later release time might today be given priority over another rail car \( k^* \) with an earlier release time. This situation can occur if the later rail car \( k \) is booked earlier than rail car \( k^* \) and they compete over the remaining capacity on a specific service. Consequently, the departure of rail car \( k^* \) will be postponed despite that it was available for transport earlier than rail car \( k \).
Our second example, illustrated by Fig. 2, also shows the short-sighted view of the base-line policy. Consider a rail car which originates from terminal yard 1 and is to be transported to terminal yard 4. Two alternative routes are available: The rail car can be transported on the upper path from yard 1 and yard 2, and continuing. Observe that the services on the upper path are available only Monday through Friday. Alternatively, the rail car could be transported on the lower path where a single direct service is offered on Saturdays (assuming this would not violate the delivery commitments). The latter option would probably require a certain degree of postponement of the rail car at its origin, while awaiting the Saturday service to depart. In case of possible risk of capacity shortage on the upper path (e.g. between yard 2 and 3) alternatively if another objective such as to reduce the number of shunting activities (i.e. terminal handlings) was applied, the Saturday service would be an attractive option. However, with the base-line policy where the first-available service always is selected, the rail car will (with only a few exceptions e.g. late Friday alternatively Saturday release times) always be routed on the upper path.

The proposed optimization-based freight routing and scheduling approach (referred to as the OFRS-policy) accounts for the above mentioned issues and can be considered pro-active since it considers e.g. overall transportation times of rail cars. The approach uses a MILP formulation, outlined below, which has been inspired by existing SND applications. We use commercial optimization software to solve the problem instances.

### 4.1. The MILP problem formulation

Since the problem formulation proposed here is foremost targeting the carload service segment, we model each transport request as one inseparable unit of freight corresponding to one single rail car (the model supports, however, also other demand representations). Each individual rail car \((k \in K)\) to be transported has a specific origin, destination, weight, volume, earliest time of departure and a latest time of arrival. The transport service network is composed of a set of terminal nodes \((n \in N)\) where each node is connected by a number of service links \((l \in L)\). On each link, a set of train services \((s \in S_l)\) operate. The services operate according to a pre-defined schedule and with a
maximum capacity with respect to length (meters) and weight (tons). Fig. 3 illustrates the service network representation corresponding to the instance used in our experiments. The solid lines are used to represent services and black circles are used to represent the terminals (i.e. shunting yards).

Fig. 3. Illustration of the service network representation for the real network used in the experiments.

Given the service network presented in Fig. 3, the main problem is to assign all rail cars \(k \in K\) to train services as to provide transport from the associated origin node to the destination node, possible via intermediary nodes. The objective is to minimize the transportation times while constrained by the limited capacity of services and the transport commitments associated with each customer and rail car. The formulation we propose adopts a continuous time representation in contrast to most previous formulations of similar problems as discussed in Section 3. We use a continuous variable \(x_{k,n}\) to represent the arrival time of rail car \(k\) at terminal node \(n\). The binary variable \(y_{k,l,s,d}\) is used to represent if rail car \(k\) is transported on link \(l\) using service \(s\) at day \(d\) (\(=1\)) or not (\(=0\)). We define two sets to represent days \(D\) and \(D_k\). The set \(D\) constitutes all days with respect to the entire planning horizon for the given problem instance, whereas the set \(D_k \subseteq D\) is used to represent days with respect to specific rail cars. Days of \(D_k\) can thereby be interpreted as the set of days within \(D\) during which the rail car \(k\) is being transported.

The complete problem formulation is outlined below and Table 1 lists the parameters used.

Table 1. List of parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta^{o}_{k})</td>
<td>-The origin node (n) of rail car (k), where (\theta^{o}_{k} \in N)</td>
</tr>
<tr>
<td>(\theta^{d}_{k})</td>
<td>-The destination node (n) of rail car (k), where (\theta^{d}_{k} \in N)</td>
</tr>
<tr>
<td>(t^{r}_{k})</td>
<td>-The release time of rail car (k), i.e. when the specific rail car becomes available for pick-up at the origin location (\theta^{o}_{k}). Given in minutes from midnight the first day (d \in D).</td>
</tr>
<tr>
<td>(t^{d}_{k})</td>
<td>-The release day (d) of rail car (k), i.e. the day when the rail car becomes available for pick-up</td>
</tr>
</tbody>
</table>
The relationship between the days \( d \in D \) and the internal days \( d' \in D_k \) can be expressed as:

\[
d = t_k^{rd} + d' - 1 \quad \text{where} \quad D_k = \{1, ..., f + 1\}.
\]

\[
\begin{align*}
\text{minimize} & \quad \sum_{k \in K} \left( x_{k,\theta_k^d} - x_{k,\theta_k^d} \right) \\
\text{subject to:} & \\
x_{k,\theta_k^d} & = t_k^{rt} \quad \forall k \in K \\
x_{k,\theta_k^d} & \leq t_k^{max} \quad \forall k \in K \\
x_{k,\theta_k^d} & \geq t_k^{rt} \quad \forall k \in K \\
x_{k,n} & \leq \sum_{l \in L} \sum_{s \in S_l} \sum_{d' \in D_k} y_{k,l,s,d'} \left( t_{ls}^{dep} + H(t_k^{rd} + d' - 2) - t_n^{yard} \right) \\
& \quad \forall k \in K, n \in N: n \neq \theta_k^d \\
x_{k,n} & = \sum_{l \in L} \sum_{s \in S_l} \sum_{d' \in D_k} y_{k,l,s,d'} \left( t_{ls}^{dep} + t_{ls}^{dur} + H(t_k^{rd} + d' - 2) \right) \\
& \quad \forall k \in K, n \in N: n \neq \theta_k^d
\end{align*}
\]
The objective function proposed (1) minimizes the total time rail cars spend in the system (w.r.t. the specific rail car release and delivery times) considering all transportation requests. We let constraint (2) force the time of availability of rail car $k$ to equal the corresponding release time. Constraint (3) specifies that a rail car $k$ may not arrive to its final destination later than a specific time, which can be assigned a value given certain customer commitments. In the same way constraint (4) ensures that a rail car cannot arrive to its destination earlier than its release time. Constraint (5) specifies that a rail car cannot be transported using an outgoing service from a certain node before it has arrived at that same node and has been subject to the corresponding shunting delay. Constraint (6) specifies that a rail car $k$ arrives at certain node $n$ at a time which equals the arrival time of the selected service $s$. Constraint (7) states that each rail car $k$ can at most use one service $s$ per link $l$. Constraint (8) is a subtour elimination constraint which restricts the sum of services used to transport a rail car $k$ to a maximum, which equals the number of nodes of the service network minus one. Constraint (9) ensures that a service $s$ cannot be used a certain day if it is not scheduled for operation during that same day, while constraint (10) and (11) define the capacity restrictions of each service $s$ with respect to weight and length, respectively.

The objective function of an optimization-based routing and scheduling policy might not only be to minimize the overall transportation times for rail cars, as modeled by objective function (1). Other objectives could e.g. be to minimize shunting activities alternatively to make better use of the fleet while still satisfying customer commitments. With this in mind, we have evaluated a set of alternative objectives of the proposed optimization-based policy with respect to a set of key performance measures. In total, four policies have been applied. We denote the initially proposed MIP formulation which minimizes the transportation times, as policy 1. In policy 2, we instead minimize the total number of required shunting activities. We also investigate the possibility to reduce the service frequency (i.e. the number of train departures per week) and define two policies which minimize the overall required service frequency; policy 3 and policy 4. In policy 3 it is only allowed to reduce the number of service departures with respect to the current operating plan, while policy 4 allows for both additional service departures and cancellation of already scheduled services. It should be noted that the departure times of services remain fixed according to the timetable and only the number of departures may be varied. The latter two formulations then achieve similar characteristics as some SND formulations with the exception of how the time is represented. Important to notice is that all
policies are constrained by the agreed delivery time requirements. We evaluate the effect of using each policy and make a comparison with the effect of using the base line policy (in the benchmark referred to as policy 0). The additional objective functions adopted in policy 2, 3 and 4 are formulated as follows:

\[
\text{minimize } \sum_{k \in K} \sum_{l \in L} \sum_{s \in S_l} \sum_{d \in D_k} 2y_{k,l,s,d'}
\]

In (14) where we aim to minimize the shunting activities, we multiply by a factor two to compensate for the required set-out and pick-up activities resulting from a single transport activity.

\[
\text{minimize } \sum_{l \in L} \sum_{s \in S_l} \sum_{g \in T} u_{l,s,g}
\]

\[y_{k,l,s,d-t_{k}^{rd}+1} \leq u_{l,s,((d-1) \mod T+1)}\]

\[\forall k \in K, l \in L, s \in S_l, d \in D; t_{k}^{rd} \leq d \leq (t_{k}^{rd} + f)\]

\[u_{l,s,g} \in \{0,1\} \text{ for all } l \in L, s \in S_l, g \in T\]

In policy 3, we replace objective function (1) with (15) and add a new decision variable, \(u_{l,s,g}\) which represents the availability of service \(s\) on link \(l\) for each weekday \(g\) \in \(T\), where \(T = \{1..7\}\). Thereby we also add constraints (16) and (17). Note that we keep constraint (9) to ensure that services which are currently unavailable still remain unavailable when the service frequency becomes variable in policy 3. This is motivated by the practical restrictions in the European railway networks. Especially in terms of the timetabling process, where it is possible to cancel already planned train departures but significantly more complex to arrange what is referred to as ad-hoc train slots based on potential excess capacity in the railway network. The modulo-operations are used to force schedules to become periodic (on a weekly basis). In the formulation corresponding to policy 4, we drop constraint (9) to enable for full modification of the service availability. The main structural differences of the formulation of the policies are presented in Table 2.

**Table 2. Overview of the policy formulations and their main structural differences.**

<table>
<thead>
<tr>
<th>Formulation</th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Policy 3</th>
<th>Policy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective function:</td>
<td>(1) (transportation times)</td>
<td>(14) (shunting activities)</td>
<td>(15) (service frequency)</td>
<td>(15) (service frequency)</td>
</tr>
<tr>
<td>Subject to:</td>
<td>(2)-(13)</td>
<td>(2)-(13)</td>
<td>(2)-(13), (16), (17)</td>
<td>(2)-(8), (10)-(13), (16), (17)</td>
</tr>
<tr>
<td>Decision variables:</td>
<td>(x_{k,m}, y_{k,l,s,d'})</td>
<td>(x_{k,m}, y_{k,l,s,d'})</td>
<td>(x_{k,m}, y_{k,l,s,d'}, u_{l,s,g})</td>
<td>(x_{k,m}, y_{k,l,s,d'}, u_{l,s,g})</td>
</tr>
</tbody>
</table>

**5. Outline of experiments**

The experiments and the problem instance used build on real data provided by the Swedish rail freight operator Green Cargo. The network in focus is a limited, yet significant, part of the service
network of Green Cargo and consists of four shunting yards with inter-connected railway lines as depicted in Fig. 3. The service schedule for the trains and the corresponding fleet characteristics are defined according to the operating plans of Green Cargo and can be overviewed in Table 3. Note that all services flow in the direction of yard 1 towards yard 4, with the exception of the service on link 5 which flows from yard 4 towards yard 3. This particular service is in practice the same train service that operates on link 4 between yard 1 and 4, but we have chosen to model it as two separated services where the dependencies instead are modeled by certain constraints.

Table 3. Train schedules and fleet characteristics.

<table>
<thead>
<tr>
<th>Train services (l-s)</th>
<th>Service capacities (tons; meters)</th>
<th>Service availability (weekdays)</th>
<th>Departure times (hh:mm)</th>
<th>Arrival times (hh:mm)</th>
<th>Durations (hh:mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 1</td>
<td>1200;630</td>
<td>Mon – Fri</td>
<td>13:51</td>
<td>15:04</td>
<td>01:13</td>
</tr>
<tr>
<td>1 – 2</td>
<td>1200;630</td>
<td>Mon – Fri</td>
<td>14:47</td>
<td>16:03</td>
<td>01:16</td>
</tr>
<tr>
<td>2 – 1</td>
<td>1200;630</td>
<td>Sat</td>
<td>08:36</td>
<td>12:20</td>
<td>03:44</td>
</tr>
<tr>
<td>2 – 2</td>
<td>1400;630</td>
<td>Mon – Fri</td>
<td>09:44</td>
<td>18:20</td>
<td>08:36</td>
</tr>
<tr>
<td>2 – 3</td>
<td>1400;630</td>
<td>Mon – Fri, Sun</td>
<td>19:36</td>
<td>23:00</td>
<td>03:24</td>
</tr>
<tr>
<td>2 – 4</td>
<td>1400;630</td>
<td>Mon – Fri</td>
<td>20:36</td>
<td>00:03</td>
<td>03:27</td>
</tr>
<tr>
<td>2 – 5</td>
<td>1400;630</td>
<td>Mon – Fri</td>
<td>23:34</td>
<td>02:32</td>
<td>02:58</td>
</tr>
<tr>
<td>3 – 1</td>
<td>1600;630</td>
<td>Mon – Fri</td>
<td>04:36</td>
<td>05:49</td>
<td>01:13</td>
</tr>
<tr>
<td>3 – 2</td>
<td>1600;630</td>
<td>Mon – Fri</td>
<td>06:11</td>
<td>08:42</td>
<td>02:31</td>
</tr>
<tr>
<td>4 – 1</td>
<td>1400;630</td>
<td>Sat</td>
<td>11:36</td>
<td>17:16</td>
<td>05:40</td>
</tr>
<tr>
<td>5 – 1</td>
<td>1400;630</td>
<td>Sat</td>
<td>19:36</td>
<td>22:40</td>
<td>03:04</td>
</tr>
</tbody>
</table>

The problem instance involves a total of 2583 single rail cars (i.e. transportation requests) which are to be distributed using the set of available services. See Table 4 for an overview of the accumulated amount of rail cars. Each transportation request is unique and represented by one rail car that has a release day and release time during the first 40 days. It then arrives at its destination maximum four days later (i.e. latest day 44). The complete time horizon is consequently 44 days, but we focus on the rail cars released during the first 28 days (i.e. four weeks). The remaining rail cars represent that the transport system continues to work as normal after the four weeks have passed. The associated freight volume in terms of length and weight is defined for each rail car. In total, 48 shippers and 76 consignees can be related to one, alternatively several, of the modeled transportation requests.

Table 4. Accumulated transportation requests (i.e. rail cars) by OD-pairs of the service network.

<table>
<thead>
<tr>
<th>(Origin - Destination)</th>
<th>Yard 1</th>
<th>Yard 2</th>
<th>Yard 3</th>
<th>Yard 4</th>
<th>Accumulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yard 1</td>
<td>-</td>
<td>536</td>
<td>30</td>
<td>242</td>
<td>808</td>
</tr>
<tr>
<td>Yard 2</td>
<td>-</td>
<td>-</td>
<td>1599</td>
<td>18</td>
<td>1617</td>
</tr>
<tr>
<td>Yard 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>158</td>
<td>158</td>
</tr>
<tr>
<td>Yard 4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Accumulated</td>
<td>-</td>
<td>536</td>
<td>1629</td>
<td>418</td>
<td>2583</td>
</tr>
</tbody>
</table>

All transportation request specific parameters (such as origin, destination, release time, weight and length) have all been set according to historical data. However, since the service network considered here is a part of a larger railway network, the arrival times of rail cars to our service network had to be adjusted accordingly. This was done using the commercial simulation software MultiRail in which simulations was performed to derive a detailed trip plan for each rail car (representing how the trip
plans would have been established in practice). We use historical transportation request data spanning over a four-week period (28 days) as input to the simulation software. The resulting trip plan for each request specifies set-out and pick-up time stamps as well as routing information. The simulations were performed on a network representing the complete Swedish railway network and the trip plans also include local shifting activities. The data generated by MultiRail was then used to assign the release times of rail cars according to when they entered our part of the network (within our network or from outside). That is, the simulated set-out time at the first node represented in our service network became the release time of the corresponding rail car. Fig. 4 provides an overview of the accumulated number of rail cars to be transported within our network during the planning period. Observe that the planning period now spans over 40 + 4 days. This can be explained by the fact that rail cars released by MultiRail on the 28th day (i.e. the last day of the input period) has to be delivered and therefore certain additional days of transportation and delays need to be accounted for. Also capacity shortages on services during the simulation prolong the time period given, as a result from the simulation. However, during the analysis of the results, we exclude the rail cars released after the first 28 days (indicated by the dashed line in Fig. 4) as mentioned earlier. This reduces the accumulated rail car quantity from 2583 to 2241. The decision and effect of isolating the demand during the first four weeks has been given a lot of thought and discussed with Green Cargo.

![Graph](image)

**Fig. 4.** Accumulated transportation requests per OD-pair and release day.

Due dates for latest possible set-out at final destinations are defined based on a *standard condition* at Green Cargo. The condition implies that rail cars are to be delivered no later than 60 hours counting from midnight after which the first distribution activity was initiated. This type of generalized delivery requirement has also previously been adopted by e.g. to Anghinolfi et al. [30].

Minimum shunting times are in our case required at each yard before rail cars are allowed to depart on outgoing services; yard (1) 10 minutes, yard (2 – 4) 140 minutes. Yard 1 has a significantly lower shunting time and is to be considered a small terminal where only minor set-out and pick-up activities are performed. However, exceptions apply for rail cars that enter the service network on
continuing services and these cars will therefore be excluded from the shunting requirement. To enable this kind of exception, we introduce an exception parameter $E_{k,n}$ for rail car $k$ at node $n$ which is set to one if no shunting is required (0, otherwise). An example would be rail cars with destination yard 2 that enters yard 1 on a service which in practice continues further in the network, e.g. service (1 – 1). These rail cars will be forced to stay on that same service until arrival at yard 2 (where the service ends) and are also excluded from shunting times at yard 1.

The resulting number of continuous and binary variables for each policy can be overviewed in Table 5, where also the general formulas used to derive the number of variables are presented. Given our problem instance, the formulation of policy 1 and 2 generates a total of 152,397 decision variables.

Table 5. The resulting number of continuous and binary variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable-types</th>
<th>Number of variables</th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Policy 3</th>
<th>Policy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{k,n}$</td>
<td>Continuous</td>
<td>[X] $</td>
<td>x</td>
<td>Y$</td>
<td>10332</td>
<td>10332</td>
</tr>
<tr>
<td>$y_{k,s,d,r}$</td>
<td>Binary</td>
<td>[X] $\sum_{i \in T}</td>
<td>S_i</td>
<td>x[D_k]$</td>
<td>142065</td>
<td>142065</td>
</tr>
<tr>
<td>$u_{t,s,g}$</td>
<td>Binary</td>
<td>$\sum_{i \in T}</td>
<td>S_i</td>
<td>x[T]$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

A selection of key performance indicators (KPIs) has been defined in cooperation with Green Cargo and based on a literature review. The KPIs can be categorized to either represent measures of operations alternatively measures of traffic performance [33]. Measures of operations can be argued being of particular interest for the rail freight operators, whereas traffic measures lie in the interest of both rail freight customers and rail freight operators. Shunting activities in terms of the number of set-out and pick-up activities and service frequencies are adopted to represent measures of operations. Service frequencies are derived as the accumulated number of departures by standard week, assuming periodic schedules. Note that in policy 0-2, the operating plan and the train schedules stand as an input and the service frequency is to be considered fixed. For measures of traffic performance we adopt transportation times, dwell times and train fill rates. Transportation times are derived as the sum of distribution times (the time cars spend on services) and dwell times (the time cars spend in yards). Train fill rates are derived by considering accumulated loaded rail car volume (tons and meters) in relation to the capacity of the services.

6. Results and discussion
The results from applying each of the four studied policies are presented in Table 6 below.

Table 6. Key performance measures related to the transportation activities.

<table>
<thead>
<tr>
<th>Traffic (KPI)</th>
<th>Policy 0</th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Policy 3</th>
<th>Policy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation times ([dd]:tt:mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Minimum</td>
<td>0:02:41</td>
<td>0:02:14</td>
<td>0:02:14</td>
<td>0:02:14</td>
<td>0:02:14</td>
</tr>
<tr>
<td>- Maximum</td>
<td>5:00:01</td>
<td>2:14:24</td>
<td>3:05:46</td>
<td>3:05:52</td>
<td>3:06:52</td>
</tr>
<tr>
<td>- Average</td>
<td>0:14:19</td>
<td>0:09:31</td>
<td>1:07:52</td>
<td>1:08:37</td>
<td>1:04:49</td>
</tr>
<tr>
<td>Dwell times ([dd]:tt:mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From Table 6 we observe a reduction in accumulated transportation times comparing policy 1 and policy 0. The reduction is of such magnitude that on average each rail car experiences a reduced transportation time of more than four hours. Certain reduction in transportation times could be argued natural due to the objective of policy 1. Especially when considering the ratio between overall transportation times and overall dwell times comparing policy 1 and 0 we observe that rail cars spend less time at yard locations under policy 1. We would like to point out that the minimum transportation time during policy (1, 2 and 3) is less compared to policy 0. The reason is that the OFRS-policies enable the service (1 - 1) for all rail cars originating from yard 1, while in practice (policy 0) this service normally is dedicated to rail cars entering the represented service network from outside the network on the certain train service which runs via yard 1. Green Cargo confirms, however, that making service (1-1) available for all relevant rail cars is practically possible and may in the future bring operational benefits. We can also observe that the flow of rail cars on service (5 - 1) is zero for all four policies. This service is in reality operated as an extension of service (4 - 1) and has been included to provide an accurate representation of the service network. Given another problem instance, rail cars could very well be transported with this service.

Under the assumption that service commitments are kept, reduced transportation times can be motivated from several perspectives. The service quality can be argued improved, since rail cars are able to arrive at customer locations given shorter time periods and consequently capital intensive resources, products and materials can be released. The profitability of the rail freight operator can be enhanced, as rail cars tend to spend shorter time in the production system, occupying valuable and limited infrastructures. In addition, overall robustness and performance improvements in the service network and in particular at shunting yards can be achieved.

For policy 2 and 3, we observe a significant increase in both overall transportation times and in average transportation times. The percentile share between overall transportation times and overall dwell times for policy 2 and 3 increases up to 89.25% and 84.96% respectively comparing 63.78% of policy 1, and rail cars tend to be postponed at origin- and intermediate locations. All transportation times are though still within the pre-defined time window which implies that the suggested trip plans are feasible and we can instead observe benefits from the policies in terms of improved performance of operations.

Results on measures of operations (see Table 7) indicate that under policy 2 reductions in the amount of required shunting activities are achieved comparing policy 0. This reduction of 172 activities can be argued to be of low magnitude, in particular if weighed against the observed increase in transportation times that follows the policy. However, the policy as such should not be neglected. Especially since the achieved result is highly dependent on the adopted service network and the existing level of service flexibility. In addition, shunting activities stand as an essential cost element which from a profitability point of view ought to be minimized. Along with increased transportation times, it is also possible to perceive benefits of policy 3 and 4 along with significantly reduced overall service frequencies. In fact, the service frequency reduces from 44 departures per
standard week (policy 0, 1 and 2) down to 28 departures (policy 3) and 26 departures (policy 4), which corresponds to a reduction of 36% and 40%, respectively. It seems that the benefits in terms of reduced service frequencies as perceived from a full modification of the train service schedule (policy 4), does not significantly exceed the benefits of only allowing cancellations of service departures (policy 3).

Table 7. Results by key performance measures of operations.

<table>
<thead>
<tr>
<th>Operations (KPI)</th>
<th>Policy 0</th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Policy 3</th>
<th>Policy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of shunting activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Accumulated</td>
<td>5348</td>
<td>5356</td>
<td>5176</td>
<td>5252</td>
<td>5036</td>
</tr>
<tr>
<td>Service frequency</td>
<td>(Fixed)</td>
<td>(Fixed)</td>
<td>(Fixed)</td>
<td>(Variable)</td>
<td>(Variable)</td>
</tr>
<tr>
<td>- Accumulated</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>- Link 1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>- Link 2</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>- Link 3</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>3*</td>
</tr>
<tr>
<td>- Link 4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>- Link 5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Service (1 – 3) eliminated, *Accumulated number of scheduled services per week.

Table 7 summarizes the weekly service frequencies per link and the number of required shunting activities by each policy. Note that for policy 0, 1 and 2 the service frequency is fixed and defined according to the current operating plan. When observing the result in Table 7 it also becomes apparent that services not only face significant reductions in departure frequencies, but for certain services also to an extent that the service is eliminated, e.g. service (3-1) and (5-1). Being able to reduce service frequencies is a valuable option for rail freight operators since operating trains is associated with noteworthy operational costs, but also since securing access to certain train slots in a shared railway network is becoming capital intensive. The number of required shunting activities also seems to decrease comparing policy 4 and policy 2. This might seem strange considering the objective of policy 2 but can though be explained by the increased service frequency on service (4-1) under policy 4, which is a direct service between yard 1 and yard 4. Percentile daily averages of train fill rates (with respect to service availabilities) for each policy can be overviewed in Fig. 5 and Fig. 6 for weight and length, respectively.
From the results presented in Fig. 5 and Fig. 6, we observe that depending on the policy adopted, and in particular the objective of the policy, different services tend to be preferred. This behavior can be highlighted by considering the two services (3 – 1) and (3 – 2). Depending on the objective of the
OFRS-policy, average train fill rates tend to vary notably. When transportation times are to be minimized (policy 1) it comes natural that services with earlier departure times are preferred. This is e.g. obvious when considering the increase in train fill rates on service \((3-1)\). Though, when the objective is set to optimize measures of operations (shunting activities and service frequencies) train fill rates are notably reduced on that same service and instead increased on service 2. Policy 4 further underlines this behavior, where the frequency of service \((3-1)\) is set to zero (i.e. the service is eliminated). Consequently, all rail cars on link 3 are routed on to service \((3-2)\) with a resulting increase in overall transportation times, see further Table 6. This trade-off is of particular importance for the rail freight operators, since eliminating services might significantly reduce costs of operations. However, the resulting service quality always has to be sufficient to be able to meet the already established customer commitments. Also services operating on link 1 indicate that the service preferred is dependent on the policy employed, where e.g. service \((1-1)\) with earlier departure time is preferred when the objective is set to minimize transportation times (policy 1). For all OFRS-policies this is, however, a permitted routing and scheduling behavior. One could though argue the necessity for modifications of the policies. Especially modifications involving multiple and weighted objectives to jointly account for both measures of traffic (e.g. transportation times) and measures of operations (e.g. service frequencies) could be of interest. Additional modifications, in terms of pure restrictions, to increase the ability to specify where service frequencies are altered could also be of interest for rail freight operators. This becomes particularly obvious when considering the elimination of service 1 (the early service) on link 3 under policy 4. Along with the reduced service frequency (i.e. the reduced number of train departures) of approximately 36% which follows from policy 3, it is also possible to observe an overall increase in train fill rates. To a certain extent the same observations can be made for policy 4. This increase in train fill rates comes natural with the reduced service frequency and since the amount of transportation requests during the period is kept constant. The increase is particularly obvious considering service \((2-2)\) and should be considered another positive effect following the policies.

Table 8. Computational performance.

<table>
<thead>
<tr>
<th>Commercial solver</th>
<th>OFRS-policy</th>
<th>Number of Simplex iterations</th>
<th>Objective function value</th>
<th>MIP-gap (%)</th>
<th>MIP-gap (abs)</th>
<th>Solution time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLPX 1</td>
<td>63029</td>
<td>1442200</td>
<td>8.89e-05%</td>
<td>128.19</td>
<td>117.7</td>
<td></td>
</tr>
<tr>
<td>CPLEX 2</td>
<td>6890</td>
<td>5870</td>
<td>0.00%</td>
<td>0.00</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>CPLEX 3</td>
<td>2727273</td>
<td>28</td>
<td>3.86%</td>
<td>1.08</td>
<td>3600*</td>
<td></td>
</tr>
<tr>
<td>CPLEX 4</td>
<td>51706450</td>
<td>26</td>
<td>9.92%</td>
<td>2.58</td>
<td>3600*</td>
<td></td>
</tr>
<tr>
<td>Gurobi 1</td>
<td>24105</td>
<td>1442200</td>
<td>9.51e-05%</td>
<td>137</td>
<td>30.4</td>
<td></td>
</tr>
<tr>
<td>Gurobi 2</td>
<td>1180</td>
<td>5870</td>
<td>0.00%</td>
<td>0.00</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>Gurobi 3</td>
<td>3004399</td>
<td>28</td>
<td>14.30%</td>
<td>4.00</td>
<td>3600*</td>
<td></td>
</tr>
<tr>
<td>Gurobi 4</td>
<td>110346922</td>
<td>25</td>
<td>24.00%</td>
<td>6.00</td>
<td>59400*</td>
<td></td>
</tr>
</tbody>
</table>

*Time limit.

The problem formulations are solved by the commercial MIP solver CPLEX 12.2. However, to be able to provide comparative measures of the computational complexity we also solve each instance using Gurobi 4.0. We use default parameter settings and the commercial software run on an AMD Opteron™ 285 Quad Core (RAM-memory 6985484 kB) in parallel deterministic mode. Table 8 presents the computational performance when using the two different MIP solvers to solve the OFRS-policy formulations. The first column specifies the number of required simplex iterations.
following columns presents objective function values and the corresponding MIP gap comparing the LP-relaxation (which for all formulations is to be considered a lower bound estimate) as well as the required solution times. Note that we include starting solutions while attempting to solve policy 3 and policy 4, the starting solutions corresponds to the optimal solution resulting from policy 1. Consequently, the initial objective function value is 42 for both policy 3 and policy 4. This service frequency is feasible and results from the current operating plan provided by Green Cargo. With respect to CPLEX, the development in terms of objective function values has been indicated in Fig. 7.

**Fig. 7.** Computational performance in terms of relative MIP-gap for policy 4 using CPLEX and Gurobi.

We find two observations of great significance. First we can notice that policy 1 and 2 are solved much quicker than policy 3 and policy 4, which after 24 hours only perceive minor improvements of the objective function value. This increased computational complexity introduced by policy 3 and 4 is interesting when considering the relatively small increase in the number of binary variables (see Table 7). The reason is probably that the values of these few additional binary variables have such a strong dependency to the other binary variables making the search space increase significantly. The progress in terms of MIP gap with respect to solution time of each MIP solver (CPLEX and Gurobi) is for policy 4 illustrated in Fig. 7. The solution progress is in line with the observations mentioned in the literature for solving similar network flow formulations using commercial software [11].

Secondly, the column representing percentile MIP gaps show that, though policy 3 and 4 already have indicated on major reductions in service frequencies, the achieved results are not to be considered computationally optimal. However, since policy 1 and 2 always can provide a feasible starting solution for policy 3 and 4, this may speed-up the solution process and every improvement might be of interest, regardless of if it is the optimal solution or not.

In our experiment, we assume complete knowledge of future transportation requests during the considered planning horizon. However, in a practical setting, this is often not the case. Our intentions
are that the OFRS-policy could be applied dynamically to update the trip plans and the train loading plans if relevant when new transportation requests arrive. In a practical setting, also the length of the considered planning horizon must be carefully selected. Furthermore, given the current industry practice, pick-up and delivery times are communicated to the customer as trip plan are being generated, so the ability to refine trip plans is therefore today limited. One possible approach to cope with this issue is to pre-define the degree to which trip plans are allowed to be refined. The degree might e.g. be customer-specific alternatively specific for the service segment.

7. Conclusions and future work
In this paper we suggest an optimization-based freight routing and scheduling (OFRS) policy to deal with the rail freight trip plan generation problem. The main intention of the policy is to provide decision support for rail freight operators at operational planning levels. Though, we are able to demonstrate that through small extensions, the OFRS-policy can be used to address also tactical planning issues. The performance of the policy, along with a set of policy extensions, has been benchmarked against the base-line policy currently adopted within the carload service segment at the Swedish rail freight operator Green Cargo. In contrast to the base-line policy, the OFRS-policy is pro-active and considers the adjustable and pre-defined overall system objective.

This paper investigates and demonstrates certain possibilities to improve the current operational planning practice, while accounting for already establish customer commitments. In particular, in terms of reduced overall transportation times, which results from reductions in rail car dwell times at yard locations following the suggested policy. Through policy extensions, this paper also elaborate on the possibility of reducing service frequencies while still being able to keep a sufficient service quality defined by already established customer commitments. Provided complete knowledge of upcoming transportation requests, we can conclude that service frequencies can be reduced with a magnitude of up to 40% and potentially even further, bringing major cost savings. Reductions of such magnitude might, however, influence the robustness of the system and the ability to serve up-coming and unexpected transportation requests. This is something that needs to be investigated in more detail. The problem instance used in the experimental evaluation serves mainly as an indicator of the performance of the OFRS-policy and further experiments are required for being able to draw general conclusions. Green Cargo has, however, confirmed that the comparison is relevant and a useful tool to assess how certain planning policies relatively affects the service quality and operational requirements. Though, what should be pointed out is the need to further investigate the implications of cancelling or inserting certain services with respect to both the locomotive and crew circulation schemes. We also want to point out that any future reformation of the current trip plan generation process and the revision of trip plans during transportation may be a sensitive topic. It requires that the customers are unaffected or made aware of and tolerating potential changes. Especially since today rail freight customers are informed of preliminary delivery dates in advance. Still we argue that the achieved results presented within this paper provide concrete incentives for using a more dynamic planning policy - such as the OFRS-policy - to enhance the current planning practices. We are aware of that a potential implementation of the policy in practice would imply that the degree of forward planning has to be reduced, and not extend over multiple weeks as in our experimental setting. Especially since the industry practice imply that transportation requests seldom are confirmed until near the execution of transportation activities, occasionally as late as during the
same day as operations are initiated. The ability to establish trip plans for up-coming transportation requests will therefore depend upon forecast accuracy as well as the customers’ willingness to communicate transportation request related information well in advance of operations. The ability to use the OFRS-policy as a decision support tool in dynamic and operational environment with a continuously updating system state also impose certain requirements on what is considered acceptable computational times. Assuming the computational complexity to be increased while attempting to solve problem instances of realistic sizes would probably require that time limits are defined to avoid extensive computationally times. At the same time, the resulting trip plans must be confirmed sufficiently close to optimal.

Improved rail car routing and scheduling is seen as one suitable attempt for being able to better match freight volumes to the set of available services. Results also argue against the nature of the current long-term planning process where service schedules are established for a standard week of operations. Instead rail freight operators should strive to increase the flexibility in operating plans. Further increased ability to reduce alternatively increase the capacity of the service network, e.g. using ad-hoc services, during operations is one suitable example which will contribute to increased operating plan flexibility.

To conclude, we have identified a number of potential directions of future research to bring further insights and improvements to the domain of rail freight planning and these includes:

- Modify the OFRS-policy to include multiple objectives and investigate how rail car specific delivery requirements influence the existing flexibility in routing and scheduling decisions.
- Elaborate on the possibility for enhanced planning of operations provided increased accuracy of forecasting methods.
- Investigate the practical possibilities to and implications of changing the service schedules on a short term basis with respect to the locomotive and crew schedules.
- Perform additional experiments involving enlarged problem instances in order to gain experience on the computational complexity of the proposed OFRS-policy and corresponding problem formulation.

Acknowledgements

Funding for this project has been received by Vinnova (the Swedish Government Agency for Innovation) and Trafikverket (the Swedish Transport Administration) through ITS-Sweden and the Swedish National Postgraduate School in Intelligent Transport Systems. The authors thank the Swedish rail freight operator Green Cargo for valuable insights and for providing us with required data. We also thank Prof. Jan Lundgren, Dr. Stefan Engevall and the three anonymous reviewers for their comments and suggestions which helped us to further improve this paper.

References


Fig. 1. Essential rail freight planning activities in the planning process (*external events).

Fig. 2. Illustration of the train service selection resulting from the base-line policy (left) and the alternative selection available if a more flexible policy would be applied (right).

Fig. 3. Illustration of the service network representation for the real network used in the experiments.

Fig. 4. Accumulated transportation requests per OD-pair and release day.

Fig. 5. Train fill rates in terms of weight presented as percentile daily average by link (l) and service (s) with respect to service availabilities.

Fig. 6. Train fill rates in terms of length presented as percentile daily average by link (l) and service (s) with respect to service availabilities.

Fig. 7. Computational performance in terms of relative MIP-gap for policy 4 using CPLEX and Gurobi.

Table 1. List of parameters.

Table 2. Overview of formulations and main structural differences.

Table 3. Train schedules and fleet characteristics.

Table 4. Accumulated transportation requests by OD-pairs of the service network.

Table 5. The resulting number of continuous and binary variables.

Table 6. Key performance measures for the transportation scheduled with the different policies.

Table 7. Results by key performance measures of operations.

Table 8. Computational performance.