Determining and Evaluating Alternative Line Plans in (Near) Out-of-Control Situations

Rolf N. van Lieshout[1,2], Paul C. Bouman[1], and Dennis Huisman[1,3]

[1]Econometric Institute and ECOPT, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

[2]Corresponding author. Tel: +31 10-4089663. Email address: vanlieshout@ese.eur.nl

[3]Process quality and Innovation, Netherlands Railways, 3500 HA Utrecht, The Netherlands

EI2018-20 October 10, 2018

Abstract

From time to time, large disruptions cause heavily utilized railway networks to get in a state of (near) out-of-control, in which hardly any trains are able to run as the result of a lack of accurate and up-to-date information available to dispatchers. In this paper, we develop and test disruption management strategies for dealing with these situations. First, we propose an algorithm that finds an alternative line plan that can be operated in the affected part of the railway network. As the line plan should be feasible with respect to infrastructural and resource restrictions, we integrate these aspects in the algorithm in a Benders'-like fashion. Second, to operate the railway system within the disrupted region, we propose several local train dispatching strategies requiring varying degrees of flexibility and coordination. Computational experiments based on disruptions in the Dutch railway network indicate that the algorithm performs well, finding workable and passenger oriented line plans within a couple of minutes. Moreover, we also demonstrate in a simulation study that the produced line plans can be operated smoothly without depending on central coordination.

Keywords. Railways, disruption management, line planning, dispatching.

1 Introduction

Every once in a while, railway systems suffer from very large disruptions as the result of power outages, extreme weather conditions or other severe incidents. Intensive use of the railway infrastructure and strong interdependencies between timetable, rolling stock and crew schedules cause these disruptions to easily propagate and accumulate. Since the affected number of resources can be very large and durations of disruptions are uncertain, this leads to an immensely complex problem for dispatchers. As a result, decisions made by dispatchers take a long time, and are often based on information that is already out-of-date, rendering the decision unworkable. In the end, dispatchers are confronted with a lack of accurate and up-to-date information on the current state of the system, preventing them from making viable rescheduling decisions. This can ultimately result in an *out-of-control situation*, meaning that all traffic in the affected region is terminated, even though the required resources (infrastructure, rolling stock and crew) might be available.

On the Dutch railway network, out-of-control situations happened about ten times during the period 2009-2012 because of extreme weather conditions. For this reason, Netherlands Railways (NS), the largest railway operator in the Netherlands, and ProRail, the Dutch infrastructure manager, decided to anticipate on such events by operating a reduced timetable if bad weather is expected. This reduces the probability of losing control, but the downside is that a worse service is provided to the passengers. Moreover, this does not prevent all out-of-control situations, since they also occur with completely unexpected causes, such as the power outages in Amsterdam in 2015 and 2017 and the terrorist attack in Amsterdam Central Station in 2018. In these cases, control was lost during the process of restarting operations after the forced shutdowns. As such, there is a clear need for effective countermeasures in (near) out-of-control situations.

In spite of the many recent advancements in disruption management, currently existing disruption management techniques cannot be applied in out-of-control situations due to the absence of complete information and the large number of affected trips and resources. Therefore, Dekker et al. (2018) propose a new strategy for avoiding or escaping (near) out-of-control situations. The core idea of this strategy is to completely decouple the disrupted region from the rest of the railway network. Rolling stock nor crew is allowed to change over between the two regions, isolating the disruption and preventing it from spreading. The traffic outside the disrupted region can be managed using conventional disruption management techniques, because in that region complete information is available. Inside the disrupted region, the lack of accurate and up-to-date information makes it very difficult to dispatch trains according to a centrally agreed upon timetable. Therefore, it is suggested to adjust the line plan within the disrupted region, such that it becomes possible to dispatch trains in the region using local and self-organizing mechanisms, which can be applied even when the information about the resources in the system degrades. All in all, this strategy could lead to stable operations both inside and outside the disrupted region, where the current practice simply terminates all traffic within the affected region.

In this paper, we examine whether the strategy proposed by Dekker et al. (2018) indeed provides a viable alternative in (near) out-of-control situations. In particular, we investigate how the line plan within the disrupted region should be modified. An alteration of the line plan is necessary, as the regular line plan most likely becomes infeasible due to limited turning capacity at the boundary stations and a limited availability of rolling stock in the disrupted region. In addition, we investigate how trains within this region can be dispatched without relying on central coordination. Note that by addressing these two research questions, we increase the scope of railway disruption management, which traditionally involves rescheduling the timetable, rolling stock schedule and crew schedule, by also modifying the line plan and the way the system is operated.

The contribution of this paper is threefold. The first contribution is a novel line planning algorithm that provides as many passengers with as many travel options as possible, while taking into account timetabling and rolling stock constraints in a Benders'-like approach. In our application, integration of multiple planning steps is required as we alter the line plan on the day of operations, hence it needs to be guaranteed that the line plan admits a feasible timetable and rolling stock schedule. However, the same decomposition approach can also be applied in strategic planning contexts, where it avoids having to optimize a complete timetable and/or rolling stock schedule before finding out that a line plan is in fact infeasible. The second contribution is that we propose several local dispatching strategies that can be applied in (near) out-of-control situations. The proposed strategies differ in the amount of required information and flexibility, offering a range of options to dispatchers. The final contribution is the evaluation of the produced line plans and suggested dispatching strategies by simulating them on the Dutch railway network. In particular, we illustrate that by applying the proper dispatching strategies, the modified line plans indeed lead to feasible operations, as intended.

The structure of the paper is as follows. In Section 2, the problem is described in detail. In Section 3, relevant literature is discussed. Section 4 addresses the line planning algorithm. Section 5 describes the local dispatching strategies and discusses how to test the performance of line plans and dispatching strategies by means of simulation. In Section 6, the results of both the line planning algorithm and the simulations are presented. We conclude in Section 7.

2 Problem Description

The schedule of a railway operator typically consists of the following main components. The *line plan* specifies between which stations direct trains are operated, their frequencies and the stopping patterns. The *timetable* specifies the exact departure and arrival times of train services. The *rolling stock schedule* specifies the train units that are used for each trip. Finally, the *crew schedule* specifies which train drivers and conductors perform which tasks. In a (near) out-of-control situation, all components are heavily disrupted and are required to be modified. Dekker et al. (2018) describe a new approach for dealing with these situations, taking the size of the disruption and the lack of complete information into account. Their framework is visualized in Figure 1.

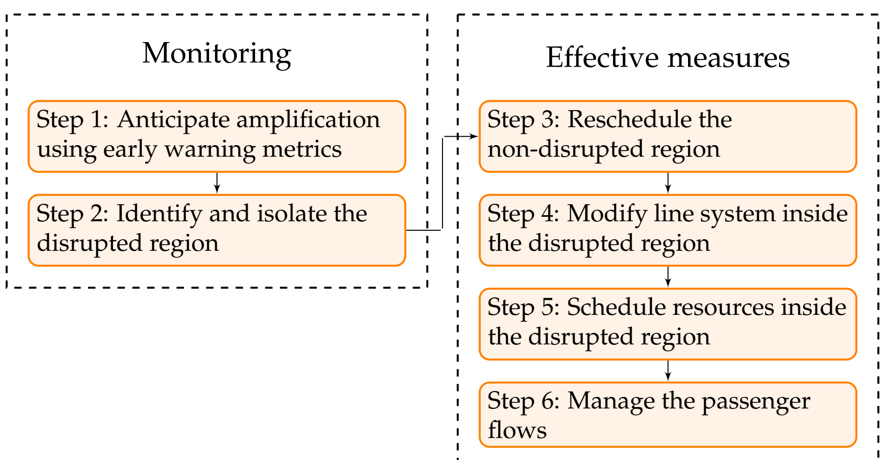


Figure 1: The framework for dealing with out-of-control situations proposed by Dekker et al. (2018).

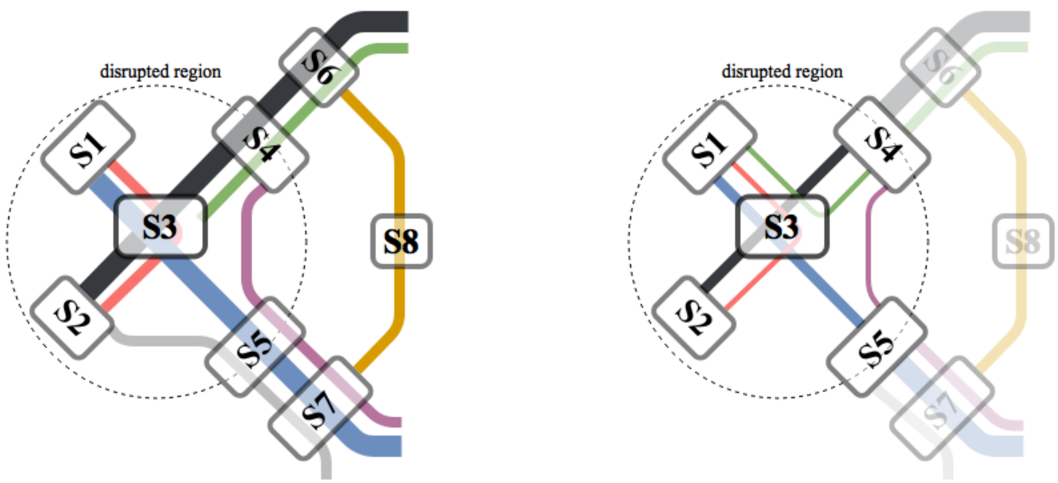
In the first two steps, a (near) out-of-control situation is detected and the *disrupted region* is identified. Once this region has been determined, no rolling stock or crew is allowed to transfer between the disrupted and the non-disrupted region, in order to encapsulate the disruption. In practice, the disrupted region often directly follows from the cause of the out-of-control situation. For example, if a temporary power outage in Amsterdam has resulted in an out-of-control situation around Amsterdam, the disrupted region will most likely consist of Amsterdam Central Station and a number of surrounding stations. The strict separation between the operations in the disrupted and non-disrupted region is maintained for the remainder of the day, as re-coupling the regions is very complex and comes with the risk of new disruptions. Over night, all resources can be set up again in order to start the regular timetable on the next day.

Besides preventing the disruption from spreading, decoupling the disrupted region makes it possible to develop tailored disruption management for both regions, which are steps 3 to 6 of the framework. Most importantly, it can be assumed that outside the disrupted region complete information is available, while inside there is not. For this reason, it is argued that the non-disrupted region can be rescheduled using conventional disruption management techniques, see Cacchiani et al. (2014) for a recent review. To operate the disrupted region, Dekker et al. (2018) propose to adjust the line plan and to schedule the resources using local, self-organizing strategies. Such strategies strongly reduce the dependence on central dispatchers and are robust against the lack of complete and up-to-date information that characterizes out-of-control situations.

In this paper, we focus on managing the operations inside the disrupted region, steps 4 and 5 of the framework. That is, we assume that a disrupted region has been identified and decoupled from the rest of the network, and investigate how the line plan within this region should be modified. Next to that, we propose and evaluate self-organizing, local dispatching strategies that can be used to operate the railway traffic in the disrupted region. To limit the scope of this research, we only consider the available infrastructure and rolling stock in the dispatching strategies and when determining the line plan, so train drivers and conductors are not taken into account. As such, we assume that the crew operating a train at the moment the disrupted region is decoupled continues to operate that train for the rest of the day.

2.1 Modifying the Line Plan

A modification of the line plan within the disrupted region is necessary, as the original line plan most likely becomes infeasible when the region is decoupled from the rest of the network. To illustrate this, consider the example in Figure 2a. When the disrupted region is decoupled, trains from both sides have to turn at stations S4 and S5. As such, only a part of the platforms at these stations can be used to serve the disrupted region, which is probably not sufficient to operate as many train services as in the original line plan. Next to that, there is only limited rolling stock available in the disrupted region, while the average proportion of time that trains are running decreases when all trains must be turned at the boundary (turning a train takes more time than a regular stop). Taking these two factors into account, one might have to settle for the alternative line plan in Figure 2b.



(a) Original line plan (b) Alternative line plan

Figure 2: Modification of a line plan. Relative frequencies are indicated by the line thickness.

As the above presented example illustrates, the challenge is to find a line plan that offers sufficient transport capacity, while ensuring that the line plan is feasible with respect to the available infrastructure and available rolling stock. As such, even though we will operate the disrupted region using local dispatching strategies instead of using a timetable and rolling stock schedule, the line plan should admit a feasible timetable and rolling stock schedule. Therefore, by considering line planning as an operational rather than a strategic problem, we are forced to integrate line planning with the subsequent planning stages.

During operations, trains using the same piece of infrastructure must be separated by a *headway time* of a certain number of minutes. This implies that every line consumes a certain amount of infrastructural capacity. In strategic planning, these restrictions are taken into account during the timetabling phase. As such, we can avoid finding line plans that require more capacity than available by considering timetabling restrictions when redesigning the line plan. However, in a (near) out-of-control situation it suffices to consider a relaxed version of the timetabling problem as trains cannot be dispatched according to a centrally agreed upon timetable due to the lack of complete information. Hence, requiring that a line plan exactly fits in a period of e.g. 60 minutes is too strong of a restriction (a line plan that fits in a period of slightly over 60 minutes likely performs equally well under the local, self-organizing dispatching strategies). For this reason, we only consider certain necessary conditions for the existence of a timetable.

As every line requires a certain number of trains and the available number of trains in the disrupted region is limited, it is essential to take rolling stock into account when modifying the line plan. The set of rolling stock compositions available in the disrupted region at the time the region is decoupled serves as input. Given the nature of out-of-control situations, we assume that it is not possible to couple or decouple units to and from a composition. Therefore, we can ignore the rolling stock compositions of the operating trains and simply refer to a rolling stock composition as a train. A complicating factor is that the exact number of trains required to operate a line plan depends on the timetable, which is unavailable. However, it is possible to find lower bounds based on the minimum running, dwell and turning times.

Although our methodology extends to the general case, we consider only two types of lines, *intercity* lines that only have stops at major stations and *regional* lines that have stops at every station. Both line types have dedicated rolling stock types. Furthermore, we assume that it is given which stations are *decoupling stations*, the stations that may serve as terminals for a line. Regional lines can be introduced between regional as well as intercity decoupling points, intercity lines can only be introduced between intercity decoupling points.

The main objective when modifying the line plan is to minimize passenger inconvenience. As the original line plan is optimized based on passenger flows, we measure the inconvenience of a passenger through the decrease in the number of travel options per period the passenger has in the modified line plan compared with the original line plan. For example, if a passenger has four travel options per hour in an undisrupted situation, ideally the modified line plan should offer this passenger four travel options per hour. However, given the restrictions it is inevitable to reduce the number of travel options for some groups of passengers.

2.2 Operating the Disrupted Region

Once the new line plan for the disrupted region has been determined, we can start executing the line plan by dispatching trains. In regular operations, trains are dispatched according to the timetable and adjustments are made to the timetable in case of disturbances or disruptions. Conversely, when operating trains in the disrupted region during an out-of-control situation, a timetable is not available. Therefore, radically different train dispatching strategies are required to operate the modified line plan. These strategies should specify simple rules that determine when trains depart. Next to that, as out-of-control situations are characterized by a lack of complete and accurate information, the strategies should be *local*, meaning that only information of the directly surrounding part of the railway network is required to decide what to do next, such that dispatching decisions can be made locally. We limit ourselves to determining departure times for trains, and scheduling the rolling stock. Train drivers and conductors are not taken into account.

3 Literature Review

Traditionally, railway disruption management involves finding a new timetable by rerouting, retiming (delaying) and canceling train services and rescheduling the rolling stock and crew such that the adapted timetable is compatible with the adapted resource schedules (Jespersen-Groth et al. 2009). Given that complete integration of these steps is computationally intractable, contributions focus on one of the rescheduling steps. Recently, researchers have started to incorporate aspects of multiple rescheduling phases in their models. For example, Veelenturf et al. (2015) consider the timetabling rescheduling problem, but also guarantee that every trip in the adapted timetable can be assigned a rolling stock composition. A second interesting development is the application of robust optimization methods. For example, Veelenturf et al. (2014) take the uncertain disruption duration into account when solving the crew rescheduling problem. For recent surveys on railway disruption management we refer to Cacchiani et al. (2014) and Ghaemi, Cats, and Goverde (2017).

To the best of our knowledge, we are the first to consider redesigning the line plan after disruptions. In strategic planning on the other hand, line planning has received a considerable amount of attention, see Schöbel (2012) for a general survey. However, existing methods are not directly applicable to our problem, since in strategic contexts computation time is of less importance and it is usually assumed that infrastructural and rolling stock restrictions can be dealt with later. In this paper however, the goal is to find a modified line plan that is feasible with respect to the infrastructure and actual available rolling stock, in real time. Therefore, we limit ourselves to discussing papers that (partially) integrate line planning with timetabling or vehicle/rolling stock scheduling.

Schöbel (2017) argues that the conventional approach of solving line planning, timetabling and vehicle scheduling problems in the traditional sequential manner can be seen as a greedy and therefore suboptimal algorithm. By considering the problem in an integrated way, one should be able to find solutions that are better for both operators and passengers. As no efficient algorithm for the complete integrated problem currently exists, the author presents a way to design new heuristics by iterating between line planning, timetabling and vehicle scheduling.

Kaspi and Raviv (2013) develop a metaheuristic to solve the integrated line planning and timetabling problem. In every iteration of their heuristic, trains are randomly scheduled based on certain distribution parameters. The resulting operator costs and passenger inconvenience is used to change the parameters for the next iteration. Burggraeve et al. (2017) also iterate between the line planning and timetable phase, but use feedback from the timetable to make deterministic changes to the line plan. In the line planning problem, constraints are included that increase the likelihood a timetable exists with large enough buffer times between trains. With this approach they are able to construct line plans that allow for timetables with larger minimum buffer times, thereby increasing the robustness. In this paper, we use a similar feedback loop connecting line planning and timetabling, but do not compute a complete timetable to limit the computation times.

Some cost-oriented line planning approaches consider aspects from rolling stock scheduling in the line planning problem. Claessens, van Dijk, and Zwaneveld (1998) and Goossens, Van Hoesel, and Kroon (2004) decide which lines to operate as well as how many carriages should be assigned to each line. Based on the driving, dwell and turnaround times, this can be used to compute a lower bound on the number of train units that is necessary to operate the line plan (the exact number depends on the timetable). A similar approach is taken by Pätzold et al. (2017), who construct line plans that can likely be operated with a small number of vehicles by only considering lines that can be operated efficiently with a fixed circulation. Specifically, a line can only be selected if a vehicle continuously performing round trips of the line has a short downtime between round trips in order to maintain periodicity. Van Lieshout and Bouman (2018) extend these ideas and compute the minimum number of vehicles required to operate a line plan, while allowing vehicles to perform circulations consisting of multiple lines. In this paper, we also consider this option as it results in more efficient use of the available rolling stock, and integrate these decisions into the line planning problem.

There are also studies on integrating line planning and rolling stock considerations into the timetabling problem. Liebchen and Möhring (2007) partially incorporate line planning by assuming that line segments are fixed and need to be matched at a certain station to constitute lines. Next to that, the authors show that for certain cases the required amount of train compositions for a line can be taken into account in the timetabling problem. Kroon et al. (2013) are able to generalize this idea. Under the assumption that arriving and departing trains are sufficiently spread, they are able to find timetables requiring the smallest possible number of rolling stock compositions.

Local dispatching strategies, which we use to evaluate the modified line plans, have to the best of our knowledge not yet been applied in railway settings. In bus networks on the other hand, there is some work on this topic. For example, Dessouky et al. (1999) simulate different strategies for holding buses at transfer stations in order to find a good compromise between missed connections and passenger delays. Berrebi, Watkins, and Laval (2015) consider the problem of maintaining stable headways and avoiding bus bunching on a high frequency line. An optimal dispatching strategy is found by formulating the problem as a stochastic decision process. As such an optimal policy requires a lot of information, we do not consider this approach viable for our problem. Instead, we use simulation to assess the performance of a number of predefined strategies, as is done by Dessouky et al. (1999).

4 Line Planning Algorithm

Our line planning algorithm rests on a decomposition of the three considered problems (line planning, timetabling and rolling stock scheduling) in a *master* problem and a *slave* problem, which can be described as the integer or combinatorial variant of Benders' decomposition (Codato and Fischetti 2006, Vanderbeck and Wolsey 2010). The master problem amounts to finding the optimal line plan subject to certain capacity and rolling stock restrictions. The line plan produced by the master serves as input for the the slave problem, which checks whether the line plan admits feasible *partial* timetables, which we define as timetables for each station independently. If the line plan is feasible, the algorithm terminates. If not, we identify a combinatorial cut in terms of the variables of the master problem. The cut is then added to the master, after which the process iterates.

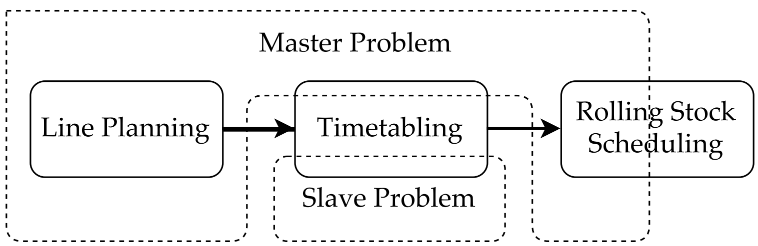


Figure 3: Decomposition of the planning problems into a master problem and a slave problem.

Figure 3 visualizes the decomposition. The master problem contains the entire line planning problem, but, as we include several necessary conditions for the existence of a timetable and rolling stock schedule, also includes parts of the timetabling and rolling stock scheduling problems. The master problem is discussed in Section 4.2. In the slave problem, the timetable feasibility of the line plan is evaluated, but since we only compute partial timetables and not a complete timetable for the whole disrupted region, not the entire timetabling problem is integrated in the algorithm. This is explained in detail in Section 4.3.

4.1 Definitions and Concepts

We represent the railway network using a connection network $G = (S, E)$. The set S contains the stations in the disrupted region and the set E contains an edge for every pair of stations between which a train can run without dwelling at an in-between station. In particular, we have *basic* *edges* connecting all pairs of stations that have no other station in-between, and *intercity edges* connecting all pairs of adjacent intercity stations, see Figure 4 for an example. We let g_e denote the original edge-frequency of edge e (the frequency at which trains are operated on the edge in the original line plan).

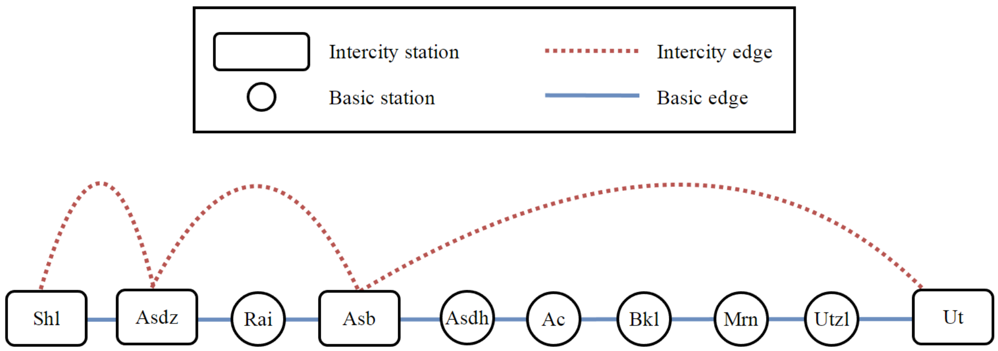


Figure 4: Connection graph of a part of the Dutch railway network.

A line l is defined as a tuple (π_l, f_l), with $\pi_l = (e_1, e_2, ..., e_m)$ a path in G and f_l an integer representing the frequency. If l is a regional line, all edges in the path must be basic edges and if l is an intercity line, all edges must be intercity edges. The terminal stations of a line must be decoupling stations compatible with the type of the line. We let S_l denote the set of stations that line l passes (with or without a stop). We assume lines are always operated in both directions.

The set Π contains all paths that can be used to form lines. To limit the number of allowed lines, we only consider paths that either are a (sub)path of a line in the regular line plan or that constitute a shortest path between the terminal stations (shortest in terms of the running times). The line pool $\mathcal{L}$ contains all feasible lines. We use subscripts to denote certain subsets of the line pool: the set $\mathcal{L}_e$ contains all lines covering edge $e \in E$, the set $\mathcal{L}_s$ contains all lines attending station $s \in S$ and the set $\mathcal{L}_\pi$ contains all lines with path $\pi \in \Pi$.

We let $\mathcal{OD} \subseteq S \times S$ denote the set of all origin-destination (OD) pairs and let $n_{o,d}$ denote the number of passengers traveling from station o to station d. We let $\rho_{o,d} \subseteq E$ denote the edges that appear in the shortest path from o to d. The shortest paths are computed using the running times on all edges. The actual shortest paths of passengers depend on the line plan and timetable, but this is a reasonably good approximation. Next to that, we let $\mathcal{OD}_{\text{dir}}$ denote the set of all OD pairs that have a direct connection in the regular line plan and let $\mathcal{L}_{o,d}$ denote the set of lines offering a direct connection between stations o and d. Finally, we let $g_{o,d}$ and $g^{\text{dir}}_{o,d}$ denote the total number of travel options per hour and the number of direct travel options between o and d in the original line plan, respectively. We also refer to $g_{o,d}$ and $g^{\text{dir}}_{o,d}$ as the original *OD-frequency* and the original *direct OD-frequency*.

4.2 Solving the Master Problem

We start by explaining the canonical form of the master problem. Thereafter, we describe the timetabling and rolling stock constraints that are included.

4.2.1 Basic Model

The binary decision variables x_l indicate whether line l is selected. Next, the integer decision variables $z_{o,d}$ represent the reduction in OD-frequency between o and d. Similarly, the integer decision variables $z^{\text{dir}}_{o,d}$ represent the reduction in the direct OD-frequency between o and d. The canonical form of the master problem is then given by:

$$
\begin{aligned}
\text{minimize} \quad & w_1 \sum_{(o,d)\in\mathcal{OD}} n_{o,d}\left(\frac{z_{o,d}}{g_{o,d}}\right)^2 + w_2 \sum_{(o,d)\in\mathcal{OD}_{\text{dir}}} n_{o,d}\left(\frac{z^{\text{dir}}_{o,d}}{g^{\text{dir}}_{o,d}}\right)^2 + w_3 \sum_{l\in\mathcal{L}} x_l && (1)\\
\text{s.t.} \quad & z_{o,d} + \sum_{l\in\mathcal{L}_e} f_l x_l \geq g_{o,d} \qquad \forall (o,d)\in\mathcal{OD}, \quad \forall e\in\rho_{o,d}, && (2)\\
& z^{\text{dir}}_{o,d} + \sum_{l\in\mathcal{L}_{o,d}} f_l x_l \geq g^{\text{dir}}_{o,d} \qquad \forall (o,d)\in\mathcal{OD}_{\text{dir}}, && (3)\\
& \sum_{l\in\mathcal{L}_e} f_l x_l \leq g_e \qquad \forall e\in E, && (4)\\
& \sum_{l\in\mathcal{L}_\pi} x_l \leq 1 \qquad \forall \pi\in\Pi, && (5)\\
& \text{(timetabling constraints)}, && (6)\\
& \text{(rolling stock constraints)}, && (7)\\
& x_l \in \{0,1\} \qquad \forall l\in\mathcal{L}, && (8)\\
& z_{o,d}, z^{\text{dir}}_{o,d} \geq 0 \text{ and integer} \qquad \forall (o,d)\in\mathcal{OD}. && (9)
\end{aligned}
$$

The objective is to minimize the weighted sum of the decrease in OD-frequency, the decrease in direct OD-frequency and the number of lines. The two latter terms are included with relatively smaller weights to favor line plans that offer more direct connections and have fewer lines. A quadratic objective is used for the first two terms to distribute the inconvenience fairly over all passengers. That is, if two OD pairs have the same original OD-frequency, we prefer to reduce both frequencies by one instead of reducing one of the frequencies by two. Moreover, if two OD pairs have a different original OD-frequency we prefer to reduce the OD-frequency of the OD-pair with the higher frequency, as this results in a lower relative decrease. As the $z-$variables take on a limited number of values, the quadratic formulation can easily be transformed into an equivalent linear formulation by defining some auxiliary variables. We provide this linearization in Appendix A.

As for the remainder of the formulation, constraints (2) and (3) make sure that the decreases in OD-frequency and direct OD-frequency are measured correctly. Constraints (4) guarantee that for every edge, the total frequency of the selected lines covering that edge is at most the original frequency of the edge (we do not want to operate more trains than in the original line plan). Constraints (5) impose that for every path, at most one line can be selected (so we cannot operate two lines with the same path but a different frequency). In the following sections, we provide constraints (6) and two variants for constraints (7).

4.2.2 Timetabling Constraints

In this section, we derive necessary conditions for the existence of a feasible timetable in terms of the line planning variables. The idea behind these conditions is that every line consumes a certain amount of the capacity at stations and a timetable can only exist if not all capacity is exhausted. The novelty of our approach is that when determining how much capacity a line requires, we take into account that in periodic timetables trains often have longer stops and turnaround times to enforce the periodic pattern. To illustrate this, consider a line between stations A and B with a frequency of 2 and a travel time between A and B of 31 minutes. Assume all lines turn on themselves, meaning that when a train arrives at its terminal station, the train turns and starts performing the reverse trip of the trip the train just finished. In a periodic timetable with a period of 60 minutes, without loss of generality trains depart at station A at minutes 0, 30, 60 and so on.

Figure 5 depicts an example of a timetable for this line. The train leaving A at 0 arrives at B at minute 31. Assuming a minimal turning time of 5 minutes, the soonest the train can again arrive at A is at minute 67. Likewise, the soonest the train is ready to depart again from A is at minute 72. However the earliest next trip it can perform starts at minute 90. This implies that on top of the minimum dwell times, we are certain the train has to dwell an additional 18 minutes. In the displayed timetable, the additional dwell time or *downtime* is spent entirely at station A, but it is of course possible to spread this time over the stations where the train dwells.

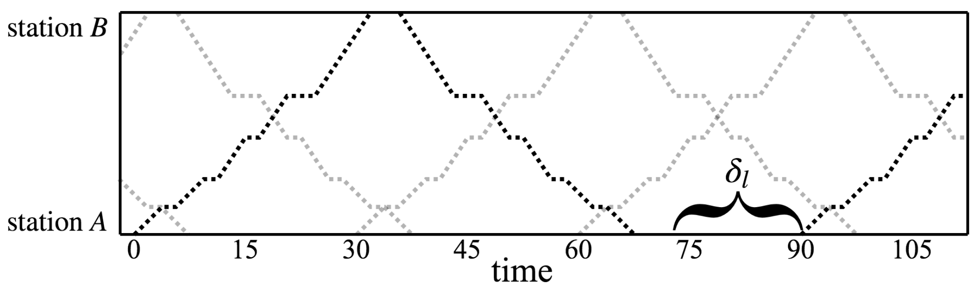


Figure 5: Time space diagram depicting the timetable of a line between A and B with frequency of 2 per hour. The downtime is denoted by δ_l.

Clearly, a timetable can only exist if there is sufficient platform capacity to serve all selected lines, taking into account that the total dwell times must be long enough to enforce the periodic pattern. To formalize this, let P^s denote the set of platforms at station s and let $P^s_{ld} \subseteq P^s$ denote the set of platforms that can be used by a train of line l arriving from direction d, taking into account infrastructural restrictions. The set D^s_l contains the directions of line l at station s. If s is a terminal station of line l, this set only contains one element, otherwise two. Next, the parameter b^s_l denotes the minimum time a platform at station s is blocked when a train from line l attends station $s \in S_l$. This time includes the headway time that needs to separate two trains using the same resource, and the minimum dwell time if l stops at s. Furthermore, we let T denote the period of the timetable and let t_l denote the the minimum time it takes a train to perform a complete circulation of line l, i.e. twice the sum of all driving and minimum dwell times. Then, the downtime δ_l of line l is the smallest positive number satisfying $t_l + \delta_l = 0 \bmod \frac{T}{f_l}$. That is, the downtime is the additional dwell time needed to enforce periodicity. The parameter $\delta^{\max}_{ls}$ indicates the maximum allowed additional dwell time of line l at station s.

To incorporate these conditions in the master problem, we declare variables and constraints that describe both the platform assignment of the selected lines and how the downtime is divided over the stations. We introduce the variables w_{lsd} representing the downtime of line l at station s in direction d. In a periodic timetable, this value is the same for every train of the line in the same direction. We also introduce the binary decision variables $y_{l_i dp}$, which are equal to 1 if the i'th train of line l in direction d is assigned to platform p, where $i = 1, 2, ..., f_l$. Then, the following set of constraints are necessary conditions for the existence of a periodic timetable:

$$\sum_{s \in S_l} \sum_{d \in D_l^s} w_{lsd} = \delta_l x_l \qquad \forall l \in \mathcal{L}, \tag{10}$$

$$\sum_{p \in P_{ld}^s} y_{l_i dp} = x_l \qquad \forall l \in \mathcal{L}, \quad \forall s \in S_l, \quad \forall i \in \{1, ..., f_l\}, \quad \forall d \in D_l^s, \tag{11}$$

$$\sum_{l \in \mathcal{L}_s} \sum_{d \in D_l^s} \sum_{i=1}^{f_l} (b_l^s + w_{lsd}) y_{l_i dp} \leq T \quad \forall s \in S, \quad \forall p \in P^s, \tag{12}$$

$$0 \leq w_{lsd} \leq \delta_{ls}^{\max} \qquad \forall l \in \mathcal{L}, \quad \forall s \in S_l, \quad \forall d \in D_l^s, \tag{13}$$

$$y_{l_i dp} \in \{0, 1\} \qquad \forall l \in \mathcal{L}, \quad \forall s \in S_l, \quad \forall i \in \{1, ..., f_l\}, \quad \forall d \in D_l^s, \quad \forall p \in P_{ld}^s. \tag{14}$$

Constraints (10) ensure that if a line is selected, the downtime is divided over the stations. Constraints (11) ascertain that the train services of selected lines are assigned to platforms. Constraints (12) impose that the total time a platform is blocked per period is less than the total available time. Constraints (13) guarantee that the downtimes are nonnegative and less than the specified upper bounds. Constraints (12) can be linearized by introducing the decision variables $w_{l_i dp} = w_{lsd} y_{l_i dp}$. This relation is enforced by adding the following linear constraints:

$$w_{l_i dp} \leq \delta_{ls}^{\max} y_{l_i dp} \qquad \forall l \in \mathcal{L}, \quad \forall i \in \{1, ..., f_l\}, \quad \forall s \in S_l, \quad \forall p \in P_{ld}^s, \quad \forall d \in D_l^s, \tag{15}$$

$$w_{l_i dp} \leq w_{lsd} \qquad \forall l \in \mathcal{L}, \quad \forall i \in \{1, ..., f_l\}, \quad \forall s \in S_l, \quad \forall d \in D_l^s, \quad \forall p \in P_{ld}^s, \tag{16}$$

$$w_{l_i dp} \geq w_{lsd} - \delta_{ls}^{\max}(1 - y_{l_i pd}) \quad \forall l \in \mathcal{L}, \quad \forall i \in \{1, ..., f_l\}, \quad \forall s \in S_l, \quad \forall d \in D_l^s, \quad \forall p \in P_{ld}^s. \tag{17}$$

4.2.3 Rolling Stock Constraints

We take rolling stock scheduling into account when redesigning the line plan by imposing that for every selected line, there must be a sufficient number of trains assigned to circulations covering to the line. We distinguish two cases. In the first case we assume that all rolling stock circulations must be *fixed*, such that all trains turn on themselves. In the second case we relax this assumption and allow for *flexible* rolling stock circulations, in which trains can switch lines when they reach their terminal station. This leads to shorter turning times, which allows more trains to be operated as (i) trains spend less time dwelling and more time running and (ii) pressure is released at the turning stations, such that there is capacity for additional trains. Moreover, flexible circulations increase flexibility during operations. As a timetable is unavailable, the number of trains required to operate circulations is determined based on the minimum trip times in both cases.

Note that the downtime constraints (10) defined in the previous section are valid under the assumption of fixed circulations. It is possible to generalize these constraints, such that the additional dwell time is not distributed over the stations of all selected lines, but over the stations of the selected circulations. However, preliminary experiments showed that this constraint is rarely violated when flexible circulations are allowed. The reason for this is that in order to use the trains as efficiently as possible, the model selects circulations with short downtimes. Therefore, we ignore constraints (10) if we allow for flexible circulations.

Fixed Rolling Stock Circulations

We let n_l denote how many trains are at least necessary to operate line l when operated solely with fixed circulations. This value can be computed as follows:

$$n_l = \left\lceil \frac{t_l}{T} f_l \right\rceil, \text{ or alternatively } n_l = \frac{t_l + \delta_l}{T} f_l. \tag{18}$$

Next, we let R denote the set of available trains located in the disrupted region. The parameter a_{rl} indicates whether train r can be assigned to line l, based on the location and type of the train. We introduce decision variables v_{rl} indicating whether train r is assigned to line l. The allocation of rolling stock can be included in the master problem by adding the following constraints:

$$\sum_{r \in R} a_{rl} v_{rl} = n_l x_l \quad \forall l \in \mathcal{L}, \tag{19}$$

$$\sum_{l \in \mathcal{L}} v_{rl} \leq 1 \quad \forall r \in R, \tag{20}$$

$$v_{rl} \in \{0, 1\} \quad \forall l \in \mathcal{L}, \quad \forall r \in R. \tag{21}$$

These constraints state that if a line is selected, n_l trains should be assigned to the line and that every train can only be assigned to a single line.

Flexible Rolling Stock Circulations

The advantage of flexible circulations is that multiple lines with long downtimes can be combined into a circulation with a short downtime. Consider for example two lines l and m with $f_l = f_m = 1$ and $t_l = t_m = 85$ minutes such that $\delta_l = \delta_m = 35$ minutes. Combining these lines in a circulation gives a travel time of 170 minutes and a downtime of only 10 minutes, which is clearly more efficient.

We formally define a rolling stock circulation c as a sequence of lines $c = l_1, l_2, ..., l_{|c|}$, such that all consecutive lines, and the first and last line, have a shared terminal station and all lines are either all regional or all intercity lines. A train performing this circulation continuously traverses the sequence from left to right: first the train performs a round trip of l_1, then a round trip of l_2 and so on. We let f_c denote the frequency of circulation c, which equals the minimum frequency of the lines in the circulation. Analogously to lines, we let t_c denote the minimum travel time of circulation c, which equals $\sum_{l \in c} t_l$, and we let δ_c denote the downtime of a circulation, which equals the smallest positive number satisfying $t_c + \delta_c = 0 \bmod \frac{T}{f_c}$.

To provide a mathematical formulation of flexible rolling stock circulations, we let $\mathcal{C}$ denote the set of allowed circulations and introduce the decision variables γ_c, indicating whether circulation c is selected, and θ_c, representing how many trains perform circulation c. We also have the assignment variables v_{rc}, indicating whether train $r \in R$ is assigned to circulation c. We let the parameter a_{rc}

indicate whether such an assignment is possible. The formulation now reads as follows:

$$\sum_{c\in\mathcal{C}|l\in c} \frac{T}{t_c+\delta_c}\theta_c = f_l x_l \quad \forall l \in \mathcal{L}, \tag{22}$$

$$l_c\gamma_c \le \theta_c \le u_c\gamma_c \qquad \forall c \in \mathcal{C} \tag{23}$$

$$\gamma_c \le x_l \qquad \forall c \in C, \quad \forall l \in c, \tag{24}$$

$$\sum_{r\in R} a_{rc}v_{rc} = \theta_c \qquad \forall c \in \mathcal{C}, \tag{25}$$

$$\sum_{c\in\mathcal{C}} v_{rc} \le 1 \qquad \forall r \in R, \tag{26}$$

$$\theta_c \ge 0 \text{ and integer} \qquad \forall c \in \mathcal{C}, \tag{27}$$

$$\gamma_c, v_{rc} \in \{0,1\} \qquad \forall c \in \mathcal{C}, \quad \forall r \in R. \tag{28}$$

Constraints (22) guarantee that if a line is selected, a sufficient number of trains is assigned to circulations covering the line. The crucial observation is that every period, a train assigned to circulation c performs $\frac{T}{t_c+\delta_c}$ trips of each line in c. Constraints (23) assure that if a circulation is selected, the number of trains assigned to the circulation is between a certain upper and lower bound. These bounds are given by

$$l_c = \left\lceil \frac{t_c}{T} \right\rceil \quad \text{and} \quad u_c = \frac{t_c}{T} f_c. \tag{29}$$

The lower bound ensures that every selected circulation accounts for at least one train service of each line in every period. The upper bound is derived from constraints (22). Constraints (24) impose that a circulation can only be selected if all lines in the circulation are selected. These constraints are not necessary for a valid formulation, but slightly strengthen the linear programming relaxation. The assignment part of the formulation is covered by constraints (25) and (26). These constraints make sure that the number of trains assigned to a circulation equals the number of times the circulation is selected and that every train is assigned to at most one circulation.

To keep the number of circulations limited, we only allow fixed circulations and circulations with two lines that satisfy $\delta_c < \sum_{l\in c}\delta_l$, i.e. the downtime of the combined circulation is strictly smaller than the sum of the downtimes of the individual lines. Doing so ensures that the size of the model does not increase too much, while still including the most promising circulations.

4.3 Solving the Slave Problem

The goal of the slave problem is to evaluate whether the line plan produced by the master problem admits a feasible timetable and to identify one or more violated inequalities or cuts if this is not the case. As explained in Section 2, it is not required to consider the complete timetabling problem since trains will be dispatched using self-organizing strategies rather than according to a timetable. Therefore, we do not try to compute a timetable for the entire network, but rather compute partial timetables for every station independently. Additional advantages of this approach are that it gives timetabling instances that can be solved quickly and that it allows us to identify small sets of inconsistent lines (and therefore strong cuts that can be added to the master problem). Conversely, computing a timetable for the entire network is very time consuming and has the disadvantage that only a single line plan is ruled out by the generated cut if the line plan is found to be infeasible.

We formalize the timetabling problem using the Periodic Event Scheduling Problem (PESP). The PESP is originally introduced by Serafini and Ukovich (1989) and can be used to model a wide variety of timetabling constraints. In the PESP we include safety, dwell, synchronizing and capacity constraints. The capacity constraints lie outside the standard formulation of the PESP but are necessary given the limited capacity of the boundary stations. We refer to the extended model as the C-PESP. A formulation of the C-PESP is provided in Appendix B. For a more extensive discussion of the (C-)PESP we refer to Peeters (2003).

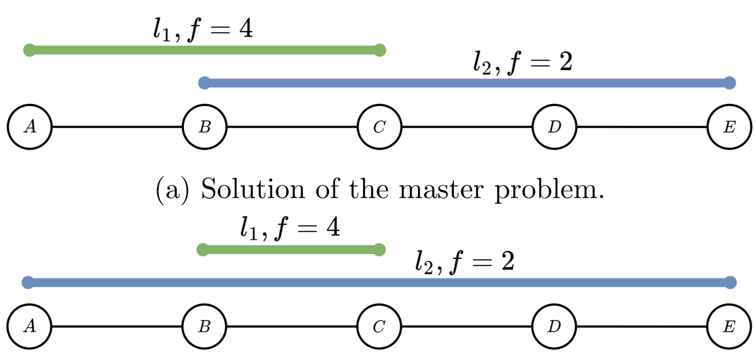
In our implementation of the C-PESP, we increase the minimum dwell times by the headway time to strengthen the capacity constraints (if both the dwell time and headway time are 2 minutes, every dwelling train effectively blocks the platform for 4 minutes). Next to that, we apply constraint propagation techniques before we split up the timetabling problem into the timetabling problems for all stations. This way, the local constraints at the stations are strengthened using constraints outside the station. Furthermore, junctions between stations and points in the network where the number of tracks changes are considered as dummy stations.

When we detect an infeasibility at a station, it is first checked whether the set of lines attending the station is a minimal inconsistent set by iteratively removing the lines in the set and checking whether the C-PESP now has become feasible. If removing one of the lines still results in an infeasible C-PESP, the line is removed from the set. This process is repeated until the set of lines corresponds to a minimal inconsistency.

To give a mathematical formulation of the cut, let $\mathscr{L}^s$ denote a minimal inconsistent set of lines attending station s in the current solution of the master problem. Then, the following cut rules out this line combination in the next iterations of the master problem:

$$\sum_{l \in \mathscr{L}^s} x_l \leq |\mathscr{L}^s| - 1. \tag{30}$$

However, this is not the only cut we can derive from the discovered infeasibility, since we know all line plans that generate a C-PESP that is at least as difficult as the C-PESP at s do not have a feasible timetable. To illustrate this, consider the line plan visualized in Figure 6a, containing the lines l_1 and l_2 with frequencies 4 and 2, respectively. Now assume this line plan leads to an infeasible C-PESP at station C. After adding the corresponding cut to the master and resolving, we might find the line plan in Figure 6b in the next iteration. However, this line plan must also result in an infeasibility at station C, as from the perspective of this station, the solution has not changed (assuming that the minimum dwell time only depends on whether trains turn or not). In other words, the C-PESPs generated for station C are equivalent. Therefore, we could have excluded both line plans upon finding the infeasibility at C.



(a) Solution of the master problem.

(b) Alternative line plan.

Figure 6: Two line plans that generate the same C-PESP at station C.

More generally, let $\mathcal{M}_i^s$ denote the set of lines that have the same frequency, the same in- and outbound edge and the same minimum dwell time at station s as the i'th line in $\mathscr{L}^s$. Then, from a station infeasibility at station s, we can derive the following cuts:

$$\sum_{l \in M} x_l \leq |\mathscr{L}^s| - 1, \quad \forall M \in \mathcal{M}_1^s \times \mathcal{M}_2^s \times ... \times \mathcal{M}_{|\mathscr{L}^s|}^s. \tag{31}$$

The advantage of adding multiple cuts per iteration is that it is likely to reduce the total number of iterations before all C-PESPs are feasible. On the other hand, the time spent solving the master problem might increase, as adding all cuts (31) increases the size of the master problem, especially when the disrupted region and/or the set $\mathscr{L}^s$ is large. In preliminary experiments, we observed that the benefits of adding multiple cuts greatly outweigh the disadvantages. Therefore, we always add all cuts (31) when running the algorithm.

5 Operating the Disrupted Region

In this section, we propose dispatching strategies that can be used to operate the line plans that are produced by the model discussed in Section 4. Next to that, we describe the simulation framework and evaluation criteria used to assess the performance of the line plans and dispatching strategies.

5.1 Train Dispatching Strategies

The train dispatching strategies that we develop specify what to do next when a train arrives at a station. Specifically, the strategies state (i) *when* the arrived train will depart and (ii) *where to* the train will depart. The information that is allowed to be used to make these decisions are previous departure times at the station and information from trains directly surrounding the station.

As for the *when* aspect of strategies, we consider three *timing principles*, referred to as ASAP, SYNC and SYNC + COOR. When trains are operated using the ASAP (as soon as possible) principle, arriving trains always leave as soon as possible. This may be reasonable, as the station capacity is limited, hence trains should not occupy platforms longer than necessary.

When trains are operated using the SYNC (synchronize) principle, trains do not depart as soon as possible at terminal stations. Instead, the departure time at these stations is decided based on the previous departure time of the line in order to promote the regularity of the departure times.

For example, if a train from a certain line with a frequency of 4 per hour arrives at the terminal station and the previous train of the line departed 7 minutes ago, the train will depart in 8 minutes. However, we do impose that if a train is unable to enter a terminal station because of a waiting train, the waiting train will depart immediately and not wait until its desired departure time, to free a platform for the entering train.

The SYNC + COOR (synchronize and coordinate) principle extends the SYNC principle by coordinating the departure times of different types of trains. Firstly, the principle imposes that if a regional train has departed from a station on a part of the network that has one track per direction, intercity trains can only depart when enough time has passed to make sure the faster intercity train does not have to wait for the slower regional train. Secondly, at stations where overtaking is possible, regional trains wait at the station if an intercity train is coming within 3 minutes and the regional train would have otherwise blocked the incoming intercity train. Note that it is also possible to take a different value than 3 minutes or let the maximum waiting time depend on the decrease in travel time of the intercity train.

As for the *where to* aspect of dispatching strategies, we consider two *turning principles*, STAT and DYN. In the STAT (static) principle, trains reaching their terminal station start performing the same line in the reverse direction. In the DYN (dynamic) principle, trains can be reassigned to a different line when they reach their terminal station. Trains are reassigned based on the type of the train and the previous departure times. The line that needs a departure the earliest gets assigned the first compatible train. The advantage of the DYN principle is that it results in shorter turning times at the terminal stations. Even more, this principle leads to more efficient use of the trains, such that it is possible to operate more trains per hour. Clearly, the STAT and DYN principles are related to the respectively fixed and flexible rolling stock constraints discussed in Section 4.2.3. If the line plan is optimized using flexible rolling stock constraints, it is expected that the STAT principle is too restrictive and will not lead to satisfactory operations. Conversely, if fixed rolling stock circulations are imposed in the line planning algorithm, it might still be beneficial to use the DYN principle, as it is more flexible.

The three timing principles and two turning principles give rise to six different strategies, with ranging degrees of coordination. For instances without intercity trains we have only four strategies, as the SYNC+COOR principle is only different from the SYNC principle in the way it deals with intercity trains.

5.2 Simulation Framework

We evaluate the line plans and dispatching strategies by simulating the railway traffic in the disrupted region. The simulation takes as input a disrupted region, the modified line plan, the regular timetable, the time instant control the disrupted region is decoupled from the rest of the network and the dispatching strategy that should be applied. Using these inputs we can retrieve the current position of the trains in the disrupted region from the regular timetable. The simulation is initialized by assigning the trains to lines. We use a basic model to perform this initialization, assigning the trains in such a way that trains assigned to the same line are roughly spread out over the network. In the case that a train not assigned to any line, it is assumed that this train can be routed to a shunt yard without interfering with other trains.

For the simulation we use a macroscopic representation of the railway network where nodes are stations and edges are tracks. Junctions are modeled as dummy stations with the number of tracks as the number of platforms. In the actual operations, tracks are subdivided into block sections and a train is allowed to enter a section if the previous train is no longer occupying the section. Otherwise, the train needs to wait before the red signal placed at the beginning of the block. In the simulation, the block sections are not taken into account and trains therefore only wait upon arriving at or departing from a (dummy) station. Between stations, trains run at a constant speed. Dummy stations are also introduced between stations to prevent faster trains from overtaking slower trains on the same track.

Trains can enter a station if there is a platform available that is compatible with the in- and outgoing track of the train. A platform is available if the predefined headway time has passed since the previous train used the platform. If multiple platforms are available, the simulation picks one at random. If there is no available platform, the train is added to the arrival queue at the station. When a train arrives at a station, the departure is scheduled according to the dispatching strategy that is being used. When a train wants to depart, the simulation checks whether the outgoing track is available for the type of the train. The outgoing track is available if the headway time has passed since the last train of the same type has departed. If the outgoing track is still blocked, the train is added to the departure queue of the track. After a train has departed, the simulation checks whether the newly available platform can be used by one of the trains in the arrival queue and an arrival is scheduled if that is the case.

5.3 Evaluation Criteria

As the main objective when optimizing the line plan is to offer sufficient travel options to the passengers, it is expected that a good execution of the line plan also leads to a good service for the passengers. Therefore, we introduce three *operational measures* that quantify how close the realized timetable is to a perfect execution of the line plan. The considered aspects in these measures are the realized frequencies of lines, the regularity of the inter-departure times and the train delays. Next to the operational measures, we also directly assess the impact of the different line plans and strategies on the journeys of passengers with two *travel measures*, taking into account the realized number of travel options per hour offered to passengers and travel times.

Operational measures. The measures are first defined for the operation of a line in a certain direction and can be computed at every departure at the associated terminal station. Later we describe how the measures can be evaluated for an entire line plan at any given time. The measures are defined in such a way that if and only if a train line is operated perfectly (i.e. trip times are at their minimum and the trains depart from the terminal stations according to a perfect synchronized pattern), the train line scores exactly 1 for all measures at all times. This allows us to clearly observe deviations from the ideal scenario.

For the sake of notation, we start counting the departures from zero. The i'th realized departure time of line l in direction d is denoted as $\text{dep}_l^d(i)$. Furthermore, we let p_l equal T/f_l, the period of a line. At the time of the i'th departure of line l in direction d, the frequency measure is given by:

$$\text{Frequency}_l^d(i) = \frac{i \times p_l}{\text{dep}_l^d(i) - \text{dep}_l^d(0)}. \tag{32}$$

This measure relates the desired time between the first and i'th departure to the realized time between the first and i'th departure. For example, if the measure is 1.1, there have been 10 percent more departures than indicated by the line plan and if the measure is 0.9, there have been 10 percent fewer departures.

The regularity measure is defined as follows:

$$\text{Regularity}_l^d(i) = 1 - \frac{\sum_{j=1}^{i} |\text{dep}_l^d(j) - \text{dep}_l^d(j-1) - p_l|/p_l}{i}. \tag{33}$$

The measure relates the cumulated relative deviation from the optimal departure pattern to the number of departures. For example, for a line with a period of 15 minutes, if the measure equals 0.8, the average absolute deviation from the ideal time between departures equals $0.2 \times 15 = 3$ minutes. If the measure equals 0.6, the average absolute deviation is 6 minutes. Relative deviations are considered such that a deviation of 10 minutes on a line with a period of 30 minutes is penalized less than a deviation of 10 minutes on a line with a period of 15 minutes.

For the delay measure, we let d_i denote the relative delay of the i'th trip. Normally, delays are computed with respect to a timetable, but since a timetable is not available, we compute delays relative to the minimum trip times, the sum of minimum running and dwell times from terminal to terminal. For example, if the minimum trip time of line l is 20 minutes, and the i'th trip takes 25 minutes, d_i equals 0.25. The delay measure is given by:

$$\text{Delay}_l^d(i) = \frac{\sum_{j=0}^{i}(1 + d_j)}{(i+1)}. \tag{34}$$

This measure equals the average normalized trip time of the line. Evidently, this measure cannot have a value below one. For example, if the minimum trip time of a line is 30 minutes, and the measure equals 1.1, the average realized trip time is $30 \times 1.1 = 33$ minutes. If the measure is 1.2, the average realized trip time is 36 minutes.

To compute the performance of a line in a direction at any given time t, we compute the measures at the last departure before t and the first departure after t and interpolate. Next, the frequency, regularity and delay measure of an entire line plan at any given t can be computed by taking the average of the measures over all lines and directions. Note that averaging over lines can result in a frequency measure of 1.00, even though the performance of the individual lines may be very bad, for example one line has frequency 1.5 and another line 0.5. However, in such a case we are still able to detect that the line plan has poor performance as both lines will have very bad scores on the regularity measure.

Travel measures. To define these measures, it is necessary to make assumptions on the paths travelers take from their origin to their destination. To ease computation, and to enable a proper comparison to the objective attained by the line plans, we assume travelers take the same path as the path that is used in optimizing the line plan (so travelers from o to d take path $\rho_{o,d}$, the shortest path based on the running times). Furthermore, to evaluate the long run average performance, we compute the measures for every whole minute during the simulation and take averages.

The first travel measure equals the average travel time of travelers entering the system at time t. If we let $\text{time}_{o,d}(t)$ denote the travel time from o to d using path $\rho_{o,d}$ for a passenger entering station o at time t, the measure is defined as follows:

$$\text{Travel Time}(t) = \frac{\sum_{(o,d)\in\mathcal{OD}} n_{o,d}\text{time}_{o,d}(t)}{\sum_{(o,d)\in\mathcal{OD}} n_{o,d}}. \tag{35}$$

Besides using this measure to compare the performance of different line plans and dispatching strategies, it also gives an indication of the overall performance of our approach by comparing it with the average travel time in an undisrupted situation.

The second travel measure corresponds to an empirical analog of the objective (1) used in the line planning algorithm and hence referred to as the realized objective. We let $\mathscr{L}$ denote the set of selected lines in the line plan and let $q_{o,d}^{(\text{dir})}(t)$ denote the number of travel options (direct travel options) on $\rho_{o,d}$ in the interval $(t, t+T)$. Then, the realized objective at time t is simply obtained by replacing $\sum_{l\in\mathcal{L}} x_l$ with $|\mathscr{L}|$ and by substituting the theoretical reduction in (direct) OD-frequency $z_{o,d}^{(\text{dir})}$ in the objective (1) with $g_{o,d} - q_{o,d}^{(\text{dir})}(t)$, the experienced reduction in the next T minutes:

$$\text{Real. Obj.}(t) = w_1 \sum_{(o,d)\in\mathcal{OD}} n_{o,d} \left(\frac{g_{o,d} - q_{o,d}(t)}{g_{o,d}}\right)^2 + w_2 \sum_{(o,d)\in\mathcal{OD}_{\text{dir}}} n_{o,d} \left(\frac{g_{o,d}^{\text{dir}} - q_{o,d}^{\text{dir}}(t)}{g_{o,d}^{\text{dir}}}\right)^2 + w_3|\mathscr{L}|. \tag{36}$$

By comparing this measure to the theoretical objective of the line plan, we can assess to what extent dispatching strategies result in the same number of travel options per period as intended in the line plan.

6 Computational Results

To test our approach, we conducted experiments on multiple parts of the network operated by NS. Disrupted regions of different sizes are considered, such that we can examine the practicability of the algorithm in different scenarios.

6.1 Problem Instances

We test the developed algorithms on two disrupted regions, using the 2017 line plan and timetable. The original line plan of the part of the Dutch railway network that we consider is presented in Figure 7. NS operates a dense line system in this part of the Netherlands, with six intercity lines and nine regional lines. Figure 7 also indicates which stations serve as decoupling stations, as specified by NS and ProRail.

The largest part of the considered railway network is double-tracked. To accommodate for higher frequencies of intercity and regional trains, there are four tracks between Utrecht Centraal and Amsterdam-Zuid and between Utrecht Centraal and Utrecht Overvecht. Furthermore, intercity trains can overtake regional trains at Amsterdam Muiderpoort, Weesp and Naarden-Bussum. The part between Baarn and Den Dolder is single-tracked, with a passing possibility at Soest.

For the passenger data we use an origin destination matrix with the average daily number of passengers between stations provided by NS. As the network that we consider is only a part of the Dutch railway network, there are many passengers that travel through the considered network without having both the origin and destination in this network. This is taken into account by including the passenger counts from and to major intercity stations outside the considered network. If possible, we derive the number of available platforms at boundary stations for serving the disrupted region from available contingency plans used by NS and ProRail. If no applicable contingency plan is available, we estimate the number of available platforms.

The first disrupted region we consider is the region bounded by Utrecht Centraal, Den Dolder, Baarn and Hilversum. An out-of-control situation in this region could occur due to a power outage at Amersfoort, directly impacting five of the eight lines in this region. We assume buses are transporting passengers from Baarn and Den Dolder to Amersfoort, and vice versa. An interesting

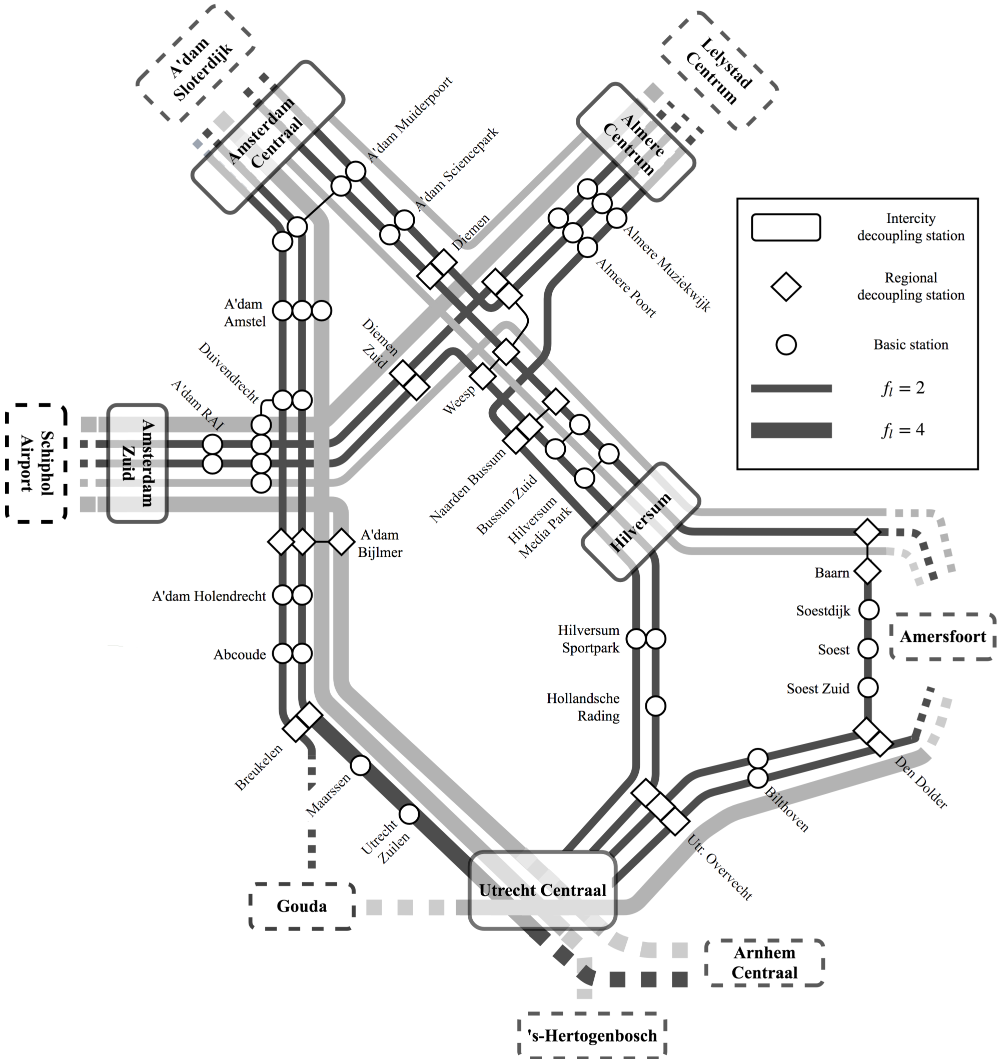


Figure 7: The regular line plan in a part of the Dutch railway network. Line frequencies are indicated by the thickness of the lines. Dark lines are regional lines, bright lines are intercity lines.

aspect of this disrupted region is that while the most important lines (in terms of the number of passengers) in this region are the intercity lines between Hilversum and Amersfoort and between Utrecht Centraal and Amersfoort, it is not possible to operate any intercity trains in the disrupted region. The reason for this is that Baarn and Den Dolder are regional decoupling stations, such that intercity trains cannot turn at these stations. Furthermore, it is not allowed to operate an intercity line between Utrecht Centraal and Hilversum because there is no such line in the regular line plan. Hence, this poses the challenge to optimally use the available regional trains in order to also serve the passengers that normally take intercity trains.

The large instance that is considered is the region bounded by Amsterdam Centraal, Amsterdam

Zuid, Almere Centrum, Utrecht Centraal, Den Dolder, Baarn and Hilversum. An out-of-control situation in such a large part of the Dutch railway network could occur after a combination of major disruptions at Amsterdam and Utrecht. In this instance it is possible to plan both regional and intercity lines. Besides its size, an interesting aspect of this instance is that there is limited capacity for turning trains at Amsterdam Zuid, Utrecht Centraal and Almere Centrum. Given that a large number of trains attend these stations in the regular line plan, the limited station capacity is most likely one of the bottlenecks in the large instance.

6.2 Parameter Settings and Experimental Setup

The running times between stations are derived from the regular timetable of NS. For the dwell times, we use 5 minutes if the train turns and 2 minutes otherwise. We use a headway time of 2 minutes. Hence, the blocking time of a line at a station is 7 minutes if the line has a turn at the station and 4 minutes otherwise. The parameter T, the period of the timetable is set to 60 minutes, the cycle time used by NS.

We impose the timetabling constraints (10-13) in their most stringent form. That is, we set the maximum additional dwell time $\delta_{ls}^{\max}$ of line l at station s to 0 minutes unless s is a terminal of line l, in which case we set $\delta_{ls}^{\max}$ to 60 minutes. In other words, we impose that the downtime of every line must be entirely spent at the terminal stations. This way we avoid finding line plans that exhaust all capacity at terminal stations, causing trains to queue up in front of these stations when operating the line plan without a timetable.

After initial experiments we decided to set the weights in the objective function as follows. The first term, representing frequency decrease of travel options is given weight $w_1 = \frac{1}{\sum_{(o,d)\in\mathcal{OD}} n_{o,d}}$, i.e. we normalize by dividing by the number of travelers. The second term, representing the decrease of direct travel options is given weight $w_2 = \frac{1}{2\sum_{(o,d)\in\mathcal{OD}_{\text{dir}}} n_{o,d}}$, i.e. we divide by two times the number of travelers with a direct travel option. The last term, the line cost, is given weight $w_3 = 0.001$. These settings ensure that the most emphasis is put on providing all passengers with sufficient travel options ("getting everyone home"), and that direct connections are maintained unless this strongly harms the first objective. Ties are broken based on the number of lines.

For a thorough analysis of the performance of the line planning algorithm, we perform tests with three variants of the rolling stock constraints. In the first setting, no rolling stock constraints are included. This setting serves as a reference and is used to illustrate which line plan would be possible if ample rolling stock would be present. In the two other settings, *fixed* and *flexible* rolling stock constraints are included, respectively.

The algorithm takes as input the trains that are present in the disrupted region when the region is decoupled. Given this time instant, the positions of trains can be derived from the timetable. Since this of course varies over the hour, we run the algorithm for five equidistant time instants (minutes 1, 13, 25, 37 and 49). As NS operates most lines every 15 or 30 minutes, this approach ensures that the majority of the variation in the number of available trains and their positions is captured. When determining the performance of the different dispatching strategies on the line plans, we use a simulated time of 4 hours and average the measures over the five time instants.

The line planning algorithm and the simulation are implemented in Java 8 on a HP EliteBook 8460p running Windows 10 with an Intel Core i5-2520M processor at 2.5 GHz and 4 GB of RAM. CPLEX 12.8.0 is used to solve both the master and slave problems. In every iteration of the line planning algorithm, we let the master problem terminate either when the optimal solution is found, or when a time limit of 120 seconds is reached.

6.3 Results Small Disrupted Region

In Table 1, the objective, number of lines, number of iterations, number of cuts and computation times are presented for the three settings. The results of the *fixed* setting and *flexible* setting depend on the time instant, hence for these settings we present the averages over the 5 time instants. It can be seen that without any rolling stock constraints we find a line plan with an objective very close to 0, indicating that hardly any travelers have a reduction in their indirect or direct OD-frequency. However, as rolling stock constraints are neglected, it turns out that this line plan requires 14 trains whereas only 6 or 7 trains (depending on the exact time the disrupted region is decoupled) are available. The line plans obtained with the *fixed* and *flexible* rolling stock constraints can be operated with the available rolling stock, but this of course results in a worse objective. The *flexible* line plans achieve a much better objective than the *fixed* line plans.

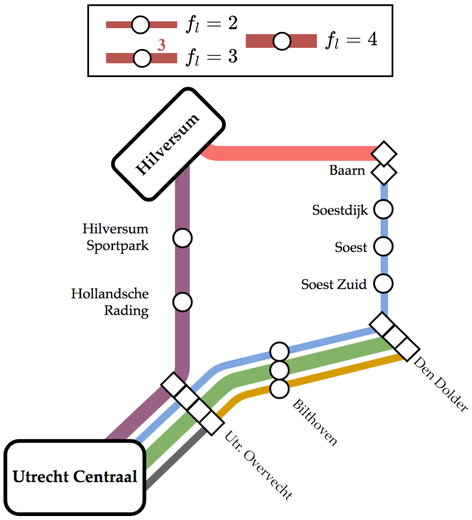
With regards to the computational performance, it can be seen that all solutions are found within a few seconds. The solution with the *basic* setting is found after 2 iterations of the line planning algorithm, in which 6 cuts are added to the master problem. For the *fixed* setting and *flexible* setting, the number of iterations varies between 1 and 2, and much fewer violated inequalities are encountered. This demonstrates that in this disrupted region, the amount of rolling stock available is the main limiting factor for designing an attractive line plan.

Table 1: Results of the line planning algorithm on the small disrupted region using the three types of rolling stock (RS) constraints.

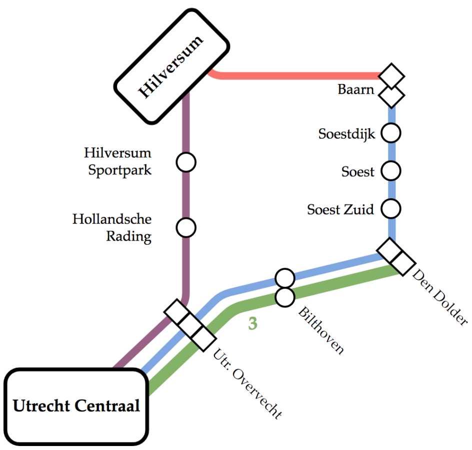
RS constraints	Objective	Lines	Iterations	Cuts	Master CPU (s)	Total CPU (s)	Max Total CPU (s)
None	0.033	6	2	6	0.3	2.2	2.2
Fixed	0.447	3.8	1.2	0.4	0.2	0.4	0.9
Flexible	0.351	4	1.4	0.8	0.7	1.0	2.2

To illustrate the impact of the different settings on the solution, the line plans for one of the time instants are visualized in Figure 8. As can be seen, the line plan obtained without rolling stock constraints is much denser than the *fixed* and *flexible* line plans, once again showing that not taking into account rolling stock restrictions results in line plans that cannot be operated with the available rolling stock. The *fixed* and *flexible* line plans have the same lines, but with different frequencies. By combining lines into circulations, the frequency between Utrecht Centraal and Hilversum and between Baarn and Hilversum can be increased from 2 to 3. The frequency of the line between Utrecht Centraal and Den Dolder however is decreased from 3 to 2. Overall, this results in a theoretically better solution.

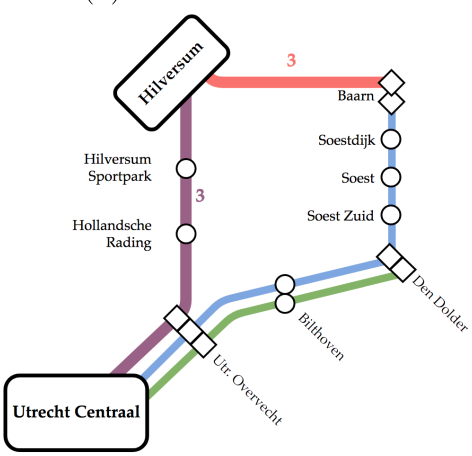
The simulation results are presented in Table 2 and Figure 9. Table 2 contains the travel measures, averaged over the entire simulation and Figure 9 shows the values for the operational measures plotted over time. The *fixed* setting, which theoretically attains an objective of 0.447, results in realized objectives of about 0.39 and 0.48 for strategies with the ASAP and SYNC timing principle, respectively. This can be explained using Figure 9(a), where we observe that the ASAP strategies have normalized frequencies that are much larger than 1, whereas the SYNC strategies have normalized frequencies very close to 1. This means that the ASAP strategies lead to 'more trains than promised' and consequently, a better objective than promised. Next to that, the ASAP strategies also result in shorter travel times. The turning principle has a much smaller effect on the performance of the *fixed* line plans. The DYN strategies have slightly shorter travel times.



(a) No RS constraints

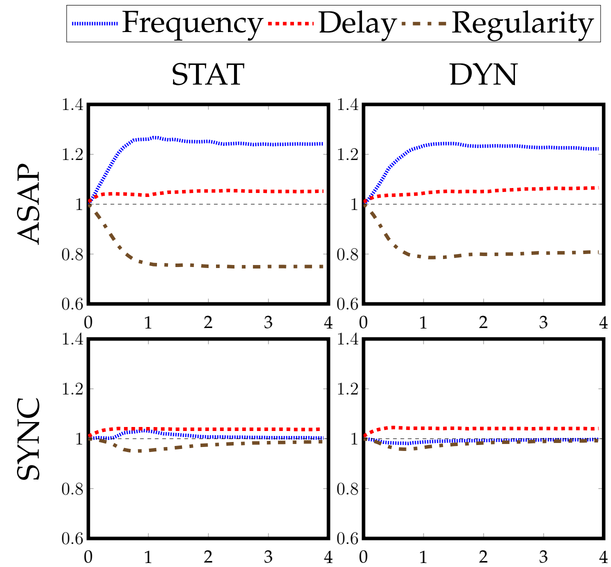


(b) Fixed RS constraints

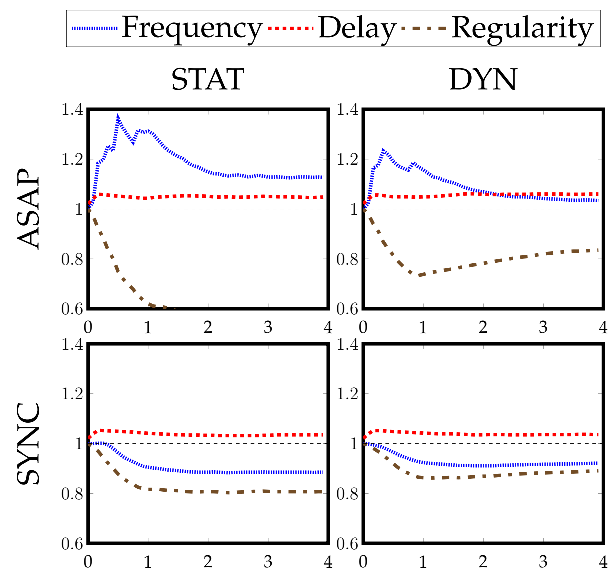


(c) Flexible RS constraints

Figure 8: Line plans for the small disrupted region obtained with different rolling stock (RS) constraints included. All lines are regional lines.



(a) Fixed RS constraints



(b) Flexible RS constraints

Figure 9: Operational measures in the small disrupted region obtained with (a) fixed and (b) flexible rolling stock (RS) constraints. The four strategies are structured in a 2x2 matrix. In the figures, the horizontal axis denotes the time in hours and the vertical axis the score on the three measures. The closer a measure is to the horizontal dashed line, the better the performance.

Table 2: Values of the travel measures for the different strategies and rolling stock (RS) constraints in the small disrupted region, averaged over the five time instants. In the undisrupted situation, the average travel time (without waiting time) is about 18 minutes.

	Fixed RS constraints		Flexible RS constraints	
	Real. Obj.	Travel Time (min.)	Real. Obj.	Travel Time (min.)
ASAP-STAT	0.388	28.1	0.363	29.0
SYNC-STAT	0.478	31.1	0.436	29.9
ASAP-DYN	0.388	27.9	0.361	27.9
SYNC-DYN	0.486	30.9	0.411	29.3

When the *flexible* setting is used, the turning principle now plays a more important role in the performance. The DYN principle clearly outperforms the STAT principle, resulting in both a lower objective and shorter travel times. This was to be expected, since the *flexible* setting generates line plans that cannot be operated if only fixed turnings are allowed. As with the *fixed* setting, the strategies with the ASAP principle attain a better realized objective, although the theoretical objective of 0.351 is not reached.

Comparing the results for the line planning settings, we can observe that for any dispatching strategy, the *flexible* setting has a better realized objective, and that for both line planning settings, the objective and travel times are lowest using the ASAP-DYN strategies. On the other hand, Figure 9 shows that the SYNC strategies generate a more stable and predictable service, which can also be deemed important when dealing with an out-of-control situation.

6.4 Large Disrupted Region

In Table 3, the objective, number of lines, number of required trains under fixed circulations, number of iterations, number of cuts and computation times are presented for the three settings. The results of the *fixed* and *flexible* constraints are again averages over the 5 time instants. As in the small disrupted region, neglecting rolling stock constraints leads to a line plan that achieves a very low objective, but is actually infeasible. However, the difference is smaller than in the small region, showing that next to the number of available trains, the available infrastructure also poses a large restriction on the line plans that can be realized in the large disrupted region.

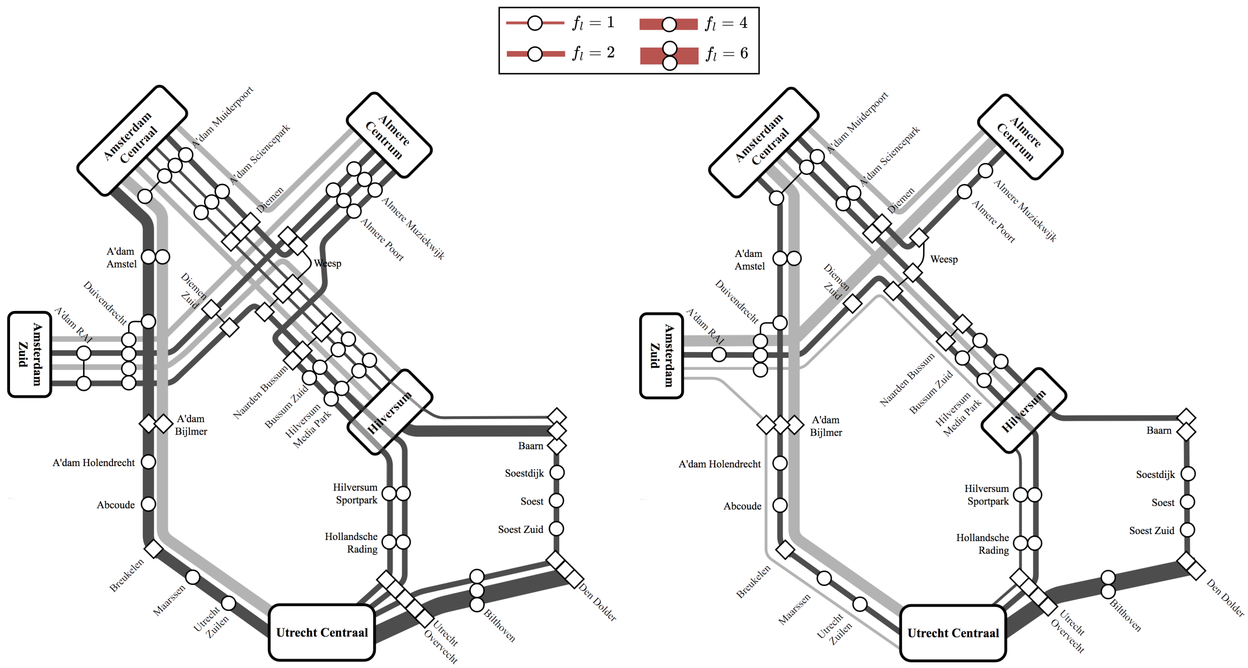
As for the computational performance, the computation times strongly depend on the setting. Using the *fixed* setting, the optimal solution is found on average within 16 seconds. The *flexible* setting takes considerably more time, on average about 90 seconds. The *basic* setting even takes more than 8 minutes, but this setting is unlikely to be used in real-time applications as it produces overly optimistic line plans. Both the *fixed* and *flexible* setting require on average 2 iterations during which 160 cuts are added. In fact, it turns out that the same 160 cuts are added for both settings with all 5 time instants.

The line plans for one of the time instants are visualized in Figure 10. As can be observed, the line plan without rolling stock constraints is similar to the regular line plan that is operated. The frequency of regional trains is increased between Hilversum and Baarn and between Utrecht Centraal and Den Dolder to compensate for the intercity trains that cannot run between these stations. As for the intercity lines, the limited turning capacity at Amsterdam Zuid, Utrecht and Almere Centrum leads to the cancellation of the line between Amsterdam Zuid and Utrecht Centraal and the reduction of the frequency of the line between Amsterdam Zuid and Almere Centrum from 4 to 2 per hour. In the *fixed* line plan, many regional lines are canceled or have their frequency reduced, e.g. between Amsterdam Zuid and Almere Centrum and between Amsterdam Centraal and Utrecht Centraal. The frequency of the intercity between Amsterdam Zuid and Hilversum is also reduced. On the other hand, the intercity between Amsterdam Zuid and Almere Centrum has frequency 4 compared to 2 in the basic line plan, and the intercity between Amsterdam Zuid and Utrecht Centraal, which was canceled entirely in the *basic* line plan, is included in the *fixed* line plan, albeit with frequency 1. In the *flexible* line plan, many improvements are visible over the *fixed* line plan. More and longer regional lines are operated, restoring direct connections between Utrecht Centraal and Baarn and Amsterdam Zuid and Almere. Furthermore, the intercity between Amsterdam Zuid and Utrecht Centraal is included with frequency 2, compared to 1 in the *fixed* line plan.

Table 3: Results of the line planning algorithm on the large disrupted region using the three types of rolling stock (RS) constraints.

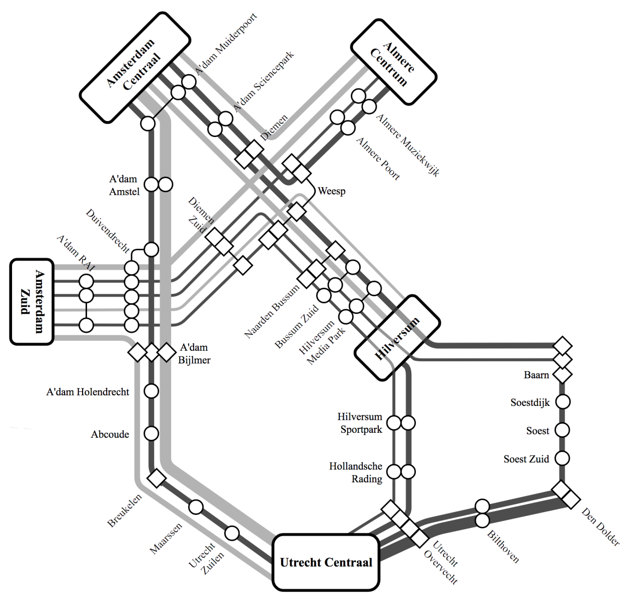
RS constraints	Objective	Lines	Iterations	Cuts	Master CPU (s)	Total CPU (s)	Max Total CPU (s)
None	0.074	36	4	371	480.1	501.5	501.5
Fixed	0.162	22.0	2.0	160.0	11.0	15.9	18.8
Flexible	0.133	26.0	2.0	160.0	82.5	88.4	99.5

The simulation results of the large disrupted region are presented in Table 4 and Figure 11. Table 4 contains the travel measures, averaged over the entire simulation and Figure 11 shows the values for the operational measures plotted over time.As in the small disrupted region, the



(a) No RS constraints

(b) Fixed RS constraints



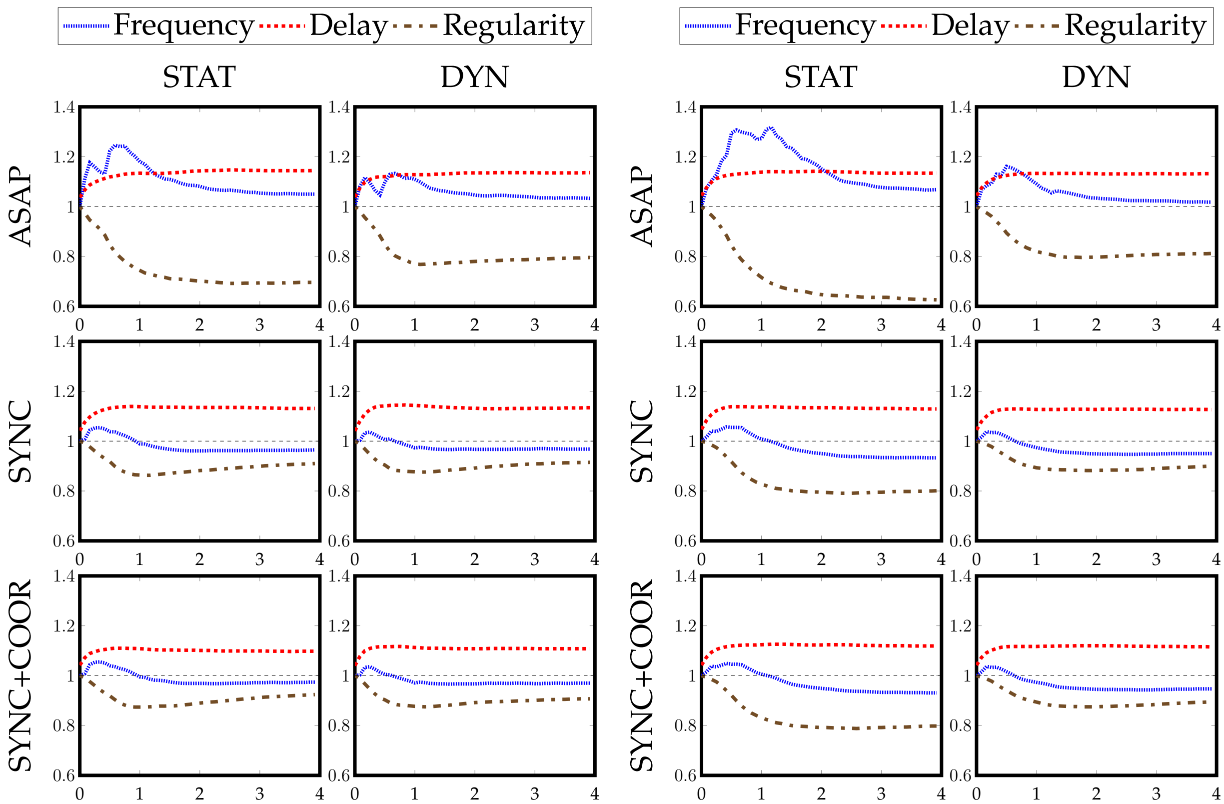
(c) Flexible RS constraints

Figure 10: Line plans for the large disrupted region obtained with the different rolling stock (RS) constraints. Dark lines are regional lines, bright lines are intercity lines.

flexible rolling stock constraints in combination with the ASAP-DYN strategy results in the best realized objective. However, note that in contrast to the small instance, the realized objectives are worse than the theoretical objectives that are attained. The figures with the operational measures provide an explanation. It is visible that the delay measure is over 1.1 for all line planning settings

Table 4: Values of the travel measures for the different strategies and rolling stock (RS) constraints in the large disrupted region, averaged over the five time instants. In the undisrupted situation, the average travel time (without waiting time) is about 25 minutes.

	Fixed RS constraints		Flexible RS constraints	
	Real. Obj.	Travel Time (min.)	Real. Obj.	Travel Time (min.)
ASAP-STAT	0.222	38.8	0.218	38.4
SYNC-STAT	0.222	37.6	0.254	39.1
SYNC+COOR-STAT	0.221	36.8	0.251	38.3
ASAP-DYN	0.196	37.1	0.195	37.3
SYNC-DYN	0.217	37.2	0.215	37.2
SYNC+COOR-DYN	0.216	36.9	0.217	36.7



(a) Fixed RS constraints (b) Flexible RS constraints

Figure 11: Operational measures in the large disrupted region obtained with (a) fixed and (b) flexible rolling stock (RS) constraints. The six strategies are structured in a 3x2 matrix. In the figures, the horizontal axis denotes the time in hours and the vertical axis the score on the three measures. The closer a measure is to the horizontal dashed line, the better the performance.

and dispatching strategies, which is simply caused by the larger size of the network, with more potential places for conflicts between trains. This implies that the actual travel times are about 10 percent larger than travel times used in optimizing the line plans. As a consequence, the realized frequencies are below the 'promised' frequencies, causing the realized objective to be worse.

Next to that, it is visible that the large difference in the attained theoretical objective between the *fixed* and *flexible* line plans does not translate in the same difference in realized objective. On the other hand, the dispatching strategies that apply the DYN principle, which allows flexible turnings during operations, do achieve significantly better objectives, travel times and also score better on the operational measures. This shows that flexible turnings partially compensate for the longer than anticipated travel times.

A last observation regarding Table 4 and Figure 11 concerns the SYNC+COOR principle. While the difference with the SYNC principle is very subtle for the operational measures, the SYNC+COOR principle does lead to a significantly lower average travel time. In fact, if combined with the *flexible* setting, the SYNC+COOR-DYN strategy results in the lowest travel time over all strategies, illustrating that adding small coordination mechanisms can greatly improve the experience of passengers. However, this does increase the complexity of the operations, potentially harming the controllability of the system, which conflicts with the original goal of preventing out-of-control situations.

7 Conclusion

In this paper, we addressed new disruption management strategies for (near) out-of-control situations occurring in railway systems. We developed a novel algorithm for redesigning the line plan, such that the railway system within a disrupted region can be operated using self-organizing principles. In order to ensure that the resulting line plan is feasible with respect to the available railway infrastructure and rolling stock, the proposed algorithm partially integrates line planning with timetabling and rolling stock scheduling using the integer version of Benders' decomposition. Besides investigating which lines should be operated when the system gets out-of-control, we also analyzed how the adapted line system should be operated. To this end, we developed several dispatching strategies that only require local coordination.

Computational experiments on the Dutch railway network indicate that the algorithm performs well. Optimal alternative line plans are provided in short amounts of time, making the approach applicable for use in practice. Using simulation, we also demonstrated that by applying the appropriate dispatching strategies, the produced line plans can be operated smoothly without relying on central coordination. The results show that allowing flexible turnings leads both to better line plans and to a better performance during operations, resulting in more travel options for passengers and shorter travel times. Next to that, we observed that adding simple coordination mechanisms between slower and faster trains also significantly improves the experience of passengers, at the cost of increased complexity of the operations. All in all, our work certainly highlights the opportunity to offer limited service in out-of-control situations, where the current practice leads to termination of all traffic within the affected region.

As a next step, it is interesting to also consider train drivers and conductors when operating the disrupted region using self-organizing principles. In addition, we see an application of the proposed decomposition approach to integrated models for strategic railway planning.

Acknowledgements This research was funded by NWO, the Netherlands Organisation for Scientific Research, as part of the research programme Complexity in Transport & Logistics (project number 439.16.111). We also would like to thank Stefanie Schouten and Alfons Schaafsma for their contributions to the early stages of this research.

References

Berrebi SJ, Watkins KE, Laval JA, 2015 *A real-time bus dispatching policy to minimize passenger wait on a high frequency route. Transportation Research Part B: Methodological* 81:377–389.

Burggraeve S, Bull SH, Vansteenwegen P, Lusby RM, 2017 *Integrating robust timetabling in line plan optimization for railway systems. Transportation Research Part C: Emerging Technologies* 77:134–160.

Cacchiani V, Huisman D, Kidd MP, Kroon LG, Toth P, Veelenturf LP, Wagenaar J, 2014 *An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological* 63:15–37.

Claessens MT, van Dijk NM, Zwaneveld PJ, 1998 *Cost optimal allocation of rail passenger lines. European Journal of Operational Research* 110(3):474–489.

Codato G, Fischetti M, 2006 *Combinatorial Benders' cuts for mixed-integer linear programming. Operations Research* 54(4):756–766.

Dekker MM, Van Lieshout RN, Ball RC, Bouman PC, Dekker SC, Dijkstra HA, Goverde RMP, Huisman D, Schaafsma AAM, Van den Akker JM, 2018 *A next step in disruption management: combining Operations Research and Complexity Science. Econometric Institute Research Papers* No. EI2017-23.

Dessouky M, Hall R, Nowroozi A, Mourikas K, 1999 *Bus dispatching at timed transfer transit stations using bus tracking technology. Transportation Research Part C: Emerging Technologies* 7(4):187–208.

Ghaemi N, Cats O, Goverde RMP, 2017 *Railway disruption management challenges and possible solution directions. Public Transport* 9(1-2):343–364.

Goossens JW, Van Hoesel S, Kroon LG, 2004 *A branch-and-cut approach for solving railway line-planning problems. Transportation Science* 38(3):379–393.

Jespersen-Groth J, Potthoff D, Clausen J, Huisman D, Kroon LG, Maróti G, Nielsen MN, 2009 *Disruption management in passenger railway transportation. Robust and online large-scale optimization*, 399–421 (Springer).

Kaspi M, Raviv T, 2013 *Service-oriented line planning and timetabling for passenger trains. Transportation Science* 47(3):295–311.

Kroon LG, Peeters LW, Wagenaar JC, Zuidwijk RA, 2013 *Flexible connections in PESP models for cyclic passenger railway timetabling. Transportation Science* 48(1):136–154.

Liebchen C, Möhring RH, 2007 *The modeling power of the periodic event scheduling problem: railway timetables and beyond. Algorithmic methods for railway optimization*, 3–40 (Springer).

Pätzold J, Schiewe A, Schiewe P, Schöbel A, 2017 *Look-Ahead Approaches for Integrated Planning in Public Transportation. 17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, 17:1–17:16.

Peeters L, 2003 *Cyclic Railway Timetable Optimization*. Ph.D. thesis.

Schöbel A, 2012 *Line planning in public transportation: models and methods. OR spectrum* 34(3):491–510.

Schöbel A, 2017 *An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. Transportation Research Part C: Emerging Technologies* 74:348–365.

Serafini P, Ukovich W, 1989 *A mathematical model for periodic scheduling problems. SIAM Journal on Discrete Mathematics* 2(4):550–581.

Van Lieshout RN, Bouman PC, 2018 *Vehicle Scheduling Based on a Line Plan. 18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, 15:1–15:14.

Vanderbeck F, Wolsey LA, 2010 *Reformulation and decomposition of integer programs. 50 Years of Integer Programming 1958-2008* 431–502.

Veelenturf LP, Kidd MP, Cacchiani V, Kroon LG, Toth P, 2015 *A railway timetable rescheduling approach for handling large-scale disruptions. Transportation Science* 50(3):841–862.

Veelenturf LP, Potthoff D, Huisman D, Kroon LG, Maróti G, Wagelmans APM, 2014 *A quasi-robust optimization approach for crew rescheduling. Transportation Science* 50(1):204–215.

A Linearizing a Quadratic Program

Here we describe how to linearize a quadratic program of the type

$$\text{minimize} \quad \sum_i \left(\frac{z_i}{g_i}\right)^2 \tag{37}$$

$$\text{s.t.} \quad \sum_i a_{ij} z_i \leq b_j, \qquad \forall j, \tag{38}$$

$$0 \leq z_i \leq g_i, \quad z_i \text{ integer}, \quad \forall i. \tag{39}$$

To this end, we introduce the auxiliary binary variables $u_{i1}, u_{i2}, ..., u_{ig_i}$, add the constraints

$$z_i = \sum_{k=1}^{g_i} u_{ik}, \quad \forall i, \tag{40}$$

$$u_{ik} \in \{0, 1\}, \quad \forall i, \forall k, \tag{41}$$

and replace the objective by

$$\text{minimize} \quad \sum_i \sum_{k=1}^{g_i} c_{ik} u_{ik} \tag{42}$$

where the cost coefficients c_{ik} are defined as follows:

$$c_{ik} = \left(\frac{k}{g_i}\right)^2 - \left(\frac{k-1}{g_i}\right)^2. \tag{43}$$

Clearly the original formulation has a feasible solution if and only if the linearized formulation has a feasible solution. Furthermore, note that the cost coefficients are increasing in k. Hence, if $z_i = m$, it is optimal to set $u_{e1} = u_{e2} = ... = u_{em} = 1$ and $u_{e(m+1)} = u_{e(m+2)} = ... = u_{eg_e} = 0$. This results in an objective contribution of $\sum_{k=1}^m c_{ik} = \sum_{k=1}^m \left(\frac{k}{g_i}\right)^2 - \left(\frac{k-1}{g_i}\right)^2 = \left(\frac{m}{g_i}\right)^2 - \left(\frac{m-1}{g_i}\right)^2 + \left(\frac{m-1}{g_i}\right)^2 - ... - \left(\frac{1}{g_i}\right)^2 + \left(\frac{1}{g_i}\right)^2 = \left(\frac{m}{g_i}\right)^2$. In terms of the original objective, the contribution equals $\left(\frac{z_i}{g_i}\right)^2 = \left(\frac{m}{g_i}\right)^2$. As the contribution to the objective value is the same for both formulations, they are equivalent.

B Formulation of the C-PESP

The PESP can be formulated concisely using an event-activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$. The arrival of the i'th train of line l at station s is represented by an arrival node $(l, s, \text{arr}, i) \in \mathcal{E}^{\text{arr}}$. Similarly, departures are represented by nodes $(l, s, \text{dep}, i) \in \mathcal{E}^{\text{dep}}$. In the basic form, there are four types of activities (arcs) linking two nodes:

- Dwelling activities link arrival nodes (l, s, arr, i) to departure nodes (l, s, dep, i).
- Driving activities link the departure node (l, s_1, dep, i) at station s_1 to the arrival node (l, s_2, arr, i) at the next station s_2.
- Safety activities link departure nodes (l, s, dep, i) or arrival nodes (l, s, arr, i) with departure nodes (l', s, dep, j) or arrival nodes (l', s, arr, j).
- Synchronizing activities link departure nodes (l, s, dep, i) with departure nodes (l, s, dep, j), with $i \neq j$.

Every activity corresponds to a constraint stating that the duration of the activity should be in a certain interval. For example, the safety activities correspond to constraints stating that the time between two trains using the same piece of infrastructure should be at least the headway time.

We let $\pi_i \in [0, T-1]$ denote the decision variable representing the time instant assigned to node i. The periodic constraint for an activity $(i,j) \in \mathcal{A}$ is then given by

$$l_{ij} \leq \pi_j - \pi_i + T p_{ij} \leq u_{ij}, \quad p_{ij} \in \{0,1\}, \tag{44}$$

where l_{ij} is the lower bound of the duration of activity (i,j) and u_{ij} the upper bound. The decision variables p_{ij} are introduced to compute the duration of activities correctly when $\pi_j < \pi_i$ and referred to as the *modulo* parameters.

Station Capacity in the PESP

A station capacity constraint can be formulated using the following interpretation of the modulo parameters:

$$p_{ij} = \begin{cases} 1, & \text{if event } j \text{ takes place before event } i, \\ 0, & \text{if event } j \text{ takes place after or at the same time as event } i. \end{cases}$$

Here, *before* or *after* refers to the sequence of the events on the linear axis $[0, T-1]$. Now assume the event-activity network of a station $\mathcal{N}_s = (\mathcal{E}_s, \mathcal{A}_s)$ contains all arcs

$$(i,j), \text{ where } i \in \mathcal{E}_s^{\text{arr}} \text{ and } j \in \mathcal{E}_s^{\text{dep}},$$
$$(i,j), \text{ where } i \in \mathcal{E}_s^{\text{arr}} \text{ and } j \in \mathcal{E}_s^{\text{arr}} \text{ and } i < j.$$

Let $\mathcal{A}_s^{\text{dwell}}$ denote the set of dwell activities at station s, let $\mathcal{A}_i^a$ denote the set of outgoing activities to arrivals from event i and let $\mathcal{A}_i^d$ denote the set of outgoing activities to departures from event i. Then, we can enforce that the capacity of station s is never violated by adding the following constraints:

$$1 + \sum_{(k,l)\in\mathcal{A}_s^{\text{dwell}}} p_{kl} + \sum_{(j,i)\in\mathcal{A}_j^a} (1 - p_{ji}) + \sum_{(i,j)\in\mathcal{A}_i^a} p_{ij} - \sum_{(i,j)\in\mathcal{A}_i^d} p_{ij} \leq |P^s| \quad \forall i \in \mathcal{A}_s^{\text{arr}}. \tag{45}$$

Usually, the platforms at a station are subdivided into groups that are assigned to the different lines and directions. In such cases, the capacity constraint can be included for every group of platforms.

A slight inaccuracy in the constraints is that when events occur concurrently, this is not dealt with consistently. This issue can be resolved for by introducing a second modulo parameter for every pair of events. We refer to Peeters (2003) for details.