WCRR BIRMINGHAM UK | 2022

A simulation framework for the operation of automated small rail vehicles in rural areas

Stephan ZIEGER[1], Nils NIESSEN[1]

[1]RWTH Aachen University, Institute of Transport Science, Aachen, Germany
Corresponding Author: Stephan Zieger (zieger@via.rwth-aachen.de)

Abstract

Rail transport in rural areas has often been negatively affected by political as well as demographic changes in recent decades. A vicious circle was initiated by the increasing individual motorisation, which created less demand, which in turn reduced the service. The end of this process is often the decommissioning or even dismantling of such regional tracks.

One possibility to reactivate these lines is the use of small highly automated vehicles, which are operated based on demand. In this way, only trips that are in demand will be performed and the residents will have the opportunity to reach destinations by rail again. This idea can also have a network effect and should therefore not be evaluated in isolation.

In addition to legal and economic aspects, operational challenges also play a major role. The annual timetable planning process is often lengthy and requires a lot of information over a long period of time. In the case considered in the paper, the requests are collected and then a timetable is created in short time, for example for the following day, which tries to accommodate as many requests as possible.

For this purpose, a simulation environment is presented in this paper, which makes it possible to transfer the trip requests into a conflict-free timetable using an insertion heuristic. The process steps are described in brief and an application example shows the possibilities of utilisation. Various Key Performance Indicators are discussed, which can be determined from the simulation.

Keywords: Demand-Responsive Transport, Railway Operations, Scheduling, Automated Railway Units, Rural Area Railway Service

1. Introduction

The shift towards greater urbanisation [2] in several countries has led to less emphasis being placed on railway lines in rural areas. In some cases, it is very difficult to sustain a railway service on these lines that is both demand-responsive and economically viable. As a result, services on many lines have been limited, e.g. from an hourly service to a service every two hours, or even disbanded. These service limitations lead to a vicious cycle which results in a higher modal share of individual transport, i.e. often by car, and consequently to further reductions in the usage of the rail service.

The work addresses the timetabling problem with small highly automated demand responsive rail units. Such units are planned already in a test phase in France for example [3]. These are not supposed to run on a cyclic basis but depending on how the passenger's requests are set up. The users of the system specify their desired origin and destination as well as the departure time for the upcoming transport period, e.g. the next day. The operator must then process the requests and create a cost-optimised timetable according to the constraints such as infrastructure or vehicle capacity. Since optimal solutions can only be found for small instances [10] due to the high complexity of the system and sheer size of possible combinations, a simulation framework is used to compare different operating strategies and to draw conclusions concerning various Key Performance Indicators (KPIs) such as number of vehicles, utilisation or mean transportation and waiting time.

2. Preliminaries

2.1 Demand-Responsive Rail Transport in Rural Areas

Demand-Responsive Transport [1] focusses on the needs of the passengers when using fleets of vehicles to pick-up and drop-off them according to their requests. This type of operation is typical, for rural areas, but currently only for rubber-tyred transport. Very recently, the system has been subject of investigation for the railway setting [6], too, and was even put in testing [3]. The authors of this paper analysed several technical and operational requirements as well as current short-comings of Demand-Responsive Rail Transport in rural areas [9].

For the operators of such a system, the flexible requests pose a significant challenge. One way to address this is to specify that all requests must be received by a certain time for the next planning period (for example, one day) which is the focus of this paper. Then the requests are routed and scheduled. This variant is called the offline version in the literature and can be tackled in part with the help of mathematical optimisation methods [10] which yield provably optimal solutions, i.e. usually cost-minimal. In the online version of the problem, ad-hoc requests are also processed and integrated into the current schedule. In this case, either the variant that the previous plan is fixed and requests are only processed if they do not lead to any changes, or a higher degree of flexibility even allows minor changes to be made to the original schedule. In order for the system to be (economically) feasible at all, ride-sharing must be implemented almost compulsorily, which means that the efficacy of the system can be increased.

2.2 Discrete-Event Simulation

Applying Discrete-Event Simulations (DES) to railway systems has a long tradition and many attempts were already made in the 1980s [5] although the expressiveness and computational power did not allow for the same standard modern simulations permit. It has to be noted that simulations aim to predict the behaviour of a system according to a specific input set [7], but do not aim to find a provably optimal solution for a problem. As simulations produce different results when the input changes, they are very useful tools to answer "what-if" scenario questions [7].

The operation of railway systems is an interplay of many puzzle pieces and thus DES are often used to describe the behaviour of larger and more complex systems. One of the most fundamental pieces in railway systems is the simulation of interlocking and train movements, see e.g. [4]. Depending on the required precision of the use-case the models can range from rough estimates to very detailed descriptions of reality. The focus in this paper will be on timetable construction, routing and evaluation of the impact on the operation and infrastructure.

3. Implementation

3.1 Constraints

The proposed implementation on rural railway lines is subject to several constraints. Firstly, the infrastructure in rural areas often consists of or might be completely composed by single-track sections with stations of a certain size in between. Thus, a capacity constraint exists for stations and lines which is determined by the number of available tracks. If a vehicle uses a single-track section in one direction, no other vehicle can travel in opposite direction at the same time. Therefore, separation times have to be taken into account. When travelling in the same direction arbitrarily chosen headway times such as 1 minute are assumed reflecting on potential changes in infrastructure and vehicle communication. Furthermore, the vehicles have a passenger capacity which allows for ride-sharing, but must not be exceeded. Finally, passengers can specify timings for departure or arrival for their journey which have to be obeyed. This is reflected by departure and arrival time windows of a certain length, e.g. 10 minutes, in which the passengers have to be picked-up or dropped-off. For each passenger the shortest path between origin and destination is calculated as well. When a detour factor, e.g. 1.2 (120%), is given, then the ride of the passenger is not allowed to exceed the detour factor times the length of the shortest path.

3.2 Infrastructure Graph

The existing infrastructure is often abstracted by a representation using graphs. Different layers of accuracy can be used. These range from a microscopic representation with all infrastructure elements and details to a macroscopic representation, which is more suitable for representing larger networks as it summarises many elements. The distinction is illustrated in Figure 1.

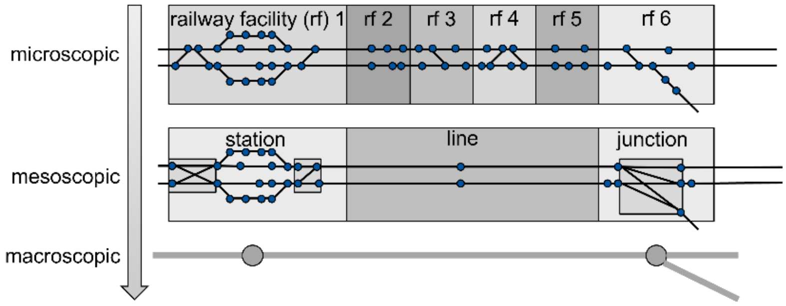

Figure 1: Detail levels of the infrastructure abstraction from microscopic to macroscopic level

The vertices and edges can be attributed with a range of characteristics. The vertices are for example usually attributed with the station capacity for rather macroscopic models or dwell times for vehicles. The edges often have the capacity, i.e. single- or double-track, or the running times as an attribute. Thus, for mesoscopic models, which abstract but still take some more details into account, details of the signalling system are exemplarily abstracted into running times and fixed headway times. These values should be chosen in such a way that zooming into a microscopic model again still leads to a feasible solution [8].

3.3 Procedure

In the following subsections, the general procedure of the simulation is described. Due to the space limitations not all details can be discussed, but the working principle and challenges should become clear. Figure 2 presents the general flow and explanations follow below.

Figure 2: General steps of the procedure

3.3.1 Setup

In the setup stage several tasks have to be performed. First of all, an infrastructure graph has to be generated, i.e. the vertices and edges as well as their incidences have to be produced. Following, the vertex attributes, i.e. capacity, dwell and drive through time, and edge attributes, i.e. capacity and running time, are assigned. The other points are the determination of the investigation period – usually one operating day – and the number of vehicles. In principle, either the number of vehicles is given and the objective is to satisfy as many requests as possible or the requests are given and the amount of vehicles to address all, or a certain percentage, of them is asked for. In this paper, the former is the case. Next, a detour factor has to be specified which determines how much the shortest path between origin and destination of a request can be extended without dissatisfaction of the passenger. Finally, a maximum waiting time for the passengers is to be determined.

3.3.2 Request Generation

The time is discretised to, for example, 30-second or 1-minute increments. Subsequently, a probability for every request is determined. Then all time steps in the investigation period are reviewed. If a request pops up for this time step, the origin, destination and the desired start time are saved in the job queue. The distribution of the requests can be freely defined to reflect either spatial (transportation hubs) or temporal (morning and evening traffic peaks) fluctuations in demand, for example.

3.3.3 Insertion Algorithm

In the algorithm, each vehicle is processed individually, so it is an asynchronous simulation, which should keep the conflicts between the vehicles as low as possible. The advantage over a synchronous simulation is further upcoming requests can already be considered when making a decision and so the number of requests served by each vehicle is aimed to be maximised. The job requests can be sorted in ascending order, so that it makes sense to process them one after the other. Each vehicle aims to include as many requests as possible in the timetable. The local decisions can have a considerable influence on the overall performance of the system. On the one hand, the requests could be pre-sorted according to certain criteria. On the other hand, the simplest method is to include the requests in the schedule at the earliest possible time or to try all combinations and take the lowest cost one. The second approach is obviously very computationally intensive. Figure 3 shows an example of how the insertion into the already existing schedule can work.

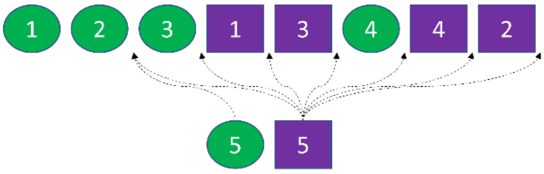

Figure 3: The trips from origin (circle) to destination (rectangle) of requests 1-4 have already been planned and are feasible. Now, the fifth job is checked for potential insertion. For the first two positions, the insertion has already been checked for feasibility. Now, if the origin of request 5 is inserted between the origins of request 2 and 3, the arrows mark the potential positions for the insertion of the destination of request 5.

As the sequences and occupancy times change constantly during the insertion, these assignments are not yet recorded, but only tested for the fulfilment of the constraints. These are first and foremost the absence of conflicts in the timetable, i.e. that no infrastructure capacities are exceeded and minimum headway times are respected. In addition, there is a maximum waiting time for the requests, which may not be exceeded during the pick-up if the request is to be accepted. Furthermore, the detour between the origin and destination of a request must not be too large. Finally, the capacity of the vehicles must be taken into account, so that requests cannot be bundled in unrestricted fashion, but the number is limited by the size of the vehicles.

3.3.4 Timetable Generation

The last step of the insertion heuristic results in the jobs that the vehicle can carry and their corresponding start times. These are now converted into the actual occupancy of the infrastructure (stations and sections). Afterwards, the insertion heuristic can be executed for the next vehicle, which now has new information and is subject to more restrictions. Therefore, the first vehicles usually handle significantly more requests than the following vehicles. These then only have requests in the job queue that are difficult to combine and can no longer occupy the infrastructure so freely, so that many opportunities for travel are blocked and no longer exist.

3.3.5 Key Performance Indicators (KPI)

In order to be able to use the output of the simulation, KPI must be defined which are as meaningful as possible for drawing conclusions for the implementation. Depending on the objective, these variables can be weighted differently in the evaluation, as the area of conflict between operator and passenger often leads to opposing interests. For example, the operator wants to transport as many passengers as possible in one vehicle, whereby the travel times of the individuals are of little relevance. The passengers, however, generally want to arrive as quickly as possible and accept only rather small detours.

These considerations lead directly to two targets: the individual travel time of passengers and the occupation ratio of the trains. Another objective for operators could be to minimise costs, which means that as few vehicles as possible cover as short a distance as possible while serving the demand. Furthermore, the number or proportion of requests handled plays a major role in the evaluation. These can also be weighted with the respective distance travelled.

4. Application Scenario

As an application example, the infrastructure of a typical line in a rural area is chosen. There is usually a major hub either at the end or, as here, in the middle of the line. Figure 4 shows this example with some double-track sections in the middle of the line. In the outer branches there are only single-track sections. Furthermore, station and section occupations are also shown for an exemplary simulation run.

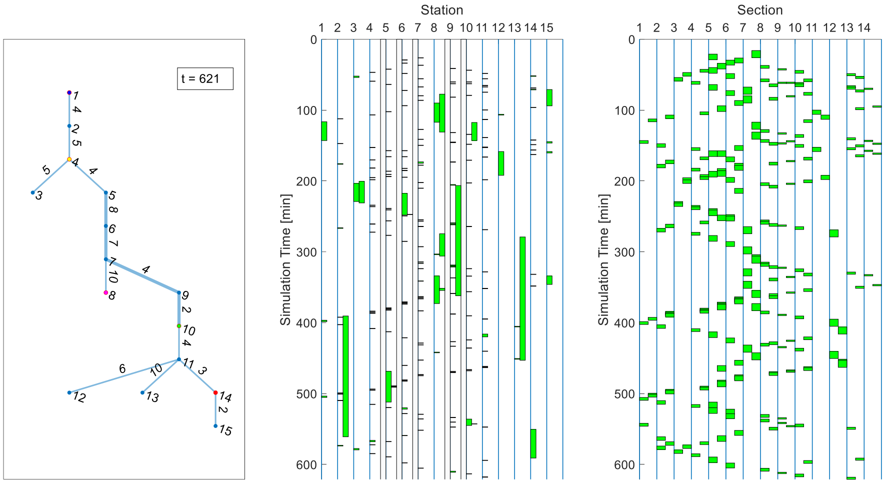

Figure 4: Left: Example infrastructure graph with single-track sections (thin edge) and distances. Middle: Station capacity (up to three tracks) usage for all vertices. Right: Section capacity (single- or double-track) usage for all edges.

One use case of the simulation is to evaluate the KPIs for a set of random requests, i.e. the investigation of robustness for the proposed service. For this purpose, 365 different days with fluctuating demand are simulated with 4 different strategies. The strategies are:

1. All requests are inserted as early as possible. The job queue is sorted by departure time in ascending order.
2. All requests are inserted as early as possible. The job queue is shuffled.
3. All possible insertions are checked. The insertion with the lowest induced costs, i.e. driving times of the vehicle, is chosen. The job queue is sorted by departure time in ascending order.
4. All possible insertions are checked. The insertion with the lowest induced costs, i.e. driving times of the vehicle, is chosen. The job queue is shuffled.

Counter-intuitively, the second pair of strategies could be expected to produce schedules with fewer total costs, but the opposite can happen. In the first case, it's possible to transport more passengers, then the total costs might be higher. Secondly, the choice of local optima can lead to higher global costs as turned out in some instances.

The simulation is run on the above infrastructure. The investigation period is 12 hours and 6 vehicles are available in service. The probability for a request is uniformly distributed with a 5 % chance for each minute. Up to 5 passengers are uniformly generated for each request. The dwell times are either 1 or 2 minutes, passing times at a station are 30 seconds and the headway times are 1 minute assuming an advanced communication and safety system. Passengers wait for up to 20 minutes after their specified departure and allow a detour factor of 1.3. Each vehicle has a capacity of 20 passengers.

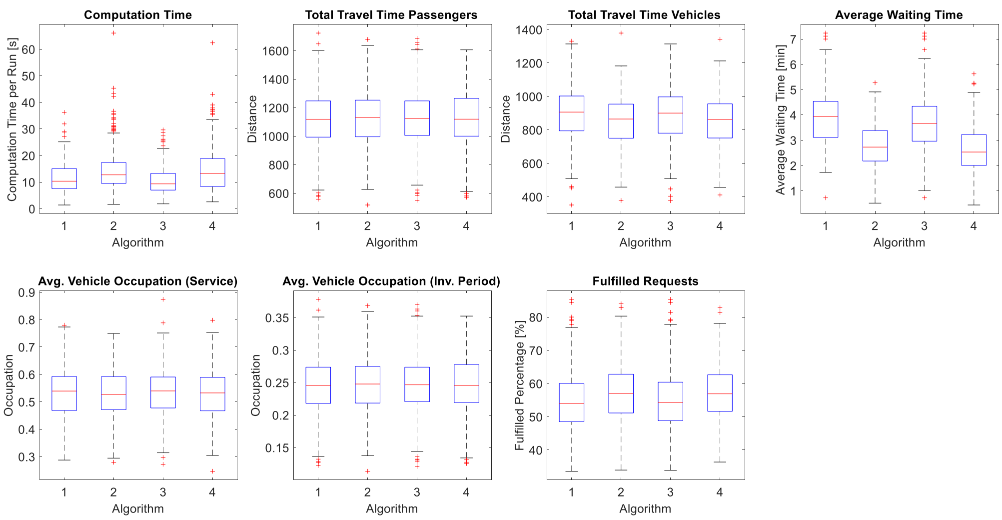

Figure 5: The output of the 4 algorithms are presented as a boxplot for the different KPIs over the 365 runs.

The results of the four algorithms appear quite similar at first glance. Contrary to expectations, the computation time of the algorithms that examine all insertions (3 and 4) is not always greater than that of the other algorithms. There is hardly any difference in the total travel times for the passengers, but algorithms 2 and 4 seem to generate lower costs for the operator. Although the sum of the passenger travel times does not seem to differ much, algorithms 2 and 4 provide passengers with schedules that are closer to their desired departure times.

The vehicle occupation is on average approximately the same for all algorithms, regardless of whether they are measured only over the service times of the vehicles or over the entire investigation period. The share of fulfilled requests, on the other hand, again differs significantly. Here, too, algorithms 2 and 4 perform better.

Subsuming for this one infrastructure, which has a large influence on the results, it can be stated that a certain portion of randomness when inserting the requests can lead to significantly better results, i.e. fulfilment of requests. In this case, it apparently makes no great difference whether local optima are searched for or not. Due to the rapidly increasing number of possible insertions, this application at least gives rise to the suspicion that the simplest method of inserting the requests at the first possible moment does not do so badly - at least in comparison with the other algorithms.

5. Conclusion

In this paper a simulation environment capable of scheduling and routing on-demand requests in an offline-setting is presented. The focus is on rural areas with dispersed spatial and temporal passenger demands in which a cyclic timetable is neither demand-responsive nor economically feasible. The simulation is meant to be a viable decision support tool for employing such services. It is highly flexible in terms of infrastructure adjustments, demand and additional constraints. Different algorithms can be used to perform the actual scheduling which is then checked for feasibility.

If necessary the infrastructure model can be further zoomed in for microscopic infrastructure considerations. The approximation of running times can be enhanced by pass-pass, stop-pass, stop-stop and pass-stop times or by means of the laws of motion. Furthermore, the algorithms should be checked for theoretical bounds which might provide a fixed approximation guarantee.

Acknowledgment

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/1.

References

[1] Brand, C.; Raje, F.; Preston, J.; Grieco, M.: *Transport and Access to Health Care: The Potential of New Information Technology, Final Report*, 2004

[2] European Commission: *Transforming Europe's Rail System*, 2020

[3] Geerts, E.: *Taxirail: the future for regional rail?* URL: https://www.railtech.com/rolling-stock/2021/07/27/taxirail-future-for-regional-rail/ - received on 27.08.2021

[4] Hill, R. J.; Bond, L. J.: Modelling moving-block railway signalling systems using discrete-event simulation. In: *Proceedings of the 1995 IEEE/ASME Joint Railroad Conference*: IEEE, 1995 — ISBN 0780325567, p. 105–111

[5] Petersen, E. R.; Taylor, A. J.: A structured model for rail line simulation and optimization. In: *Transportation Science* Bd. 16, INFORMS (1982), Nr. 2, p. 192–206

[6] Ritzer, P.; Flamm, L.; Scheier, B.; Mönsters, M.: Demand Responsive Rail Transport auf Regionalstrecken – Konzept, Machbarkeit und Wirtschaftlichkeit. In: *Eisenbahntechnische Rundschau* Vol. 70, DVV Media Group, Eurailpress (2021), Nr. 6

[7] Robinson, S.: *Simulation: the practice of model development and use*. Vol. 50

[8] Schlechte, T.; Borndörfer, R.; Erol, B.; Graffagnino, T. ; Swarat, E.: Micro–macro transformation of railway networks. In: *Journal of Rail Transport Planning & Management* Vol. 1 (2011), Nr. 1, p. 38–48

[9] Zieger, S.; Niessen, N.: Opportunities and Challenges for the Demand-Responsive Transport Using Highly Automated and Autonomous Rail Units in Rural Areas. In: *2021 IEEE Intelligent Vehicles Symposium (IV)*, 2021, p. 77–82

[10] Zieger, S.; Niessen, N.: The Dial-a-Ride Problem in Railways. In: *9th International Conference on Railway Operations Modelling and Analysis (ICROMA)*, 2021