Simulation of planning strategies for track allocation at marshalling yards

M. Bohlin[1], S. Gestrelius[1] & F. Khoshniyat[1,2]
[1]SICS Swedish ICT, Kista, Sweden
[2]Royal Institute of Technology, Stockholm, Sweden

Abstract

Planning the operational procedures in a railway marshalling yard is a complex problem. When a train arrives at a marshalling yard, it is uncoupled at an arrival yard and then its cars are rolled to a classification yard. All cars should eventually be rolled to the classification track that has been assigned to the train they're supposed to depart with. However, there is normally not enough capacity to compound all trains at once. In Sweden, cars arriving before a track has been assigned to their train can be stored on separate tracks called *mixing tracks*. All cars on mixing tracks will be pulled back to the arrival yard, and then rolled to the classification yard again to allow for reclassification. Today all procedures are planned by experienced dispatchers, but there are no documented strategies or guidelines for efficient manual planning. The aim of this paper is to examine operational planning strategies that could help dispatchers find a feasible marshalling schedule that minimizes unnecessary mixing. In order to achieve this goal, two different online planning strategies have been tested using deterministic and stochastic simulation. The Hallsberg marshalling yard was used as a case study, and was simulated for the time period between December 2010 and May 2011. The first tested strategy simply assigns tracks to trains on a first come-first served basis, while the second strategy uses time limits to determine when tracks should be assigned to departing trains. The online planning algorithms have been compared with an offline optimized track allocation. The results from both the deterministic and the stochastic simulation show that the optimized allocation is better than all online strategies and that the second strategy with a time limit of 32 hours is the best online method.
Keywords: railways, marshalling, marshalling yards, simulation.

1   Introduction

Railway freight plays a key role in the transportation chain for many companies, and has benefits such as low cost and low environmental impact. To improve freight transportation services, minimizing delays in the railway network is essential. There are different factors that can cause delays, but it is clear that marshalling is often a source of delay for freight in Sweden, Fakhraei Roudsari [1].

Planning the operational procedures in a railway marshalling yard is a complex problem. Currently all the classification procedures in Swedish marshalling yards are planned manually by highly experienced dispatchers. According to our investigations, there are no documented or systematic rules or guides to help operators with this planning task, and in this paper we therefore investigate how different planning strategies affect the marshalling. The Hallsberg marshalling yard, which is the largest freight yard in the Nordic countries, and arguably one of the most important marshalling yards in Sweden, is used as a case study.

The aim of this paper is to apply discrete event simulation to evaluate different planning strategies in marshalling yards with respect to efficiency and robustness. In this paper, optimizing the operational procedures with respect to efficiency is defined as decreasing the number of unnecessary car movements. Further, a planning strategy is considered robust if it generates feasible allocations with no or few missed cars, both in the deterministic simulation and when stochastic delays are added to the arrival times. The simulation is macroscopic and does not simulate the dynamic motion of wagons or the interlocking system and switches, but instead average times for task durations are used. In accordance with previous literature, we assume that any car ordering within a train is acceptable (Bohlin *et al.* [2–4]).

The paper is organized as follows. First, we give a brief overview of the problem and previous work. We then describe the methods used, including the deterministic and stochastic simulation models. We continue with an experimental evaluation and an analysis of the results. Finally there is a conclusion including suggestions for future research.

2   Background

2.1 The marshalling process in general

In general customers of rail freight transportation can be divided into two major categories. The first category contains customers that need to transport such large amounts of freight that they can buy or hire complete train sets for the transportation. These trains are called "unit trains", and all the cars in such a train will have the same origin and destination (Fröidh *et al.* [5]). Unit trains do not require marshalling. The second category contains customers that have smaller amounts of freight to be delivered, and that are interested in the transportation of individual cars rather than complete trains. Trains transporting such freight will

consist of cars from different origins and/or different destinations. These trains do require marshalling, and will travel to and/or from marshalling yards where the cars are sorted into new trains based on their destinations.

There are two types of marshalling yards: hump yards and flat yards. Most marshalling yards consist of three major sub-yards; an arrival yard, a classification yard and a departure yard. Each sub-yard has a set of tracks of different lengths. Further, hump yards have a hump between the arrival and the classification yard, and rely on gravity and switching systems to transport the cars from the top of the hump to the desired classification track (see Figure 1).
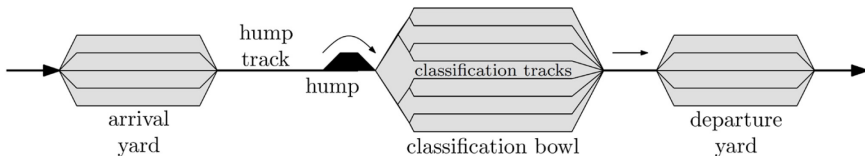

Figure 1:   A typical layout of a marshalling yard with a hump (Bohlin *et al*. [2]).

When a train arrives to a hump yard it is parked on the arrival yard, and its cars are uncoupled and the brakes released. The cars are then pushed over the hump and rolled to the classification tracks. However, before pushing the cars over the hump, a decision has to be made about which classification track each car should be rolled to. In the Hallsberg marshalling yard, when a train is being compounded on a classification track no cars belonging to other trains are allowed on that track. As a consequence, the classification yard needs at least one classification track for each departing train being compounded. Normally there is not enough capacity to compound all trains at the same time, and therefore special tracks, called *mixing tracks*, are used for cars whose trains have not yet been assigned to a classification track. This means that we have to decide when a track should be booked for a certain train, and cars arriving before the start of this booking period should be rolled to a mixing track.

Cars on mixing tracks have to be reclassified. This is accomplished by pulling the mixed cars back to the arrival yard and then pushing them over the hump again so that they may be directed to their assigned classification tracks. Pulling a car back to the arrival yard and rolling it in again is an unnecessary car movement that wears on the car and yard, and causes extra work. Therefore, the number of cars being sent to mixing tracks should be kept low.

When all cars of a departing train have arrived to the assigned classification track, the cars are coupled and the train is pulled out to the departure yard where it waits for its departure time. In the Hallsberg marshalling yard, trains can also depart straight from the classification yard.

2.2 Booking systems in Sweden

A car booking system is used in Sweden (Heydenreich *et al*. [6]). This means that when a car arrives at the marshalling yard, it has already been decided which

departing train this car should join. Booking systems give the freight operators better control over their cars and trains. However, booking systems also impose planning and operation constraints on the marshalling yards. If a booking system is not used, operators can classify a car by simply assigning it to the earliest departing train that passes through the car destination. In contrast, when a booking system is used, operators have to send each car to its predetermined departing train, even if there are other suitable trains leaving earlier. This drawback can be remedied by re-booking cars in situations where this makes sense, but this option has to be exercised with care since re-booking might affect agreements with the customers, and might also cause problems in other yards that are not expecting the car until later. In Europe, there are currently railway freight booking systems in Sweden, Belgium, the Netherlands and the Czech Republic.

2.3 Marshalling in Hallsberg

Hallsberg marshalling yard is the biggest marshalling yard in the Nordic countries. It is located in the centre of the Swedish transportation network where all the main tracks coming from Germany, Denmark, Norway and the northern parts of Sweden merge. The strategic location of Hallsberg has made the marshalling yard crowded, and optimized use of capacity is therefore of interest.

   The arrival yard in Hallsberg consists of 8 tracks with different lengths from 590 m to 690 m. The arrival yard is connected to the classification yard via a double hump, but only one hump is used at a time due to safety constraints. The classification yard has 32 tracks with different lengths from 374 m to 760 m. Finally, the departure yard consists of 12 tracks with lengths from 562 m to 886 m (Alzén [7]). A thorough description of the operations and timings of various marshalling tasks can be found in Bohlin *et al.* [2] and Alzén [7].

2.4 Optimized classification track allocation

Several mathematical programming models for finding an optimal classification track allocation have already been developed for the Hallsberg marshalling yard (Bohlin *et al.* [2–4]). The goal of the optimization models is to minimize the number of cars being sent to mixing tracks to reduce the number of car pull-backs, and in Bohlin *et al.* [4] an optimal track allocation for five days is found within 13 minutes. However, the optimizing models are complex and require a computer implementation to be used in practice. It is therefore of interest to see if less powerful but simpler rules for classification track allocation could be found, since such rules would be more easy to apply in practice.

3   Simulation

Two simulation models have been developed to evaluate the online planning strategies. The first simulation is deterministic and uses the planned arrival and departure times, while the second one introduces stochastic delays in the arrival times. Apart from the arrival times, the two models are exactly the same. GNU

Octave 3.0.5 was used to run the stochastic simulation and MATLAB R2010a was used to run the deterministic simulation. The models were partially validated by checking that the sequence of events followed the implemented rules using a visualization tool.

3.1 Input data

The case study was based on planned arrival and departure times of trains using the Hallsberg marshalling yard during the time period between December 11, 2010 and May 10, 2011, as well as car assignments for these trains. A planning horizon of seven days was used, and Saturday was chosen to be the first day of each planning period. The data was pre-processed as outlined in Bohlin *et al*. [3], and the heuristics in Bohlin *et al*. [3] were used to determine the hump schedule (initial roll-in and pull-back times) as well as the times when the newly formed trains should be rolled out to the departure yard. Therefore, in the remainder of this paper, we only consider the classification bowl operations. This reduces the allocation problem to determining which classification track the departing trains should be assigned to, and when this booking period should start. In other words, for each car that is rolled over the hump, we need to decide whether to send it to a mixing track or to a normal classification track, and if we decide to send it to a normal classification track we have to decide which one.

The output from the heuristic pre-processing in Bohlin *et al*. [3] is an ordered list of time-stamped events. The events are the following:

1. *Roll-in*: A car group (i.e. cars that arrived with the same train, and that will also depart with the same train) is pushed over the hump from the arrival yard to the classification yard. The car group needs to be directed either to its train's classification track, or to a mixing track.
2. *Roll-out:* A train in the classification yard undergoes departure preparations and is rolled out to the departure yard. All car groups belonging to the train must be at the classification track by this time. If a car group has not arrived to the track by this time it is *missed*, i.e. it will not depart with its assigned train. When a car group was missed, its identity was recorded and it was removed from the simulation. That is, we did not try to re-assign the car group to a new train.
3. *Pull-back:* All car groups on mixing tracks are pulled back to the arrival yard and rolled over the hump again to allow for reclassification.

The *mixing* tracks can be one or several tracks. In the experimental setup, two tracks with a total length of 1423 meters were reserved for this purpose. It was also assumed that the cars on both mixing tracks were pulled back at every pull-back event.

3.2 Output data

Several output variables were selected to evaluate the different planning strategies. As mentioned above, cars can miss their assigned trains. This is a planning failure, and therefore the number of missed cars is a reasonable measure of a strategy's aptness. A desired planning strategy should have no or

few missed cars. Further, the mixing tracks have a predefined capacity, and planning strategies that use more mixing capacity than available are clearly not feasible. Finally, the number of car pull-backs was counted as an efficiency measure.

3.3 Implementing stochastic arrival times

To evaluate how well the strategies cope with delays, random arrival times based on empirical data were generated. Although both early and late arrival times were sampled, only delays were propagated to the roll-in times used by the simulation.

3.3.1 Arrival time distributions
The variations in arrival times were sampled from an empirical distribution. The data consisted of measurements for two months, September and October 2008 and was taken from the Swedish train delay statistics database, TFÖR (Lindfeldt [8]). Extreme data points where trains had been more than 1000 minutes early or late were omitted. The cleaned data can be seen in Figure 2 where it has been mapped as a discrete cumulative probability density function. The variations were sampled from this density function.
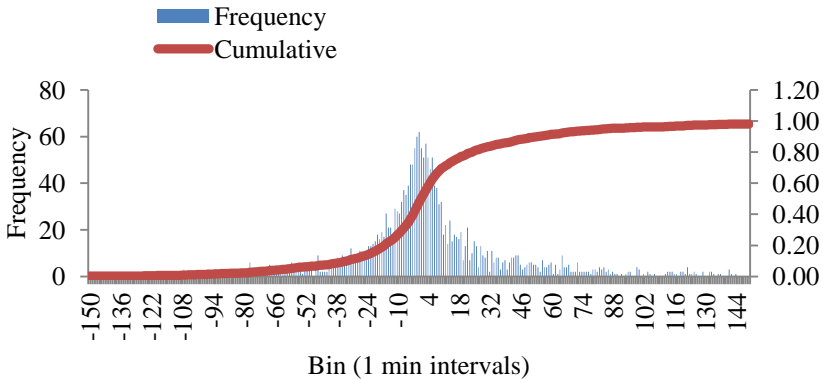

Figure 2:   The number of trains and C.D.F for the variation in arrival times, Lindfeldt [8]. A negative value means the train was early and a positive value that it was late.

3.3.2 Roll-in times
As the roll-in times are different from the arrival times further processing was needed to deduce the effect the delays had on the roll-in events. Most arriving trains had some buffer time on the arrival yard, i.e. they were parked on the arrival yard longer than what was needed for all necessary preparation work. If the sampled delay was shorter than this buffer time, no delay was added to the roll-in time. However, if the delay was longer than the buffer time, the excess delay was added to the roll-in time. Once all roll-in times had been updated to

take the sampled delays into consideration, the event list was resorted such that the events were once again in time order. Finally a sweep algorithm was used to make sure there was enough time between roll-in and pull-back events for all necessary engine movements. When there were events with too little time in-between, the later event was simply moved back to make space for the earlier event. If needed the delay was further propagated to even later events.

4   Planning strategies

In this section we outline the two online planning strategies that were tested. The results of the online methods were compared with the results from an optimized allocation for 7 days which had been constructed using the method described in Bohlin *et al*. [4].

4.1 First come-first served strategy

The first strategy is a very simple first come-first served rule (FCFS). Every time a car group is rolled over the hump (a roll-in or a pull-back event) we check if that car group's train has been assigned to a track. If the train already has an assigned track, the car group is sent there; else an attempt is made to assign a track to the car group's train. If no feasible track is available for the train, the car group is rolled to a mixing track. If more than one feasible track is available, the shortest one is chosen.

4.2 Time limit strategy

The time limit strategy works in the same way as FCFS in many ways, but it also takes the trains' departure times into consideration. When a car group is rolled in, its designated train's departure time is checked. If the departure time is more than a certain number of hours away, the car group is mixed. But if the departure time is within the time limit, we try to assign a track to the car group's train using the same rules as in FCFS. Once again, if a track has already been assigned to a car group's train, it will be sent to that track straight away.

5   Results

5.1 Deterministic results

The allocation generated by the optimizing method in Bohlin *et al*. [4] will always be feasible and never miss any cars in the deterministic simulation. Therefore these results are omitted in this section.

   First of all it is important to realise that car groups with a departure time that is earlier than the next pull-back time will miss their assigned trains if they are sent to mixing (as they will be stuck on the mixing track until the next pull-back event). We call such cars *urgent cars*. In the time limit strategy a time limit is introduced to prevent early arriving cars from occupying a classification track

during the long wait for their trains' departure times. The aim was to free up space for trains that have prompt departure times, and thereby minimize the risk of having to mix urgent cars. However, if the time limit is too restrictive urgent cars might be forced to the mixing tracks by the time limit. Therefore finding a suitable time limit is important. Further, as more and more of a train's cars ought to be rolled in as we get closer to its departure time, prioritizing trains with prompt departure times should limit the mixing track usage.

In Figure 3 the effects of the different time limits are clearly visible; a time limit of 28 hours is too restrictive while a time limit of 40 hours is not restrictive enough. 32 hours seems to be one of the best limits as it produces an infeasible allocation in only one period, and has a low percentage of missed cars. Due to the reasons stated above, it is not surprising that the first come-first served strategy misses a lot of car groups compared to the time limit strategies. However, it is worth noticing that for generating feasible allocations, i.e. schedules that use less than the available mixing capacity, a too restrictive time limit is worse than having no time limit at all.
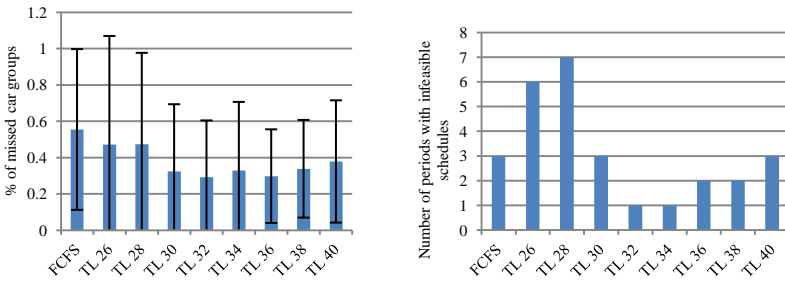

Figure 3:   *Left:* The average percentage of missed car groups in the deterministic simulation for the FCFS strategy and the time limit strategy with time limits from 26 to 40 hours. Error bars show the standard deviation. *Right:* The number of periods (out of 21) for the deterministic simulation where the strategies generated infeasible allocations.

In Figure 4 the average number of car pull-backs in the deterministic simulation is presented for the different planning methods. As can be seen the optimized schedule out-performs the other strategies when it comes to minimizing the number of car pull-backs. Further, although the maximum mixing track usage seems to be limited by setting an appropriate time limit (see Figure 3), the average number of car pull-backs decreases as the time limit is increased. This is expected as the less restrictive time limits should send fewer cars to mixing on a general basis.
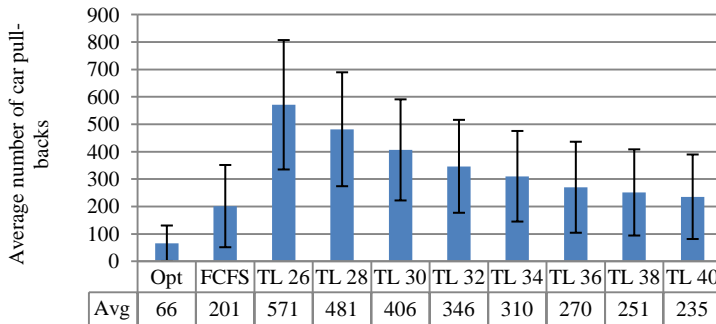
Figure 4:   The average number of car pull-backs for the different planning methods in the deterministic simulation.

5.2 Stochastic results

When the arrival times are varied it is harder to produce a schedule with no missed cars. In fact, due to delays some cars were rolled in later than their departure times, making it impossible not to miss them. In Figure 5 these results are clearly visible. However, it is worth noticing that when it comes to cars that did not arrive after their departure time, all methods missed approximately the same percentage of cars in the deterministic and stochastic runs. Most notably, the optimized allocation does not miss any cars that arrive early enough to catch their assigned trains.
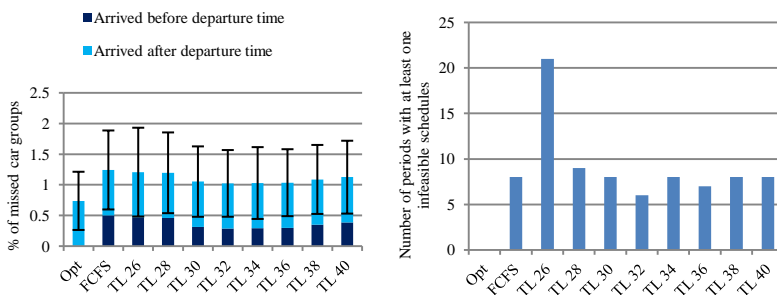

Figure 5:   *Left:* The average percentage of missed cars for the stochastic simulation for all planning methods. Error bars show the standard deviation. *Right:* The number of periods (out of 21) for the stochastic simulation where at least one of the simulation runs resulted in an infeasible allocation being generated.

As can be seen in Figure 5 the stochastic simulations resulted in an increased number of periods where at least one infeasible allocation was generated for the

online strategies, while the optimized allocations are still always feasible. Further, the stochastic arrival times seem to have decreased the average number of car pull-backs slightly (see Figure 4 and Figure 6). This might be due to the cars spending less time in the classification yard, but is probably also an effect of missed car groups being removed from the simulation. If we were to keep missed cars on the mixing tracks the average number of car pull-backs would increase for the stochastic simulation.

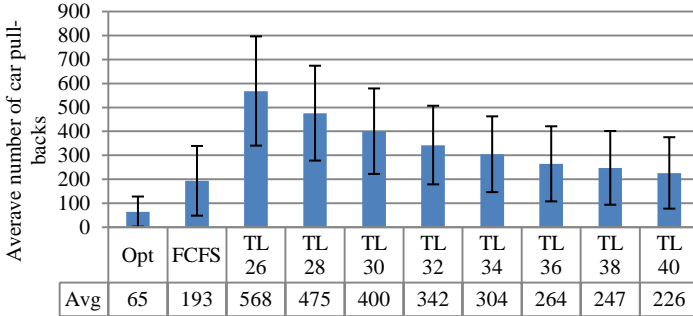|	Opt	FCFS	TL 26	TL 28	TL 30	TL 32	TL 34	TL 36	TL 38	TL 40
Avg	65	193	568	475	400	342	304	264	247	226

Figure 6:   The average number of car pull-backs for the different planning methods in the stochastic simulation.

6   Conclusions and suggested further studies

In this paper two simple online planning strategies were compared with an offline optimized classification track allocation. The Hallsberg marshalling yard in Sweden was used as a case study, and two simulations, one deterministic and one stochastic, were used to compare the different strategies. The deterministic simulation showed that the time limit strategy with 32 hours was the best online method with only one infeasible allocation and 0.29% missed cars on average. However, the optimized schedule never missed any cars nor produced infeasible allocations. Further, the optimized allocation minimized the number of extra car roll-ins, and used approximately 1/5 of the car roll-ins needed by the 32 hour time limit strategy. During the stochastic simulation runs all methods missed more cars. However, the majority of these cars were so late that they were rolled into the classification yard after their assigned trains had departed. Notably, the optimized allocation missed no cars but from the ones that were rolled in later than their departure time. Further, the number of periods resulting in infeasible allocations increased for the online methods, while the optimized allocations remained feasible in all runs. The average number of car pull-backs was reduced when stochastic arrival times were used. However, this might change if the missed cars were to remain on the mixing tracks rather than being removed from the simulation when their trains depart.

   This article presented some of the most basic planning strategies for allocating tracks in a classification yard. One of the draw-backs of the time limit strategy is that when short time limits are implemented cars are sometimes sent

to mixing tracks even though there is no pull-back event before their departure time. Including pull-backs in the strategy would hence be an interesting further development. In addition, some initial offline analysis of train lengths and expected arrival times might further improve the strategies. Comparing our results with real planning data, and making more in-depth interviews with the planning staff, would also allow us to develop and adapt our strategies.

Finally, looking at simple rules for planning the hump schedule and arrival and departure yards would be a good complement to this paper.

Acknowledgements

References

[1] Fakhraei Roudsari, F., *Spatial temporal and size distribution of freight train time delay in Sweden*. Master thesis, KTH Royal Institute of Technology, School of Architecture and the Built Environment, Division of Traffic and Logistics, Department of Transport Science, TSC-MT 11-027, pp. 27-28, 2011.

[2] Bohlin, M., Flier, H., Maue, J. and Mihalák, M., Hump yard track allocation with temporary car storage. *Proc. of the 4th Int. Seminar on Railway Operations Modelling and Analysis*, Rome, Italy, p.7, 2011.

[3] Bohlin, M., Flier, H., Maue, J. and Mihalák, M., Track Allocation in Freight-Train Classification with Mixed Tracks. *Proc. Of the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, Saarbrücken, Germany, 2011.

[4] Bohlin, M., Dahms, H.W., M., Flier, H. and Gestrelius, S., Optimal freight train classification using column generation. *Proc. of the 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, Ljubljana, Slovenia, 2012.

[5] Fröidh, O., Kordnejad, B., Kottenhoff, K., Lindahl, A., Nelldal, B. and Torche, G., *Compendium in railway traffic planning, 4th edition, KTH Royal Institute of Technology,*, School of Architecture and the Built Environment, Division of Traffic and Logistics, Railway group, pp. 282-285, 2011.

[6] Heydenreich, T. and Lahrmann, M., How to save wagonload freight. *Railway Gazette International*, p. 128, 2010.

[7] Alzén, C., *Trafikeringsplan Hallsbergs rangerbangård*. Banverket, Handbok BRÖH 313.00700, p. 11-12, 25, 32, 2006.

[8] Lindfeldt, A., A study of the performance and utilization of the Swedish railway network. *First International Conference on Road and Rail Infrastructure*, Opatija, Croatia, 2010.