Reducing Passenger Delays by Rolling Stock Rescheduling

Rowan Hoogervorst[1], Twan Dollevoet[1], Gábor Maróti[2,3], and Dennis Huisman[1,3]

[1] *Econometric Institute and ECOPT, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands*
[2] *VU University Amsterdam, 1081 HV Amsterdam, The Netherlands*
[3] *Process quality and Innovation, Netherlands Railways, 3500 HA Utrecht, The Netherlands*

Econometric Institute Report Series - EI2018-29

Abstract

Delays are a major nuisance to railway passengers. The extent to which a delay propagates, and thus affects the passengers, is influenced by the assignment of rolling stock. We propose to reschedule the rolling stock in such a way that the passenger delay is minimized and such that objectives on passenger comfort and operational efficiency are taken into account. We refer to this problem as the Passenger Delay Reduction Problem (PDRP). We propose two models for this problem, which are based on two dominant streams of literature for the traditional Rolling Stock Rescheduling Problem. The first model is an arc formulation of the problem, while the second model is a path formulation. We test the effectiveness of these models on instances of Netherlands Railways (NS). The results show that the rescheduling of rolling stock can significantly decrease the passenger delays in the system. Especially allowing flexibility in the assignment of rolling stock at terminal stations turns out to be effective in reducing the delays. Moreover, we show that the arc formulation based model performs best in finding high-quality solutions within the limited time that is available in the rescheduling phase.

Keywords— Rolling Stock Rescheduling, Disruption Management, Railway Optimization, Column Generation

1 Introduction

Delays are among the largest annoyances experienced by railway passengers. A recent study (Kennisinstituut voor Mobiliteitsbeleid 2017) estimates the annual societal costs of train delays and cancellations in the Netherlands to be more than 400 million euro. These costs are not only related to passengers arriving late at their destination, but are, e.g., also a result of the increased uncertainty that passengers feel towards their journey. In this paper, we focus on reducing the effects of relatively large delays, i.e., delays between 15 and 30 minutes in our application for Netherlands Railways (NS). In particular, we focus on reducing the effects of delays on passengers by means of rolling stock rescheduling.

Traditionally, the planning process of a railway operator consists of three sequential steps: timetabling, rolling stock scheduling and crew scheduling. In the timetabling step, one finds a timetable based on the services that need to be operated. In rolling stock scheduling we then assign train units, i.e., units of rolling stock, to the trips present in this timetable. The aim in rolling stock scheduling is to find a rolling stock assignment that offers enough seat capacity, but which is also not too expensive to operate. Moreover, it should respect the limited number of train units that is available. The found rolling stock schedule, i.e., rolling stock assignment, finally serves as input to the crew scheduling step.

By assigning rolling stock to the trips, the rolling stock schedule creates links between trips which are successively operated by the same train units. These links between trips lead to delay propagation when the assigned rolling stock becomes delayed. This dependence of delay propagation on the rolling stock schedule creates opportunities to reduce the impact of delay on passengers. By changing the rolling stock schedule, we change the links between trips and hence affect the propagation of delay through the railway system. In this way, we may decrease the faced delays by making better use of the buffers in the timetable. In addition, we may be able to move the delay to trips with fewer passengers.

We refer to this problem of decreasing the passenger delays by rolling stock rescheduling as the *Passenger Delay Reduction Problem* (PDRP). In this problem, we minimize the delays that are experienced by passengers on the different trips while also taking into account objectives on passenger comfort and operational efficiency. Compared to traditional rolling stock rescheduling, we take into account the propagation of delay through the railway system and actively try to reschedule the rolling stock schedule to decrease the propagation of delay. Moreover, we allow to change the connections between incoming and outgoing trips at some of the terminal stations to minimize this propagation of delay. These connections are generally fixed beforehand in rolling stock rescheduling to lower the computational effort needed to find a rolling stock schedule. Allowing to change such connections is referred to as *flexible turning* (Nielsen 2011).

Our contributions in this paper are threefold. First, we introduce the PDRP to minimize passenger delays by means of rolling stock rescheduling. Second, we introduce two models to solve the PDRP, which are based on two well-known models for solving the traditional Rolling Stock Rescheduling Problem. Third, we show the applicability of the proposed models to instances of NS. Each of these instances corresponds to a random delay in the timetable operated by NS. Our results show that rolling stock rescheduling can significantly decrease the passenger delays, where especially flexible turning plays a central role in reducing the delays.

The remainder of this paper is organized as follows. In Section 2, we introduce the PDRP. In Section 3, we discuss the related literature. In Section 4, we discuss two models that are commonly used in rolling stock rescheduling, which are both extended to the PDRP in Section 5. In Section 6, we introduce our solution methods for the two proposed models. Finally, we test the proposed methods on instances of NS in Section 7 and show the extent to which rescheduling for delays changes the original rolling stock schedule. We conclude the paper in Section 8.

2 The Passenger Delay Reduction Problem

In this section, we introduce the Passenger Delay Reduction Problem (PDRP) that aims to reduce passenger delays by means of rolling stock rescheduling. We describe the PDRP in the existing rolling stock (re-)scheduling setting of Fioole et al. (2006).

2.1 Rolling Stock (Re-)Scheduling

Rolling stock scheduling and rescheduling deal with assigning rolling stock to the trips in the timetable. In this paper, we restrict ourselves to rolling stock that is composed of self-powered train units, opposed to locomotive-hauled carriages, as is common for European railway operators. A railway operator generally owns train units of different types, which differ in terms of characteristics such as the number of carriages they contain and whether they are single- or double-deck trains. These characteristics affect the number of passengers these train units can carry.

Train units of compatible types may be combined to form *compositions*. A composition is an ordered sequence of the train unit types that are in a train. An example of a composition, consisting of an ICM-III and an ICM-IV train unit, is given in Figure 1. The order in a composition matters, meaning that reordering the train unit types within a composition results in a different composition. Moreover, we assume in this paper that train units of the same type are interchangeable. Therefore, we only have to be concerned with the train unit types in a composition.

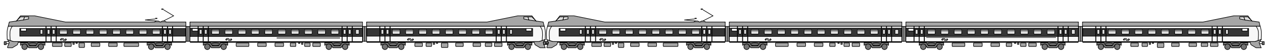


Figure 1: Example of a composition with two train units: ICM-IV (4 carriages) in front, ICM-III (3 carriages) in the back

The composition of a train may be changed at *transitions* between trips. A transition links an incoming trip at a station to an outgoing trip and hence links the rolling stock on the predecessor trip to that of the successor trip. Changes to the composition of a train may occur through shunting movements by the *coupling* of additional train units to a composition and by the *uncoupling* of train units from a composition to the shunting yard. Each such possible way of changing the composition at a transition, including that of not changing the composition at all, is captured by a *composition change*.

The total number of train units that is available is limited. The number of train units that are available at a certain station at some moment in time is referred to as the *inventory* of that station. Of particular interest are the starting inventory and ending inventory, which represent the number of train units present at the start and end of the planning horizon. In this paper, we assume that the starting inventory of train units is fixed, while a target ending inventory is known for each station.

The aim of the *Rolling Stock Scheduling Problem* is to find an assignment of compositions to the trips in the timetable, such that the composition changes implied by these compositions are feasible and such that the starting inventory of train units is respected. All this is done under a large set of mutually conflicting objectives which take into account factors on passenger comfort and operational efficiency. Based on this assignment of compositions to trips we can find an assignment of the individual train units to the trips. This allocation is referred to as the rolling stock *circulation*.

An example of a rolling stock circulation is given in Figure 2 for a timetable that includes three stations and eight trips. This rolling stock circulation includes three train units, all of which are part of the starting inventory of station *Rtd*. Moreover, two train units start with trip t_1, while the other train unit starts on trip t_5. Note how shunting takes place at station *Ut*, where a train unit is uncoupled at the transition between trip t_2 and t_3, after which it is coupled at the transition between t_6 and t_7.

The *Rolling Stock Rescheduling Problem (RSRP)* differs from the Rolling Stock Scheduling Problem as we try to reschedule an existing circulation. Rescheduling is performed in the real-time rescheduling phase when a disruption leads to changes in the original timetable. For example, rescheduling may be performed when a blockage of railway infrastructure between two stations leads to the cancellation of trips. Due to the dynamic nature of disruptions, rolling stock rescheduling is generally performed every time when new information about disruptions comes in. As each rescheduling step requires communication with the crews, an important objective in rolling stock rescheduling is to minimize the deviation from the original circulation.

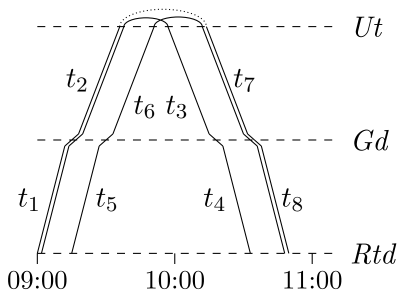


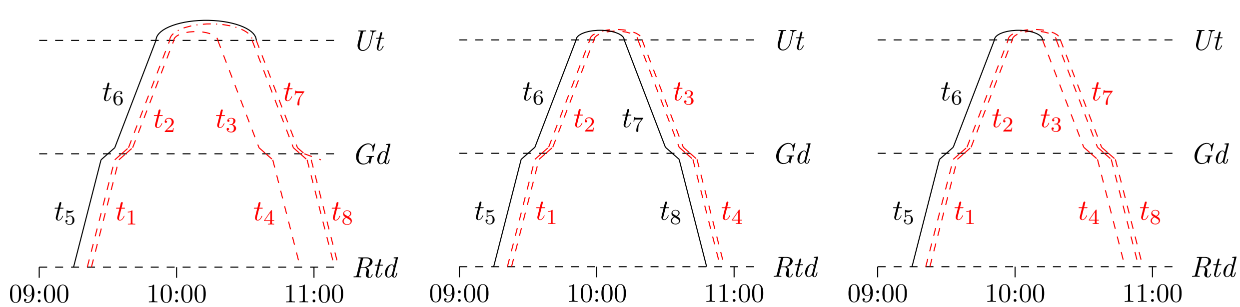
Figure 2: Time-space diagram for a timetable that includes three stations (*Rtd*, *Gd*, *Ut*) and eight trips ($t_1, \ldots, t_8$). Moreover, the rolling stock circulation of three train units is shown. Here, the dotted line indicates the shunting of a train unit.

2.2 The Passenger Delay Reduction Problem

In the PDRP, we again consider the setting of real-time rolling stock rescheduling, but now for a type of disruption that is not considered in the RSRP: a delay that has occurred for a trip, or possibly multiple trips, in the timetable. Such delays occur commonly in practice, e.g., as a result of small technical failures on train units or due to increased dwell times at stations. We focus on delays that cannot quickly be absorbed by slacks in the timetable, which are for NS assumed to be in the order of 15 to 30 minutes.

The rolling stock schedule impacts the propagation of these initial delays in two main ways. First, the transitions that are present between incoming and outgoing trips link the rolling stock that is operated on these trips. These links lead to delay propagation when the rolling stock on the predecessor trip of a transition becomes delayed. Second, the shunting of train units in the rolling stock circulation may lead to further delay propagation. This occurs when a train unit is uncoupled at some transition with a delay and afterwards coupled at some other transition with insufficient time being present between the transitions to absorb the delay.

An illustration of both influences is given in Figure 3(a), which shows how an initial delay on trip t_1 propagates in the timetable presented in Figure 2. First, note how, e.g., the transition between trips t_2 and t_3 causes delay propagation between these trips due to rolling stock that arrives too late at station Ut. Second, note how the delay propagates from trip t_2 to t_7 as an effect of the rolling stock circulation. Here, a train unit is uncoupled from trip t_2 with a delay and afterwards coupled to trip t_7. As the time between the moment of uncoupling and coupling is too short to absorb all the delay, this leads to a delay on trip t_7.



(a) No rescheduling (b) No coupling at t_7 (c) Flexible turning

Figure 3: Delay propagation based on the rolling stock circulation of Figure 2 under different rescheduling actions. The initial delay is present on trip t_1 and a dashed line indicates that a trip is delayed.

The dependence of delay propagation on the rolling stock schedule creates opportunities to reduce the impact that initial delays have on the passengers. First of all, we can change the circulation of train units in order to prevent that a coupled train unit delays a trip. For example, deciding to cancel the coupling of a train unit at trip t_7 prevents that t_7 becomes delayed. This is illustrated in Figure 3(b). The effect of this rescheduling action is that the passengers on trip t_7 and t_8 no longer face a delay. However, this rescheduling action also reduces the seat capacity on these trips.

A second way to alter the delay propagation is to change the defined transitions between incoming and outgoing trips at terminal stations, which are known as *turnings*. This assignment of rolling stock of incoming trips to outgoing trips by means of turnings is referred to as the *turning pattern* of a station. As no passengers are present in a train during a turning, opposed to transitions at intermediate stations, we can reschedule the turning pattern without impacting the passengers. This is referred to as *flexible turning* (Nielsen 2011). The effect of flexible turning on delay propagation is illustrated in Figure 3(c). Note how the turnings are changed at the terminal station *Ut*, where trip t_2 is now connected to trip t_7 and trip t_6 to trip t_3. Moreover, note how flexible turning leads to substantial delay reduction compared to performing no rescheduling and retains the same seat capacity on trips t_7 and t_8 as in the undisturbed scenario.

Just like Nielsen (2011), we assume that flexible turning is only allowed at a set of predetermined stations. Limiting flexible turning to these stations reduces the work that is needed in later stages of the rescheduling process. Moreover, we assume that uncoupling and coupling of train units in a flexible turning may occur respectively after the incoming trip and before the outgoing trip. The composition that comes from the incoming trip, after any shunting actions, then remains parked at the platform of a station until it departs on some outgoing trip.

When determining the delay propagation for a given circulation and given turning patterns at the stations, we consider a limited form of delay propagation in which we only consider delay propagation as a result of delayed rolling stock. This implies in particular that we disregard any delay that is created due to headway constraints between trains. Our motivation for doing so stems from the fact that the delays due to headway constraints are likely to be small and can generally be absorbed by running time supplements.

Moreover, we assume that part of the delay may be absorbed by slacks in the timetable. In this paper, we assume that delay absorption may happen at three different moments in time. First, running time supplements are generally present in the timetable on top of the nominal trip time. Second, supplements may be available for the time that is planned for a transition. Third, some supplement may be available for the time that is needed between uncoupling a train unit at one transition and coupling it at another transition.

The PDRP is now the problem of finding a circulation and corresponding turning patterns for those stations where flexible turning is allowed. This is done under an objective that includes costs for delays and flexible turning next to the traditional objectives in the RSRP on passenger comfort and operational efficiency. To consider the effect that delays have on the passengers, we additionally take into account the number of passengers that are expected to travel on a trip when determining the costs of a delay. In this way, delays on trips with few passengers are preferred over delays on trips with many passengers. Moreover, by penalizing delays in the objective function, opposed to only minimizing the total delay, we follow the idea that disturbing the original circulation to achieve delay reduction can be costly as well. For example, it might lead to trains operating with too little seats for passengers.

3 Related Literature

The rescheduling of rolling stock, as considered in the PDRP, is just one of the steps that need to be taken by railway operators in case a disruption occurs. An overview of the problems faced in this setting of real-time rescheduling is given by Cacchiani et al. (2014), Kroon and Huisman (2011) and Jespersen-Groth et al. (2009). These works also show the interrelation between these problems, where rolling stock rescheduling is generally performed after finding an updated timetable and is succeeded by steps such as crew rescheduling and the rescheduling of the shunting plans at the stations. Dollevoet et al. (2017) show that such a sequential approach often performs well for practical instances. Hence, we restrict ourselves to rolling stock rescheduling in this paper.

3.1 Rolling Stock (Re-)Scheduling Without Delays

Solution methods for the RSRP, that is without taking into account delays, are strongly linked to those for rolling stock scheduling. Early works on rolling stock scheduling of train units include those of Schrijver (1993), Ben-Khedher et al. (1998) and Abbink et al. (2004). One of the first papers to consider a similar setting for shunting, i.e., with the ideas of transitions and composition changes as considered here, is that of Alfieri et al. (2006). However, they assign the rolling stock for a single railway line opposed to for a whole network.

The first paper to consider the rolling stock scheduling problem as described in Section 2.1 is that of Fioole et al. (2006). They propose a Mixed Integer Programming (MIP) model to solve this problem, which is referred to as the *Composition Model*. It is based on a multi-commodity flow representation of the problem, where additionally the possible composition changes are taken into account. This approach is for this reason often described as a *flow based approach*. Nielsen (2011) has adapted the Composition Model for the setting of rolling stock rescheduling. Another flow based approach has been proposed by Borndörfer et al. (2016), who use a hypergraph-based model to solve a rolling stock scheduling problem that includes regularity and maintenance considerations.

A second solution approach that has been considered is a *path based approach*. This approach has been explored by Peeters and Kroon (2008), Cacchiani, Caprara, and Toth (2010) and Lusby et al. (2017). The main difference between these models is the decomposition that is used, where the decomposition is at a train level in the model of Peeters and Kroon (2008), but at the train unit level in the other models. The advantage of the latter is that individual train unit constraints, such as maintenance restrictions, can be taken into account. In all of these papers, column generation is used to solve the model, due to the exponential number of paths that have to be considered.

A comparison of the network flow and path-based approaches is made by Haahr et al. (2016), who compare the models of Fioole et al. (2006) and Lusby et al. (2017) on instances of NS and DSB S-tog. To achieve this, they extend the model of Lusby et al. (2017) to include the order of train units within a composition. In computational experiments, they find that both models are able to obtain solutions in reasonable time, but that the Composition Model is faster for most of the instances. However, they argue that these faster solution times for the Composition Model may be explained by the potential of the path-based approach to model constraints for individual train units. In this paper, we will consider both approaches, as it is unclear whether these results also extend to the setting of the PDRP. To do so, we extend the models of Fioole et al. (2006) and Lusby et al. (2017) to the setting of the PDRP.

3.2 Rescheduling for Delays

The main focus in railway disruption management when rescheduling for delays has traditionally been on Train Timetable Rescheduling (TTR). Rescheduling the timetable is often necessary as initial delays lead to timetable conflicts in which multiple trains require the same infrastructure at the same moment in time. The main objective in TTR is then to find a conflict-free timetable that minimizes the impact of these initial delays. An overview of the models that have been proposed for this problem can be found in Cacchiani et al. (2014).

Veelenturf et al. (2016) incorporate rolling stock requirements in TTR. In particular, they allow trips to be retimed or canceled in order to find a feasible timetable in case of large disruptions that lead to a (partial) blockage of railway infrastructure. Their objective is to minimize the total delay and the number of canceled trips. Moreover, to ensure that a feasible circulation can be found, they require that rolling stock is available to operate each trip. When compared to our paper, they do not allow changing the composition of trains during the day and thus take into account only a small part of the RSRP. Similarly, our paper does not take into account headway constraints between trains and thus takes into account only a small part of TTR.

Retiming is also dealt with by Veelenturf et al. (2012) for crew rescheduling. By retiming the moment of departure of trains slightly, they make it possible to find feasible crew schedules in cases where no feasible crew schedules can be found without retiming. These retiming decisions lead to delays in the system, which they take into account when determining the starting time for subsequent tasks. Note that also in this paper we implicitly retime trips, by changing the propagation of delay throughout the system. However, while Veelenturf et al. (2012) use retiming to make the crew schedule feasible, retiming in our paper follows from trying to minimize the impact that delays have on passengers.

Other papers have considered a more complete integration of rolling stock rescheduling and timetabling. Adenso-Díaz, González, and González-Torre (1999) determine the departure time of trips based on the assignment of rolling stock to the trips. Opposed to our paper, they consider locomotive hauled carriages of which the composition cannot be changed throughout the day. Cadarso, Marín, and Maróti (2013) consider the recovery of disruptions in rapid transit networks. They allow to insert emergency trips in the timetable and ensure that enough rolling stock is available to operate these trips and the non-canceled original trips. However, opposed to our paper, they do not consider delays for the original trips.

3.3 Rolling Stock Rescheduling for Delays

To the best knowledge of the authors, this paper is the first to consider rolling stock rescheduling with the goal of minimizing delays for passengers. In particular, we are unaware of any papers that take into account the effect that changing the chosen composition changes at transitions and that changing the turning pattern at stations has on delay propagation. In this paper, we aim to bridge this gap and employ rolling stock rescheduling as a means of minimizing passenger delays.

4 The Composition and Path Model for the RSRP

In this section, we describe two state-of-the-art models for solving the RSRP: the *Composition Model* as proposed by Fioole et al. (2006) and the *Path Model* as proposed by Haahr et al. (2016). To describe these models, we first formalize the problem description, where we follow the notation of Nielsen (2011).

Let $\mathcal{T}$ be the set of trips in the timetable and let $\mathcal{S}$ be the set of stations that trips arrive at and depart from. Let $\mathcal{C}$ be the set of transitions and let $s(c) \in \mathcal{S}$ be the station at which transition $c \in \mathcal{C}$ takes place.

Moreover, let IN_c denote the set of incoming trips at this transition and OUT_c the set of outgoing trips. Note that one of these sets can be empty for a transition that indicates the start or end of a train service. Let $\delta^-(t) \in \mathcal{C}$ and $\delta^+(t) \in \mathcal{C}$ indicate respectively the preceding and succeeding transition for trip $t \in \mathcal{T}$. Moreover, we define $\tau^-(c)$ as the moment in time at which a train unit that is uncoupled from the incoming trip of transition $c \in \mathcal{C}$ arrives at the shunting yard. Similarly, we define $\tau^+(c)$ as the moment in time at which a train unit needs to leave the shunting yard in order to be coupled to the outgoing trip for transition $c \in \mathcal{C}$.

Let $\mathcal{M}$ be the set of train unit types. Moreover, let the possible compositions that can be formed by these train unit types be given by $\mathcal{P}$. The set $\eta(t) \subseteq \mathcal{P}$ indicates the set of allowed compositions for trip $t \in \mathcal{T}$. Similarly, the set $\varrho(c)$ indicates the allowed composition changes for transition $c \in \mathcal{C}$. To describe the compositions in a composition change $q \in \varrho(c)$, let $p_{q,t} \in \eta(t)$ be the incoming composition of trip $t \in \mathrm{IN}_c$ for composition change q. Similarly, let $p'_{q,t} \in \eta(t)$ be the outgoing composition of trip $t \in \mathrm{OUT}_c$ for composition change q. Moreover, let $\iota^0_{s,m}$ be the starting inventory of units of type $m \in \mathcal{M}$ that are available at station $s \in \mathcal{S}$.

We can now formulate the shared components between the Composition Model and the Path Model. We consider the decision variables

$$X_{t,p} \coloneqq \begin{cases} 1 & \text{if composition } p \in \eta(t) \text{ is chosen for trip } t \in \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases}$$

$$Z_{c,q} \coloneqq \begin{cases} 1 & \text{if composition change } q \in \varrho(c) \text{ is chosen for transition } c \in \mathcal{C}, \\ 0 & \text{otherwise,} \end{cases}$$

$$I^{\infty}_{s,m} \in \mathbb{Z}_+ \coloneqq \text{the ending inventory of train units of type } m \in \mathcal{M} \text{ at station } s \in \mathcal{S}.$$

The shared constraints are given by

$$\sum_{p \in \eta(t)} X_{t,p} = 1 \qquad \forall t \in \mathcal{T}, \tag{1}$$

$$X_{t,p} = \sum_{q \in \varrho(\delta^+(t)) : p_{q,t} = p} Z_{\delta^+(t),q} \qquad \forall t \in \mathcal{T}, p \in \eta(t), \tag{2}$$

$$X_{t,p} = \sum_{q \in \varrho(\delta^-(t)) : p'_{q,t} = p} Z_{\delta^-(t),q} \qquad \forall t \in \mathcal{T}, p \in \eta(t). \tag{3}$$

Constraints (1) ensure that a feasible composition is chosen for each trip $t \in \mathcal{T}$. Constraints (2) and (3) link the X and Z variables by ensuring that the compositions on the incoming trip and outgoing trip of a transition match with the chosen composition change for this transition. What remains in formulating the RSRP is the connection between the chosen compositions and the availability of train units. Both models do this by means of modeling the natural flow of train units through the timetable, where the Composition Model is an arc formulation for this flow problem and the Path Model is a path formulation. We will describe these two approaches next.

4.1 The Composition Model

The Composition Model links the compositions to the available train units by explicitly modeling the inventory of train units at stations. It does so by keeping track of the number of train units that are coupled and uncoupled at each of the transitions and updating the inventory accordingly. Then, taking into account the limited availability of train units corresponds to requiring that the inventory is non-negative for every station and for every moment in time.

To define the inventory for each station, we introduce the decision variables

$$\begin{aligned}
I_{c,m} \in \mathbb{Z}_+ &\coloneqq \text{inventory of units of type } m \in \mathcal{M} \text{ at station } s(c) \text{ at time } \tau^+(c),\\
C_{c,m} \in \mathbb{Z}_+ &\coloneqq \text{amount of units of type } m \in \mathcal{M} \text{ that are coupled at transition } c \in \mathcal{C},\\
U_{c,m} \in \mathbb{Z}_+ &\coloneqq \text{amount of units of type } m \in \mathcal{M} \text{ that are uncoupled at transition } c \in \mathcal{C}.
\end{aligned}$$

Moreover, let $\gamma_{q,m}$ and $\upsilon_{q,m}$ indicate the number of train units of type $m \in \mathcal{M}$ that are respectively coupled and uncoupled at composition change q. The Composition Model is then given by

$$\begin{aligned}
& \min f(X, Z, I^{\infty}) && && (4)\\
\text{s.t. } & (1) - (3), && &&\\
& C_{c,m} = \sum_{q \in \varrho(c)} \gamma_{q,m} Z_{c,q} && \forall c \in \mathcal{C}, m \in \mathcal{M}, && (5)\\
& U_{c,m} = \sum_{q \in \varrho(c)} \upsilon_{q,m} Z_{c,q} && \forall c \in \mathcal{C}, m \in \mathcal{M}, && (6)\\
& I_{c,m} = \iota^0_{s(c),m} - \sum_{\substack{c' \in \mathcal{C}: s(c') = s(c),\\ \tau^+(c') \leq \tau^+(c)}} C_{c',m} + \sum_{\substack{c' \in \mathcal{C}: s(c') = s(c),\\ \tau^-(c') \leq \tau^+(c)}} U_{c',m} && \forall c \in \mathcal{C}, m \in \mathcal{M}, && (7)\\
& I^{\infty}_{s,m} = \iota^0_{s,m} - \sum_{c \in \mathcal{C}: s(c) = s} C_{c,m} + \sum_{c \in \mathcal{C}: s(c) = s} U_{c,m} && \forall s \in \mathcal{S}, m \in \mathcal{M}, && (8)\\
& X_{t,p} \in \{0, 1\} && \forall t \in \mathcal{T}, p \in \eta(t), && (9)\\
& Z_{c,q} \in \{0, 1\} && \forall c \in \mathcal{C}, q \in \varrho(c), && (10)\\
& I^{\infty}_{s,m} \in \mathbb{Z}_+ && \forall s \in \mathcal{S}, m \in \mathcal{M}, && (11)\\
& I_{c,m}, C_{c,m}, U_{c,m} \in \mathbb{Z}_+ && \forall c \in \mathcal{C}, m \in \mathcal{M}. && (12)
\end{aligned}$$

Constraints (1) – (3) are shared with the Path Model. Constraints (5) and (6) determine, for each train unit type, the number of train units that are respectively coupled and uncoupled at a transition. Constraints (7) define the inventory at the moments of coupling. Note that it is only needed to keep track of the inventory at the coupling moments $\tau^+(c)$, as a non-negative inventory at a coupling moment implies that the inventory was non-negative since the last coupling moment before it. Constraints (8) define the ending inventory for each station. The remaining constraints give the variable domains.

4.2 The Path Model

The Path Model links the availability of rolling stock to the compositions by considering train unit paths. Such a train unit path describes the trips that are operated by a train unit during the planning period. By assigning a single path to each train unit in the starting inventory, it is ensured that no more train units are used than are available.

To formulate the Path Model, let Π be the set of all feasible train unit paths, i.e., all those sequences of trips that can be operated by a single train unit during the planning horizon. Moreover, let $\Pi_m \subseteq \Pi$ be the set of train unit paths for a train unit of type $m \in \mathcal{M}$ and let $b(\pi)$ and $e(\pi)$ indicate respectively the starting and ending station of path $\pi \in \Pi$. Furthermore, let ω^t_π indicate whether path $\pi \in \Pi$ contains trip

$t \in \mathcal{T}$ and let μ_p^m indicate the number of train units of type $m \in \mathcal{M}$ that are present in composition $p \in \mathcal{P}$. If we consider the decision variables

$$L_\pi := \begin{cases} 1 & \text{if path } \pi \in \Pi \text{ is operated by a train unit,} \\ 0 & \text{otherwise,} \end{cases}$$

the Path Model is given by

$$\begin{aligned}
& \min f(X, Z, I^\infty) && && (13)\\
\text{s.t. } & (1) - (3), && && \\
& \sum_{\pi \in \Pi_m} \omega_\pi^t L_\pi = \sum_{p \in \eta(t)} \mu_p^m X_{t,p} && \forall m \in \mathcal{M}, t \in \mathcal{T}, && (14)\\
& \sum_{\pi \in \Pi_m : b(\pi) = s} L_\pi = \iota_{s,m}^0 && \forall m \in \mathcal{M}, s \in \mathcal{S}, && (15)\\
& \sum_{\pi \in \Pi_m : e(\pi) = s} L_\pi = I_{s,m}^\infty && \forall m \in \mathcal{M}, s \in \mathcal{S}, && (16)\\
& L_\pi \in \{0, 1\} && \forall \pi \in \Pi, && (17)\\
& X_{t,p} \in \{0, 1\} && \forall t \in \mathcal{T}, p \in \eta(t), && (18)\\
& Z_{c,q} \in \{0, 1\} && \forall c \in \mathcal{C}, q \in \varrho(c), && (19)\\
& I_{s,m}^\infty \in \mathbb{Z}_+ && \forall s \in \mathcal{S}, m \in \mathcal{M}. && (20)
\end{aligned}$$

Constraints (1) – (3) are shared with the Composition Model. Constraints (14) link the compositions and the paths, by ensuring that enough paths visit a trip to form the composition that is chosen for this trip. Constraints (15) ensure that the starting inventory is respected, by assigning a single path to each train unit present in the starting inventory. Similarly, constraints (16) determine the ending inventory at each station. The remaining constraints define the domains of the variables.

5 Modeling the PDRP

In this section, we propose two models to solve the PDRP: the *Delay Composition Model* and the *Delay Path Model*. These models extend respectively the Composition Model and Path Model, as described in the last section, to the setting of the PDRP. We present the extensions to these models in two steps: we first model the delay propagation that is caused by an initial delay and afterwards model flexible turning and its impact on delay propagation. The complete models can also be found in Appendix A and Appendix B, respectively.

5.1 Delay Propagation

We use some additional notation for the delays and delay propagation. Let $\mathcal{D} = \{0, 1, \ldots, d_u\} \subseteq \mathbb{Z}_+$ be the set of delay sizes, with d_u an upper bound on the possible delays. Moreover, let $\mathcal{T}_{\text{init}} \subseteq \mathcal{T}$ be the set of initially delayed trips, i.e., representing the primary delays in the system, with a delay of $d_t \in \mathcal{D}$ for trip $t \in \mathcal{T}_{\text{init}}$. Let $\Delta_t \in \mathbb{Z}_+$ be the running time supplement for trip $t \in \mathcal{T}$. Similarly, let $\Delta_c \in \mathbb{Z}_+$ be the supplement available in the planned dwell time at transition $c \in \mathcal{C}$. Lastly, let $\Delta_s \in \mathbb{Z}_+$ be the supplement to the time that is required at station $s \in \mathcal{S}$ between uncoupling a train unit from one composition and subsequently coupling it to another composition.

To keep track of the delay that a trip has, we introduce the decision variables

$$Y_t \in \mathcal{D} \coloneqq \text{the delay that trip } t \in \mathcal{T} \text{ has at the moment of arrival.}$$

Furthermore, we introduce the constraints

$$Y_t = d_t \qquad \forall t \in \mathcal{T}_{\text{init}} \tag{21}$$

to model the initial delays.

To model the propagation of these initial delays, we distinguish again between delay that is propagated at transitions and delay that is propagated due to the shunting of train units. We will refer to delay propagated at a transition as *predecessor propagation*, as delay is passed from a predecessor trip to a successor trip. Delay propagation due to a delayed train unit being coupled at a transition is referred to as *inventory propagation*. In this section, we cover both predecessor and inventory propagation for the Delay Composition Model and Delay Path Model.

We include predecessor propagation in both models by the constraints

$$Y_t \geq Y_{t'} - \Delta_{\delta^-(t)} - \Delta_t \qquad \forall t \in \mathcal{T}, t' \in \text{IN}_{\delta^-(t)}. \tag{22}$$

In this case, the delay of any successor trip in the transition is at least as large as that of any predecessor trip, minus any of the delay that is absorbed during either the successor trip itself or during the transition that precedes it. It remains to include inventory propagation, which is done separately for the two considered models.

5.1.1 Delay Composition Model

In the Delay Composition Model, we take into account inventory propagation by means of keeping track of the delay that train units have when they enter and leave the inventory, i.e., when being uncoupled from a composition and being coupled to another composition. The delay of a trip, as caused by the coupling of delayed train units, is then determined by looking at the delay with which these units have left the inventory. Moreover, we take the delays at uncoupling and coupling into account when determining the number of train units that are present in the inventory at any moment in time.

To model this formally, we define the decision variables

$$\begin{aligned}
I_{c,m,d} \in \mathbb{Z}_+ &\coloneqq \text{the size of the inventory of type } m \text{ at station } s(c) \text{ at time } \tau^+(c) + d,\\
C_{c,m,d} \in \mathbb{Z}_+ &\coloneqq \text{the number of units of type } m \text{ that are coupled at transition } c \text{ with delay } d \in \mathcal{D},\\
U_{c,m,d} \in \mathbb{Z}_+ &\coloneqq \text{the number of units of type } m \text{ that are uncoupled at transition } c \text{ with delay } d \in \mathcal{D},\\
D^i_{c,d} &\coloneqq \begin{cases} 1 & \text{if the incoming (uncoupled) units for transition } c \text{ have a delay of } d \in \mathcal{D},\\ 0 & \text{otherwise,}\end{cases}\\
D^o_{c,d} &\coloneqq \begin{cases} 1 & \text{if the outgoing (coupled) units for transition } c \text{ have a delay of } d \in \mathcal{D},\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

We determine the delay that is incurred at uncoupling and the number of units that are uncoupled with this delay by the constraints:

$$\sum_{d\in\mathcal{D}} dD^i_{c,d} \geq Y_t \qquad \forall c \in \mathcal{C}, t \in \mathrm{IN}_c, \tag{23}$$

$$\sum_{d\in\mathcal{D}} D^i_{c,d} = 1 \qquad \forall c \in \mathcal{C}, \tag{24}$$

$$U_{c,m,d} \leq M_1 D^i_{c,d} \qquad \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D}, \tag{25}$$

$$\sum_{d\in\mathcal{D}} U_{c,m,d} = U_{c,m} \qquad \forall c \in \mathcal{C}, m \in \mathcal{M}. \tag{26}$$

Constraints (23) and (24) link the delay at the moment of uncoupling to the delay of the predecessor trip. Constraints (25) link the number of delayed uncoupled units to the delay at uncoupling. The constant M_1 can be chosen here as the maximum number of units that can be uncoupled, which is generally no more than five in practical instances. Lastly, constraints (26) ensure that for each uncoupled train unit an appropriate delay is selected.

We introduce similar constraints for the coupling of train units to trips:

$$Y_t \geq dD^o_{\delta^-(t),d} - \Delta_t \qquad \forall t \in \mathcal{T}, d \in \mathcal{D}, \tag{27}$$

$$C_{c,m,d} \leq M_2 D^o_{c,d} \qquad \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D}, \tag{28}$$

$$\sum_{d\in\mathcal{D}} C_{c,m,d} = C_{c,m} \qquad \forall c \in \mathcal{C}, m \in \mathcal{M}. \tag{29}$$

Constraints (27) determine the delay of an outgoing trip. Constraints (28) relate the delay at coupling to the number of units that are coupled with a delay. The constant M_2 can be chosen here as the maximum number of units that can be coupled at this transition, which is again no larger than five for practical instances. Lastly, constraints (29) ensure that for each coupled train unit an appropriate delay is selected.

We link the delays that units have at the moments of uncoupling to the delays that units have at the moment of coupling by means of the inventory. To do so, we replace constraints (7) by

$$\begin{aligned} I_{c,m,d} = \iota^0_{s(c),m} - &\sum_{\substack{c'\in\mathcal{C}:\\ s(c')=s(c)}} \sum_{\substack{d'\in\mathcal{D}:\\ \tau^+(c')+d'\leq\tau^+(c)+d}} C_{c',m,d'} \\ + &\sum_{\substack{c'\in\mathcal{C}:\\ s(c')=s(c)}} \sum_{\substack{d'\in\mathcal{D}:\\ \tau^-(c')+\left(d'-\Delta_{s(c')}\right)^+\leq\tau^+(c)+d}} U_{c',m,d'} \qquad \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D}, \end{aligned} \tag{30}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. Constraints (30) record the number of available train units at a station for each moment $\tau^+(c) + d$. Taking into account the delays at coupling, these form all the time moments at which a train unit can leave the inventory of a station. Similarly to the definition of the regular inventory, the current inventory equals the number of uncoupled train units minus the number of coupled units on top of the train units that are present at the station at the start of the day. However, we now need to take into account the delays with which these train units are coupled and uncoupled to determine if they are present at the station.

5.1.2 Delay Path Model

To model inventory propagation in the Delay Path Model, we exploit the fact that we have a train unit path for each of the train units. In particular, we extend these paths in such a way that we can record at which trip a train unit is uncoupled before being coupled to another trip. The delay on the trip at which this unit is coupled then becomes related to the delay on the trip at which the train unit was uncoupled.

Keeping track of uncoupling and coupling between trips is only necessary when the time that a train unit stays in the inventory between these trips is short, as the long buffer between the trips will otherwise ensure that no delay is propagated between these trips. For this reason, let $\mathrm{SS}_t \subset \mathcal{T}$ be the set of trips that can send a unit to the inventory which can delay trip $t \in \mathcal{T}$. A trip t' is thus present in SS_t if the time between the moment of uncoupling from t' and the departure of trip t is not large enough to absorb all delay that can be present on trip t'. We refer to such a shunting movement that can propagate delay as a *short shunting*. Let the binary parameter $\zeta_\pi^{t',t}$ now indicate whether a train unit following path $\pi \in \Pi$ is uncoupled at trip $t' \in \mathrm{SS}_t$ and consecutively coupled at trip t.

We introduce the decision variables

$$S_{t',t} := \begin{cases} 1 & \text{if any train unit is uncoupled at } t' \in \mathrm{SS}_t \text{ and consecutively coupled at } t \in \mathcal{T}, \\ 0 & \text{otherwise.} \end{cases}$$

We then model inventory propagation by means of the constraints

$$\sum_{\pi \in \Pi} \zeta_\pi^{t',t} L_\pi \leq M_3 S_{t',t} \qquad \forall t \in \mathcal{T}, t' \in \mathrm{SS}_t, \tag{31}$$

$$Y_t \geq Y_{t'} - d_u(1 - S_{t',t}) + \tau^-(\delta^+(t')) - \tau^+(\delta^-(t)) - \Delta_{s(\delta^-(t))} - \Delta_t \qquad \forall t \in \mathcal{T}, t' \in \mathrm{SS}_t. \tag{32}$$

Constraints (31) determine if there are any train units that are involved in a short shunting between a trip $t \in \mathcal{T}$ and a trip $t' \in \mathrm{SS}_t$. Constraints (32) then determine the delay propagation between such trips while taking into account the absorption of the delays, where delay is passed only if short shunting occurs between the trips. Note that the constant M_3 can be chosen relatively small in practical cases, as it can be set equal to the maximum of the number of uncoupled units from t' and the number of coupled units to t. Both of these numbers tend to be smaller than five for practical instances.

5.2 Flexible Turning

To allow for flexible turning, we follow the modeling approach of Nielsen (2011), which consists of defining a pool of compositions that are at the platforms of a station. Just like the inventory defines the number of train units present at the shunting yard of a station, this pool of compositions defines the number of compositions of a certain type that are *parked* at the platforms of a station. Incoming trips at this terminal station can add compositions to this pool by *dropping off* compositions at the platform, which can then be *picked up* by outgoing trips. In the remainder of this section, we present an adaptation of the model that was proposed by Nielsen (2011) for the setting of the Composition Model. Moreover, we extend this approach to the Delay Path Model and determine the delay propagation that is incurred by a certain turning pattern.

We again need some additional notation. Let $\mathcal{S}_f \subseteq \mathcal{S}$ be the set of stations where flexible turning can occur. Let $\mathcal{T}_i^f \subseteq \mathcal{T}$ be the set of incoming trips that can engage in flexible turning at their arrival station. Similarly, let $\mathcal{T}_o^f \subseteq \mathcal{T}$ be the set of outgoing trips that can engage in flexible turning at their departure station. Let $\widetilde{\mathcal{C}}$ be the set of transitions at which flexible turning can occur. Moreover, let $\mathrm{UN}_{p,t}$ be the set of compositions that can be dropped off at the platform if trip $t \in \mathcal{T}_i^f$ arrives with composition $p \in \eta(t)$. Similarly, let $\mathrm{CO}_{p,t}$ be the set of compositions that can be picked up from the platform if trip $t \in \mathcal{T}_o^f$ departs with composition $p \in \eta(t)$. Furthermore, let $\tau_a(c)$ and $\tau_d(c)$ denote the moment that respectively the composition on the incoming trip of transition $c \in \widetilde{\mathcal{C}}$ becomes available at the platform and the moment that the outgoing trip of c departs.

We keep track of the pool of compositions by means of the decision variables

$$Q_{t,p,p'} := \begin{cases} 1 & \text{if trip } t \in \mathcal{T}_i^f \text{ arrives with composition } p \in \eta(t) \text{ before dropping off composition } p' \in \mathrm{UN}_{p,t}, \\ 0 & \text{otherwise,} \end{cases}$$

$$W_{t,p,p'} := \begin{cases} 1 & \text{if trip } t \in \mathcal{T}_o^f \text{ departs with composition } p \in \eta(t) \text{ after picking up composition } p' \in \mathrm{CO}_{p,t}, \\ 0 & \text{otherwise,} \end{cases}$$

$P_{c,p} \in \mathbb{Z}_+ :=$ the number of compositions of type $p \in \mathcal{P}$ that are parked at station $s(c)$ at time $\tau_d(c)$.

We start by specifying the constraints that are shared between the two models. To allow for flexible turning at the terminal stations we introduce the constraints

$$X_{t,p} = \sum_{q \in \varrho(\delta^+(t)): p_{q,t} = p} Z_{\delta^+(t),q} + \sum_{p' \in \mathrm{UN}_{p,t}} Q_{t,p,p'} \qquad \forall t \in \mathcal{T}_i^f, p \in \eta(t), \tag{33}$$

$$X_{t,p} = \sum_{q \in \varrho(\delta^-(t)): p'_{q,t} = p} Z_{\delta^-(t),q} + \sum_{p' \in \mathrm{CO}_{p,t}} W_{t,p,p'} \qquad \forall t \in \mathcal{T}_o^f, p \in \eta(t). \tag{34}$$

Constraints (33) replace (2) for each trip $t \in \mathcal{T}_i^f$. Similarly, constraints (34) replace (3) for each trip $t \in \mathcal{T}_o^f$. These two sets of constraints extend the original constraints by including the possibility that a flexible turning can occur respectively after or before a trip. Hence, either the normal transition should be maintained or a composition should be respectively dropped off or picked up at the platform.

We determine the number of parked compositions at any moment in time by means of the constraints

$$P_{c,p} = \sum_{\substack{c' \in \widetilde{\mathcal{C}}: s(c') = s(c), \\ \tau_a(c') \le \tau_d(c)}} \sum_{t \in \mathrm{IN}_{c'}} \sum_{p': p \in \mathrm{UN}_{p',t}} Q_{t,p',p} - \sum_{\substack{c' \in \widetilde{\mathcal{C}}: s(c') = s(c), \\ \tau_d(c') \le \tau_d(c)}} \sum_{t \in \mathrm{OUT}_{c'}} \sum_{p': p \in \mathrm{CO}_{p',t}} W_{t,p',p} \quad \forall c \in \widetilde{\mathcal{C}}, p \in \mathcal{P}. \tag{35}$$

The number of parked compositions is determined as the difference between the number of compositions that have been dropped off so far and the number of compositions that have been picked up so far. Note that it is only required to determine the number of available compositions at the moment of departure of an outgoing trip, as a non-negative size of the pool at these moments in time ensures that the size of the pool is non-negative at any moment in time.

Moreover, we need to take into account the use of flexible turning when determining the delay that is passed to a successor trip. This is achieved jointly for both models by replacing constraints (22) by the constraints

$$Y_t \ge Y_{t'} - \Delta_{\delta^-(t)} - \Delta_t - d_u \left(\sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} W_{t,p,p'} \right) \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{IN}_{\delta^-(t)}. \tag{36}$$

for each trip $t \in \mathcal{T}_o^f$. Constraints (36) state that the delay propagation from the predecessor trip to the successor trip of a transition should be taken into account unless the trip picks up a composition from the platform. In that case, no rolling stock moves from the predecessor to the successor trip of the transition and hence no delay propagation occurs between these trips. It remains to link flexible turning to the available inventory of train units. Moreover, the delay propagated through the chosen turning pattern needs to be taken into account for both models.

5.2.1 Delay Composition Model

In the Delay Composition Model, flexible turning is linked to the available number of train units by means of the inventory. First, we redefine the number of coupled and uncoupled units at a transition to take into account any shunting that is executed during a flexible turning. We replace the sets of constraints (5) and (6) by

$$C_{c,m} = \sum_{q \in \varrho(c)} \gamma_{q,m} Z_{c,q} + \sum_{t \in \mathrm{OUT}_c} \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} (\mu_p^m - \mu_{p'}^m) W_{t,p,p'} \qquad \forall c \in \widetilde{\mathcal{C}}, m \in \mathcal{M}, \tag{37}$$

$$U_{c,m} = \sum_{q \in \varrho(c)} \upsilon_{q,m} Z_{c,q} + \sum_{t \in \mathrm{IN}_c} \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{UN}_{p,t}} (\mu_p^m - \mu_{p'}^m) Q_{t,p,p'} \qquad \forall c \in \widetilde{\mathcal{C}}, m \in \mathcal{M}, \tag{38}$$

for each $c \in \widetilde{\mathcal{C}}$. Constraints (37) and (38) record respectively the number of coupled and uncoupled units, where we take into account specifically any shunting that occurs during flexible turning.

Moreover, we need to redefine the ending inventory. Here, we assume that any parked compositions that are not picked up end up in the ending inventory. Hence, we replace the set of constraints (8) by

$$\begin{aligned} I_{s,m}^{\infty} = \iota_{s,m}^0 - \sum_{c \in \mathcal{C}: s(c)=s} C_{c,m} + \sum_{c \in \mathcal{C}: s(c)=s} U_{c,m} + \sum_{c \in \widetilde{\mathcal{C}}: s(c)=s} \sum_{t \in \mathrm{IN}_c} \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{UN}_{p,t}} \mu_{p'}^m Q_{t,p,p'} \\ - \sum_{c \in \widetilde{\mathcal{C}}: s(c)=s} \sum_{t \in \mathrm{OUT}_c} \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} \mu_{p'}^m W_{t,p,p'} \qquad \forall s \in \mathcal{S}_f, m \in \mathcal{M}, \end{aligned} \tag{39}$$

for each station $s \in \mathcal{S}_f$.

We now consider the delay propagation that occurs when making use of flexible turning. Similar to what we did for inventory propagation, we explicitly keep track of the delays of dropped off and picked up compositions for flexible turning. We introduce the decision variables

$$F_{t,d,p}^i := \begin{cases} 1 & \text{if trip } t \in \mathcal{T}_i^f \text{ drops off composition } p \in \mathcal{P} \text{ at the platform with delay } d \in \mathcal{D}, \\ 0 & \text{otherwise,} \end{cases}$$

$$F_{t,d,p}^o := \begin{cases} 1 & \text{if trip } t \in \mathcal{T}_o^f \text{ picks up composition } p \in \mathcal{P} \text{ from the platform with delay } d \in \mathcal{D}, \\ 0 & \text{otherwise,} \end{cases}$$

$P_{c,p,d} \in \mathbb{Z}_+ :=$ the number of compositions of type $p \in \mathcal{P}$ that are parked at station $s(c)$ at time $\tau_d(c) + d$.

First, we determine the delay with which compositions are dropped off by

$$F_{t,d,p}^i \leq D_{\delta^+(t),d}^i \qquad \forall t \in \mathcal{T}_i^f, p \in \mathcal{P}, d \in \mathcal{D}, \tag{40}$$

$$\sum_{d \in \mathcal{D}} F_{t,d,p}^i = \sum_{p': p \in \mathrm{UN}_{p',t}} Q_{t,p',p} \qquad \forall t \in \mathcal{T}_i^f, p \in \mathcal{P}. \tag{41}$$

Constraints (40) ensure that the delay of the dropped off composition matches that of the trip that precedes the turning. Constraints (41) ensure that if a composition is dropped off at the station, then also a delay is selected with which this composition is dropped off.

Similarly, the picking up of compositions gives rise to

$$F^o_{t,d,p} \leq D^o_{\delta^-(t),d} \qquad \forall t \in \mathcal{T}^f_o, p \in \mathcal{P}, d \in \mathcal{D}, \tag{42}$$

$$\sum_{d \in \mathcal{D}} F^o_{t,d,p} = \sum_{p' : p \in \mathrm{CO}_{p',t}} W_{t,p',p} \qquad \forall t \in \mathcal{T}^f_o, p \in \mathcal{P}. \tag{43}$$

Constraints (42) ensure that the correct delay is propagated to the successor trip of the flexible turning. Constraints (43) ensure that if a composition is picked up from the station, then also a delay is selected with which this composition is picked up.

Lastly, we redefine the number of compositions that are parked at the platforms of a station to take into account any delays that are faced at dropping off and picking up compositions. We replace the set of constraints (35) by

$$\begin{aligned} P_{c,p,d} = & \sum_{c' \in \widetilde{\mathcal{C}} : s(c') = s(c)} \sum_{\substack{d' \in \mathcal{D}: \\ \tau_a(c') + \left(d' - \Delta_{c'}\right)^+ \leq \tau_d(c) + d}} \sum_{t \in \mathrm{IN}_{c'}} F^i_{t,d',p} \\ & - \sum_{c' \in \widetilde{\mathcal{C}} : s(c') = s(c)} \sum_{\substack{d' \in \mathcal{D}: \\ \tau_d(c') + d' \leq \tau_d(c) + d}} \sum_{t \in \mathrm{OUT}_{c'}} F^o_{t,d',p} \qquad \forall c \in \widetilde{\mathcal{C}}, p \in \mathcal{P}, d \in \mathcal{D}. \end{aligned} \tag{44}$$

Constraints (44) count the number of parked compositions of some type $p \in \mathcal{P}$ at each time instant $\tau_d(c) + d$. Taking into account each delay $d \in \mathcal{D}$, these are all the possible moments that a composition can leave the pool of compositions. Again, the number of parked compositions of this type is the number of dropped off compositions of this type minus the number of picked up compositions of this type.

5.2.2 Delay Path Model

In the Delay Path Model, it remains to link the train unit paths to the dropped off and picked up compositions. Again, we introduce some additional properties on the paths. Let $\phi^a_{t,\pi}$ indicate whether a train unit that operates path $\pi \in \Pi$ is dropped off at the platform after trip $t \in \mathcal{T}^f_i$. Similarly, let $\phi^d_{t,\pi}$ indicate whether a train unit that operates path $\pi \in \Pi$ is picked up at the platform before trip $t \in \mathcal{T}^f_o$. We enforce that the dropped off and picked up compositions match the selected paths by means of the constraints

$$\sum_{\pi \in \Pi_m} \phi^a_{t,\pi} L_\pi = \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{UN}_{p,t}} \mu^m_{p'} Q_{t,p,p'} \qquad \forall t \in \mathcal{T}^f_i, m \in \mathcal{M}, \tag{45}$$

$$\sum_{\pi \in \Pi_m} \phi^d_{t,\pi} L_\pi = \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} \mu^m_{p'} W_{t,p,p'} \qquad \forall t \in \mathcal{T}^f_o, m \in \mathcal{M}. \tag{46}$$

Constraints (45) and (46) ensure that enough train units are respectively dropped off and picked up from the platform to form the composition that is dropped off or picked up. In contrast to the Delay Composition Model, the ending inventory is automatically adjusted for the remaining compositions in case of the Delay Path Model.

It remains to model the delay propagation that is implied by the chosen turning pattern. We model this in a similar way as that we model inventory propagation for the Delay Path Model. Let ST_t be the set of incoming trips that can drop off a composition at a platform which can delay trip t when this composition is picked up at trip t. Let $\xi_\pi^{t',t}$ indicate whether a train unit is dropped off in a composition at trip $t' \in \mathrm{ST}_t$ and picked up from the platform before operating on trip t when this train unit operates path π. We refer to such a situation as a *short turning*.

To model the delay propagation through short turning, we introduce the decision variables

$$B_{t',t,p} = \begin{cases} 1 & \text{if a short turning occurs between trips } t \text{ and } t' \in \mathrm{ST}_t \text{ with composition } p, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, let $\Delta(t,t')$ indicate the delay absorption when a composition is dropped off at trip t' and picked up at trip t. We model short turning and the corresponding delay propagation by

$$\sum_{\pi \in \Pi_m} \xi_\pi^{t',t} L_\pi = \sum_{p \in \eta(t')} \sum_{p' \in \mathrm{UN}_{p,t'}} \mu_{p'}^m B_{t',t,p'} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, m \in \mathcal{M}, \tag{47}$$

$$Y_t \geq Y_{t'} - \Delta_t - M_4 \left(1 - \sum_{p \in \eta(t')} \sum_{p' \in \mathrm{UN}_{p,t'}} B_{t',t,p'}\right) - \Delta(t,t') \quad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, \tag{48}$$

$$B_{t',t,p'} \leq \sum_{p^* \in \eta(t')} Q_{t',p^*,p'} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, p \in \eta(t'), p' \in \mathrm{UN}_{p,t'}. \tag{49}$$

$$B_{t',t,p'} \leq \sum_{p^* \in \eta(t)} W_{t,p^*,p'} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, p \in \eta(t'), p' \in \mathrm{UN}_{p,t'}. \tag{50}$$

Constraints (47) link the train unit paths to the composition which is used in the short turning. Constraints (48) then propagate the delay between the two relevant trips in case a short turning is actually present, taking into account any absorption of delay. Constraints (49) link the composition that is involved in the short turning to the composition that is dropped off at the platform. Similarly, constraints (50) link the composition that is involved in the short turning to the composition that is picked up from the platform. Moreover, note that the constant M_4 can be chosen as $M_4 = d_u - \Delta_t - \Delta(t,t')$ to make sure that no delay is passed in the case that no short turning occurs.

Lastly, we adjust the number of parked compositions for the delays on the incoming and outgoing compositions. We can do so by replacing constraints (35) by

$$\begin{aligned} P_{c,p} = & \sum_{\substack{c' \in \widetilde{\mathcal{C}}: s(c')=s(c), \\ \tau_a(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{IN}_{c'}} \sum_{p': p \in \mathrm{UN}_{p',t}} Q_{t,p',p} - \sum_{\substack{c' \in \widetilde{\mathcal{C}}: s(c')=s(c), \\ \tau_d(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{OUT}_{c'}} \sum_{p': p \in \mathrm{CO}_{p',t}} W_{t,p',p} \\ & + \sum_{\substack{c' \in \widetilde{\mathcal{C}}: s(c)=s(c') \\ \tau_d(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{OUT}_{c'}} \sum_{\substack{t' \in \mathrm{ST}_t: \\ \tau_a(\delta^+(t')) > \tau_d(c)}} B_{t',t,p} \qquad \forall c \in \widetilde{\mathcal{C}}, p \in \mathcal{P}. \end{aligned} \tag{51}$$

Constraints (51) correct for the fact that the temporal ordering between trips may be changed when outgoing trips leave with a delay. In particular, it may occur that an outgoing trip uses a composition that would in the original timetable only arrive after the outgoing trip has already departed.

6 Solution Approaches

In this section, we propose solution methods for the Delay Composition Model and the Delay Path Model. Both of these models are Mixed Integer Programming (MIP) models, but the dimensions of these models vary considerably. While the complete Delay Composition Model can generally be stored in memory, the number of paths (variables) in the Delay Path Model is generally so large that the model does not fit in memory. As a consequence, we use an off-the-shelf MIP solver for the Delay Composition Model, while we propose a specialized method for the Delay Path Model in the remainder of this section.

Following the approach of Haahr et al. (2016), we solve the Delay Path Model by Branch-and-Price (B&P). Here, column generation is used in each node of a Branch-and-Bound tree to generate promising variables dynamically, alleviating the need of enumerating all variables. To apply column generation, the problem is split in a (restricted) master and pricing problem, where the master problem provides dual information with respect to the current optimal solution, while the pricing problem finds new variables (columns) based on this dual information. These problems are then solved iteratively until no more promising variables can be generated, proving that the current node in the Branch-and-Bound tree is solved to optimality. For a general overview of Branch-and-Price we refer the reader to Barnhart et al. (1998).

In our method, the (restricted) master problem corresponds to solving the linear relaxation of the Delay Path Model proposed in Section 5 with a subset of all train unit paths. This set is initialized in the root node with all those paths that are present in the original solution, i.e., the planned circulation for the undisturbed situation, complemented with a set of artificial columns that ensure feasibility. The pricing problem corresponds to generating new train unit paths and decomposes over the different train unit types.

6.1 Solving the Pricing Problem

The pricing problem for each train unit type corresponds to a shortest path problem in a directed graph $G = (V, A)$. This graph extends the one considered by Haahr et al. (2016). For the set of nodes V, we introduce for each trip $t \in \mathcal{T}$ a node v_t^d representing the departure of t and a node v_t^a representing the arrival of t. Moreover, we introduce for each trip $t \in \mathcal{T}$ an inventory node v_t^c that indicates the departure of train units from the inventory before being coupled to t. To model flexible turning, we introduce a picking up node v_t^{pi} and a short turning node v_t^{st} for each trip $t \in \mathcal{T}_o^f$. Finally, a source node v_s and sink node v_t are added to the graph to connect the starting and ending inventories of the different stations.

The arcs A represent actions taken by the train units. We introduce an arc (v_t^d, v_t^a) for all $t \in \mathcal{T}$ to indicate that a train unit is part of the composition on trip t. Moreover, we introduce for each transition $c \in \mathcal{C}$ and between any pair $t_1 \in \mathrm{IN}_c, t_2 \in \mathrm{OUT}_c$ a connection arc $(v_{t_1}^a, v_{t_2}^d)$ that indicates that a train unit operates on trip t_2 directly after trip t_1. Next, we construct the station arcs. These are added between those inventory nodes v_t^c that occur at the same station and that are consecutive in time. Moreover, the first inventory node for each station is connected to the source v_s and the last inventory node to the sink v_t.

We now represent the coupling and uncoupling of train units. We add an arc (v_t^c, v_t^d) for each trip $t \in \mathcal{T}$ to indicate that a train unit is coupled to t. For uncoupling, we have to take into account short shunting. Consider some trip $t' \in \mathcal{T}$. For each trip $t \in \mathcal{T}$ such that $t' \in \mathrm{SS}_t$, we introduce an arc $(v_{t'}^a, v_t^d)$ that represents that a train unit is involved in a short shunting between these two trips. In Figure 4, this is indicated by the dotted arc from $v_{t_1}^a$ to $v_{t_4}^d$. Moreover, we assign an additional arc to the first coupling moment for which no delay can be passed. Specifically, assign an arc $(v_{t'}^a, v_{t^*}^c)$, such that t^* is the first trip in time such that $\tau^-(\delta^+(t')) < \tau^+(\delta^-(t^*))$, $s(\delta^+(t')) = s(\delta^-(t^*))$, and $t' \notin \mathrm{SS}_{t^*}$. This arc indicates that the train unit is moved to the inventory and only coupled to a trip at a moment that no delay can be passed. In Figure 4, this is indicated by the dotted arc between $v_{t_1}^a$ and $v_{t_5}^c$.

We now represent the picking up and dropping off of compositions at station $s \in \mathcal{S}_f$. We introduce the arcs $(v_{t'}^a, v_t^{st})$ and (v_t^{st}, v_t^d) for each pair (t', t) such that $t \in \mathcal{T}_o^f$ and $t' \in \mathrm{ST}_t$. The dotted arcs from $v_{t_1}^a$ to $v_{t_4}^{st}$ and from $v_{t_4}^{st}$ to $v_{t_4}^d$ in Figure 5 serve as an example. These arcs represent that a train unit is involved in a

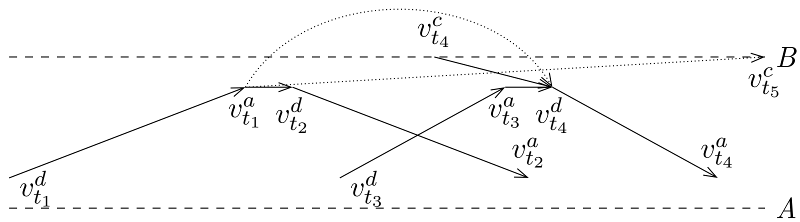


Figure 4: Example of the graph found in the pricing problem of the Delay Path Model. Note the dotted short shunting arc between trip t_1 and trip t_4. Moreover, note the dotted uncoupling arc from trip t_1 to trip t_5 that indicates that no delay can be passed.

short turning between trips t' and t. Moreover, we add an arc $(v_{t'}^a, v_{t^*}^{pa})$ to indicate that some compositions may only be picked up at a trip at which already all delay has been absorbed. Here, t^* indicates the first trip in time such that $\tau_a(\delta^+(t')) < \tau_d(\delta^-(t^*))$, $s(\delta^+(t')) = s(\delta^-(t^*))$ and $t' \notin \mathrm{ST}_{t^*}$. In addition, we connect those picking up nodes v_t^{pi} that occur at the same station and are consecutive in time. These arcs represent that a train unit is part of a composition that is parked at the platform of a station. Lastly, we connect the last picking up node for each station to the sink v_t.

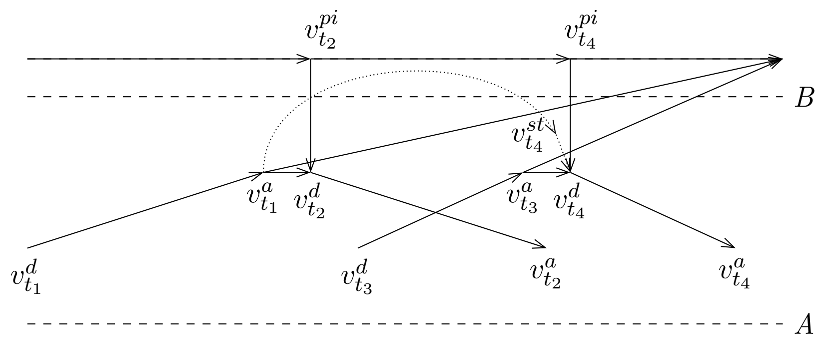


Figure 5: Example of the graph found in the pricing problem of the Delay Path Model, where flexible turning is allowed at station B. Note the picking-up arcs for respectively trip t_2 and trip t_4. Moreover, note the dotted short turning arc between trip t_1 and t_4 and the arcs that represent dropping off in case no delay can be passed.

The pricing problem can now be formulated as that of finding a shortest path from source node v_s to sink node v_t. The costs on the arcs are determined from the duals in the master problem for the considered train unit type. If a shortest path of negative length can be found, it corresponds to a column with negative reduced cost and this column is added to the master problem. Alternatively, if no such path can be found for all train unit types, the current node in the Branch-and-Bound tree is solved to optimality.

6.2 Branching Scheme

To obtain an integral solution, we additionally need to impose branching decisions. We adopt all of the branching rules that are used by Haahr et al. (2016). These include the branching on a fractional flow of units that leave the starting inventory of a station, branching on a fractional flow of units that enter the ending inventory of a station and branching on a fractional flow of units that operate on a trip. Moreover, Haahr et al. (2016) branch on subsets of compositions which together have a fractional assignment.

The branching decisions of Haahr et al. (2016) ensure that the composition variables and ending inventory variables are integral, but fractional solutions can still occur in the setting of the PDRP. First, the variables associated to short shunting and short turning may be fractional. Consider the case where some short turning variable $B_{t',t,p}$ is fractional. We then create the branches

$$B_{t',t,p} = 0 \quad \text{and} \quad B_{t',t,p} = 1.$$

We perform branching in a similar way for the short shunting variables $S_{t',t}$.

Moreover, we have to ensure that the composition change variables and the variables that represent the dropping off and picking up of compositions are integral. If no flexible turning can occur at a transition, the integrality of the composition change variables follows from integrality on the composition variables. Let us instead consider a transition where flexible turning can occur, for some trip $t \in \mathcal{T}_i^f$. By constraints (33) it holds for each integral solution that there is either some composition change $q \in \varrho(\delta^+(t))$ such that $Z_{\delta^+(t),q} = 1$ or that there are compositions $p \in \eta(t)$ and $p' \in \mathrm{UN}_{p,t}$ such that $Q_{t,p,p'} = 1$. If this is not satisfied by a fractional solution, there are specifically a set $B \subseteq \varrho(\delta^+(t))$ and a set $E \subseteq \{(p,p') \mid p \in \eta(t), p' \in \mathrm{UN}_{p,t}\}$ such that

$$0 < \sum_{q \in B} Z_{\delta^+(t),q} + \sum_{(p,p') \in E} Q_{t,p,p'} < 1. \tag{52}$$

We then branch by requiring that the above sum is equal to one in one branch and equal to zero in the other. We apply a similar branching decision to enforce that for a trip $t \in \mathcal{T}_o^f$ either a single composition change is chosen that precedes the trip or that a single composition is picked up from the station.

Note that the above branching rules do not necessarily imply that all paths are integral. However, the integrality of the composition variables, composition change variables and the variables related to shunting and turning implies the presence of an integer-valued network flow in the pricing problem graph for each train unit type. It is well known that such a flow can be decomposed into (integral) paths. Moreover, as no costs are directly related to the paths, this implies that we can always find a solution of equal objective value in which all path variables are integral.

6.3 Acceleration Strategies

To speed up the Branch-and-Price procedure, we reuse many of the acceleration strategies as proposed by Haahr et al. (2016). First, we apply delayed row generation for the transition constraints (2), (3), (33) and (34). In this approach, we first solve the restricted master problem without these sets of constraints and only include these constraints dynamically when we find that a constraint is violated in the optimal solution for this smaller problem. The aim of this approach is to reduce the total time that is spent on solving the restricted master problem, as this time spent on the master problem turned out to be substantial in our computational experiments.

A second strategy to reduce the time spent on solving the master problem is to generate multiple paths in each column generation iteration. In particular, we generate a path for each train unit type and for each possible starting point of the train units. Such a starting point either corresponds to a station or to a trip, depending on the location of train units at the start of the planning horizon. By generating multiple paths in each iteration, we generally have to solve the master problem less often.

As preliminary results showed that solving the complete Branch-and-Price procedure for the Delay Path Model was often prohibitive due to the large number of nodes explored in the Branch-and-Bound tree, we also consider a root node heuristic for this model. In this heuristic, we first solve the root node to optimality by column generation. We then solve the master problem with integrality restrictions using only the set of columns that was added in the root node. While this does not necessarily give an optimal solution, this method has been widely applied in literature to obtain high-quality solutions when the available time to solve the model is limited. In our computational experiments, we compare the performance of this method to the full Branch-and-Price procedure.

7 Computational Experiments

In this section, we test the two proposed models on instances of NS. Our aim is to compare the performance of the proposed solution methods on real-life instances and to evaluate the extent to which solving the PDRP reduces the passenger delays.

7.1 The Problem Instances

Two important factors that influence the difficulty of a rolling stock rescheduling instance are the number of trips that are present in the timetable and the train unit types that are available to operate these trips. The former mostly determines the size of the instance, while the latter determines the flexibility that is available in rescheduling the circulation at trips and transitions. To test how the proposed models behave under instances of varying difficulty, we consider two timetables that have been operated by NS in 2016: one with Intercity services for the ICM train unit family and one with Regional (Sprinter) services for the SLT train unit family. Both of these timetables concern a Tuesday, which is the day of the week with the highest passenger demand. Basic information about these timetables and the available rescheduling options, in terms of the possible number of compositions and composition changes, is provided in Table 1.

Table 1: Summary statistics for the two timetables: the number of trips, number of transitions, average number of allowed compositions per trip and the average number of allowed composition changes per transition.

	Intercity (IC)	Regional (RE)
Trips	1122	1386
Transitions	1207	1614
Compositions	19.39	6.87
Composition Changes	26.65	7.40

Another factor that influences the difficulty of an instance is the moment at which the initial delay occurs, as this determines how long the remaining planning horizon is. In particular, if the disruption occurs at time instant τ, we assume that our remaining planning horizon is from $\tau + 30$ to the end of the day, where the 30 minutes of additional time is needed to find an updated schedule and to communicate the updates to the crews. Hence, a disruption that occurs earlier corresponds to a longer planning horizon and thus to a larger problem. To test the effect of different planning horizons, we split up the problem instances further according to whether the delay occurs during the morning peak (6.30-9.00), off-peak (9.00-16.00) or evening peak (16.00-18.30) hours. Resembling their size, these classes of instances will be referred to as large (L), medium (M) and small (S), giving the six scenarios IC_L, IC_M, IC_S, RE_L, RE_M and RE_S.

For each of the six scenarios we create 25 instances by introducing initial delays into the original timetable. Each instance corresponds to an initial delay on one of the trips in the timetable. This initial delay is drawn uniformly at random between 15 and 30 minutes, while the trip on which this initial delay occurs is chosen randomly from all trips that are operated during the relevant moment of the day.

Based on the generation of delays, we assume an upper bound on the delays d_u of 30 minutes. Moreover, we measure all delays in minutes for all instances, implying that $\mathcal{D} = \{0, \ldots, 30\}$. For the delay buffers, we assume that a buffer of 1 minute is available for each trip on top of the nominal trip time. Moreover, we assume a buffer of 3 minutes for all transitions that correspond to turnings. No buffers are assumed for other transitions or for the time between uncoupling and coupling.

Flexible turning for the generated instances is allowed at the two terminal stations for the line that is affected by the initial delay, i.e., those terminal stations where the affected train departed from and will eventually arrive at. Furthermore, we assume that only trips which arrive or depart within four hours of the initial delay can participate in flexible turning. This duration of flexible turning will be varied in Section 7.5 to show the effect of limiting flexible turning to a fixed time interval.

7.2 Objective Function

As is common in rolling stock (re)scheduling, our objective consists of a large number of mutually conflicting objectives which are weighed against each other by their respective coefficients in the objective function. We consider the cost components and associated weights as given in Table 2. *Seat Shortage* indicates the cost associated with not offering a passenger a seat, which is accounted for per kilometer traveled. The next three components correspond to costs related to deviating from the originally communicated shunting plan. Here, we distinguish a new shunting movement from a changed or a canceled one, as these are likely to lead to a different level of disruption to the existing shunting plan. *Ending Inventory Deviation* corresponds to the cost of deviating in the ending inventory from the number of units that were planned to end at a station. This cost is incurred per train unit of deviation. *Mileage* refers to the cost of using a train unit, which is accounted for per kilometer of use and per carriage that is part of the train unit.

Table 2: Coefficients in the objective function.

Cost Component	Coefficient
Seat Shortage	0.2
New Shunting	1000
Changed Shunting	100
Canceled Shunting	50
Ending Inventory Deviation	10000
Mileage	0.1
Delay	10
Flexible Turning	1000

Next to these traditional cost objectives, we include two new cost components for the PDRP. To include the impact of delays on passengers, *Delay* gives the cost per minute of delay for a trip and per passenger that is expected to be on a trip. This ensures that a delay of similar size is preferred on a trip that carries few passengers over a trip that carries many passengers. Moreover, we penalize the use of flexible turning, as changes to the turning pattern of a station are likely to lead to platform changes at a station and may be disruptive for crew schedules as well. Here, *Flexible Turning* gives the cost per used flexible turning, i.e., per turning that is changed compared to the original turning pattern.

7.3 Comparing the Proposed Models

In this section, we compare the performance of the two proposed models. The results of the three solution methods that have been proposed for these models are presented in Table 3, where *Delay Composition Model* refers to the Delay Composition Model solved by a MIP solver, *Delay Path B&P* refers to the Delay Path Model solved by Branch-and-Price and *Delay Path Root Node* refers to the Delay Path Model solved with a root node heuristic. All results have been obtained on a computer with an Intel Xeon E5-2650 v2 @ 2.6GHz processor and 64GB of internal memory. Moreover, the CPLEX 12.8.0 solver has been used to solve the MIP model in the Delay Composition Model and the LP models in the Delay Path Model. Furthermore, a time limit of 15 minutes has been imposed for all solution methods to resemble the limited time that is available in rolling stock rescheduling.

Table 3: Comparison of the three proposed solution methods. All entries are averaged over the 25 instances that have been run for each scenario. Moreover, *Time* denotes the computation time in seconds, *Solved* denotes the number of instances for which the solution method completed before the time limit and *Opt. Gap* denotes the obtained optimality gap (including for instances which were not considered solved).

	Delay Composition Model			Delay Path B&P			Delay Path Root Node		
	Time	Solved	Opt. Gap	Time	Solved	Opt. Gap	Time	Solved	Opt. Gap
IC_L	37	25	0.0%	109	25	0.0%	61	25	0.0%
IC_M	31	25	0.0%	51	25	0.0%	32	25	0.2%
IC_S	22	25	0.0%	15	25	0.0%	10	25	0.3%
RE_L	100	25	0.0%	747	7	3.7%	299	24	0.3%
RE_M	56	25	0.0%	638	12	8.1%	105	25	0.2%
RE_S	41	25	0.0%	27	25	0.0%	10	25	0.4%

Table 3 shows that both the *Delay Composition Model* and *Delay Path Root Node* methods are able to consistently find solutions within the set time limit and are able to complete execution on (almost) all 150 instances. In contrast, *Delay Path B&P* uses significantly more time for some of the instances and is thus able to find an optimal solution for only 119 out of the 150 instances. This effect is especially apparent for the Regional train instances, where only 44 out of 75 instances can be solved to optimality by *Delay Path B&P*. This poorer performance of *Delay Path B&P* on these instances seems to be caused by the large number of flexible turning opportunities for the Regional trains, increasing the number of nodes that have to be explored in the Branch-and-Bound tree.

Another interesting difference between the *Delay Composition Model* and *Delay Path B&P* is the increase of running time for larger instances. While the average running time of the *Delay Composition Model* increases only moderately when going from the S to the L instances, the running time of *Delay Path B&P* increases more quickly. This behavior seems to be caused by the ability of the CPLEX solver to reduce the problem considerably by means of presolving, eliminating many of the delay variables for trips that cannot incur a delay in the optimal solution. Instead, the LP models solved by *Delay Path B&P* model grow rapidly for larger problem instances.

When comparing the models based on these above observations, the *Delay Composition Model* generally performs best when comparing it to the other exact method of *Delay Path B&P*. While it performs comparably to *Delay Path B&P* for the smaller instances, it scales better towards the larger instances. Moreover, we find that *Delay Path Root Node* is able to find solutions that are close to optimality in solution times that are lower than those for *Delay Path B&P*. However, the running times are longer than those of the exact *Delay Composition Model* for larger sized instances. Based on these results, we will use the *Delay Composition Model* throughout the remainder of this section.

7.4 Impact on Delays and Original Circulation

In this section, we look at the extent to which solving the PDRP reduces the passenger delays and at the extent to which it alters the original circulation. We do so by comparing the outcome of the PDRP to a baseline scenario in which no changes are made to the original rolling stock schedule. In the baseline scenario, we thus operate the original circulation and leave the original assignment of individual train units to the trips unaltered. The results of our experiments are given in Table 4 for the case where we consider delays, but where we do not allow flexible turning. Table 5 gives the results when we include flexible turning as well. *Baseline* refers here to the baseline approach, while *Optimal Solution* refers to the optimal solution to the PDRP as found by the *Delay Composition Model*.

Table 4: Results without flexible turning. All entries in the table are averaged over the 25 instances that have been run for each scenario. Moreover, *Objective* denotes the average objective value and *Delay* the average delay cost. The shown percentages indicate the change compared to the baseline scenario.

	Baseline		Optimal Solution			
	Objective	Delay	Objective		Delay	
IC_L	395510	278158	395510	(0.0%)	278158	(0.0%)
IC_M	319280	201929	318445	(−0.3%)	201094	(−0.4%)
IC_S	323889	206538	321136	(−0.9%)	203784	(−1.3%)
RE_L	192568	161472	192439	(−0.1%)	161432	(0.0%)
RE_M	118223	87127	116515	(−1.4%)	85482	(−1.9%)
RE_S	139434	108338	134093	(−3.8%)	102998	(−4.9%)

Table 5: Results with flexible turning. All entries in the table are averaged over the 25 instances that have been run for each scenario. Moreover, *Objective* denotes the average objective value, *Delay* the average delay cost and *Turnings* the average number of flexible turnings. The shown percentages indicate the change compared to the baseline scenario.

	Baseline		Optimal Solution				
	Objective	Delay	Objective		Delay		Turnings
IC_L	395510	278158	370755	(−6.3%)	253244	(−9.0%)	0.16
IC_M	319280	201929	290938	(−8.9%)	172361	(−14.6%)	0.48
IC_S	323889	206538	315585	(−2.6%)	197831	(−4.2%)	0.08
RE_L	192568	161472	182390	(−5.3%)	150156	(−7.0%)	1.04
RE_M	118223	87127	104519	(−11.6%)	72680	(−16.6%)	0.68
RE_S	139434	108338	133190	(−4.5%)	101935	(−5.9%)	0.16

The results in Table 4 and Table 5 show that solving the PDRP substantially reduces the delay costs, with average delay reductions of up to 17% when including flexible turning and average delay reductions of up to 5% when excluding flexible turning. However, we do see that the delay reduction differs per scenario. The delay reduction, for example, tends to be larger for the Regional (RE) instances than for the Intercity (IC) instances. Moreover, when including flexible turning, the delay reduction tends to be smaller for the Evening Peak (S) instances than for the other instances.

The results also show that flexible turning plays a central role in delay reduction. For all scenarios the delay reduction is strictly, and often significantly, larger when including flexible turning. Moreover, the use of flexible turning results in delay reduction for each scenario, while there are scenarios in which no delay reduction is achieved without flexible turning. This large impact of flexible turning can be explained by the relative frequent occurrence of flexible turning opportunities, as trains often reach a terminal station that allows for flexible turning before all delay is absorbed. On the other hand, delay reduction can only be achieved without flexible turning if train units are coupled with a delay at a transition, which turns out to be rare in the baseline scenario.

We also observe that the average number of turnings that are changed is low, where the average number of changed turnings is between 0 and about 1 for all scenarios. This indicates that delay reductions can often be achieved by breaking up only a few existing transitions at terminal stations. This is favorable, as it implies that the effect of flexible turning on the overall shunting and crew plans that are made in later rescheduling phases is limited. This also agrees with the observation that rescheduling leads to significant decreases in the overall objective, which seems to indicate that these delay reductions can be achieved without increasing other objective components too much.

Although the results in Table 4 and Table 5 show the average performance of rescheduling, it should be taken into account that the delays can only be reduced for some of the instances. To take a closer look at the performance of the methods on an instance level, Table 6 shows the percentage of instances for which the delay can be reduced when including flexible turning. Moreover, this table shows the average delay reduction for those instances in which delay reduction is achieved.

Table 6: Results at the instance level for the found optimal solutions. The columns indicate the percentage of instances where the delay cost is reduced (*Reduced*) and the average percentage of reduction in the delay costs for the instances where the delay cost is reduced (*Reduction*).

	Intercity		Regional	
	Reduced	Reduction	Reduced	Reduction
Morning Peak	8%	39%	44%	12%
Between Peak	28%	30%	36%	29%
Evening Peak	8%	44%	16%	29%

The results in Table 6 show that delay reduction is only possible for a relatively small subset of all instances, but that the delay reduction for these instances is generally large. These results confirm our intuition, as there are only a limited number of instances for which inventory propagation occurs and for which it is thus possible to reduce the delay propagation by preventing that a delayed train unit is coupled. Moreover, not every terminal station offers good flexible turning opportunities, as some terminal stations only serve a few incoming and outgoing trains per hour. On the other hand, if a flexible turning or changed shunting movement can be found, the delay is often significantly reduced or passed to trips with fewer passengers.

Table 7 gives further insight into the relation between the use of flexible turning and delay reduction. In particular, it shows for how many of the instances where delay is reduced, flexible turning is used as well.

The results in Table 7 show that flexible turning is indeed used for many of the instances where delay can be reduced. This is in line with the results in Table 4 and Table 5, which showed that flexible turning plays a central role in the reduction of delays. Furthermore, it is interesting to note that flexible turning plays especially a large role for the instances which occur earlier during the day. This can be explained by the fact that the number of terminal stations that is passed by a delayed train unit will generally be lower when the delay occurs later in the day.

Table 7: The percentage of instances for which flexible turning occurs over all the instances for which delay can be reduced.

	Intercity	Regional
Morning Peak (L)	100%	100%
Between Peak (M)	86%	89%
Evening Peak (S)	50%	50%

7.5 The Impact of Flexible Turning Opportunities

An important choice in the PDRP is on which flexible turning opportunities to include in the model. This is determined by the sets $\mathcal{S}_f$, $\mathcal{T}_i^f$ and $\mathcal{T}_o^f$ that define at which stations and at which trips flexible turning can occur. Including more flexible turning possibilities increases the potential of finding rescheduling actions that decrease the delays, but also complicates finding new shunting plans in a later step of the rescheduling process. Moreover, it increases the computation time needed to find an optimal solution to the PDRP. To investigate the impact of the number of flexible turning possibilities on the found results, we vary the duration during which flexible turning is possible. Note that we keep the assumption that flexible turning is only possible at the two terminal stations along the railway line that is affected by the delay. The results of these experiments are given in Figure 6 for the Intercity scenarios, where the flexible turning duration is varied from no flexible turning to 5 hours of flexible turning.

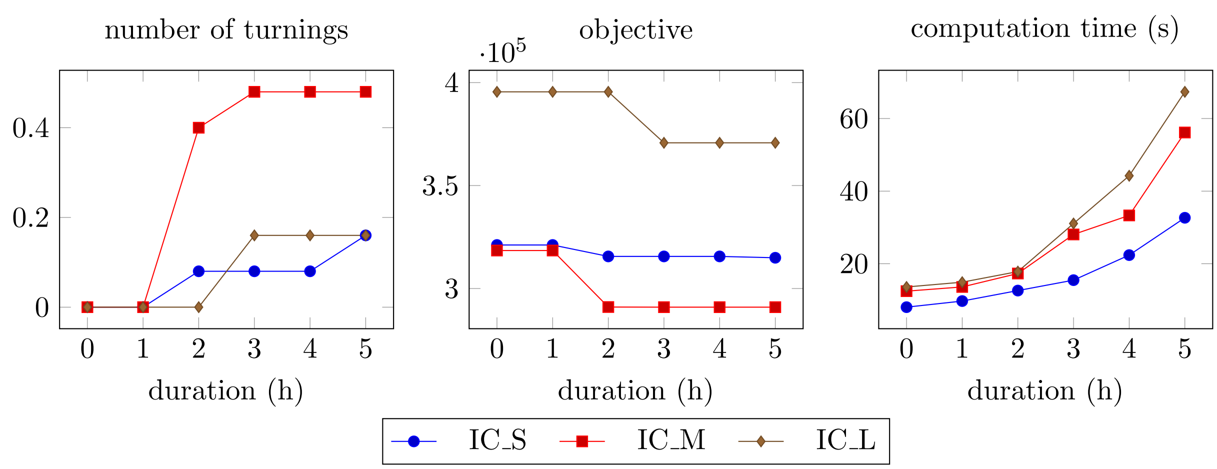


Figure 6: Effect of changing the turning duration on the number of flexible turnings, on the objective value and on the computation time.

The results show that, as expected from the increase in flexible turning possibilities, the number of used flexible turnings increases when the duration of flexible turning rises. Interesting to note is that no flexible turning is used for a duration of flexible turning below an hour. This is caused by the fact that it takes some time before a train arrives at a terminal station where flexible turning is available. Moreover, we see in all scenarios that the increase in the use of flexible turning levels off for large durations, as delays have mostly been absorbed at the moment that these turnings occur.

Furthermore, we observe that the value of the objective function decreases significantly when the flexible turning duration rises. This shows again the role that flexible turning plays in reducing the delays. However, we also see that the computation time steadily rises when the duration increases. In particular, the rate of increase in computation time seems to pick up for larger flexible turning durations. This explains the previous results of Nielsen (2011), who shows that allowing flexible turning for too many stations can quickly make rolling stock rescheduling models very difficult to solve.

8 Conclusions

In this paper, we introduced the Passenger Delay Reduction Problem (PDRP). This problem is concerned with finding a rolling stock circulation that minimizes the passenger delay as a result of some initial delays, while at the same time considering objectives on passenger comfort and operational efficiency. Moreover, it allows to change the turning pattern at a set of predefined stations in order to reduce the passenger delays.

We have introduced two models for solving the PDRP: the Delay Composition Model and the Delay Path Model. These are based on two commonly used models for the traditional Rolling Stock Rescheduling Problem. We show that for practical instances the Delay Composition Model performs best in terms of finding high-quality solutions quickly. This model was able to find optimal solutions for all considered instances within a time frame of 15 minutes. In contrast, solving the full Branch-and-Price model for the Delay Path Model turns out to be too expensive for some instances, especially for those that include many flexible turning opportunities.

Computations on practical instances of NS reveal that the impact of solving the PDRP on delay propagation is substantial, even when combining the costs of delays with traditional objectives for rolling stock rescheduling. In addition, we show that such reductions in terms of total delay can only be found for a select subset of the considered instances. Moreover, an interesting observation from our work is that not much improvement is possible without the use of flexible turning.

Our work provides a first step toward considering the interrelation that exists between the timetable and the rolling stock circulation when rescheduling for delays. An interesting direction for future research is to further integrate these two rescheduling stages. In particular, the problem could benefit from more accurately modeling the delay absorption on trips and at transitions, by considering the delay propagation caused by headway constraints. Moreover, further research may focus on finding efficient (heuristic) solution methods to solve the problem dynamically when new information comes in. These steps would bring the approach closer to the problem that is faced daily by rolling stock dispatchers.

References

Abbink E, van den Berg B, Kroon L, Salomon M, 2004 *Allocation of railway rolling stock for passenger trains. Transportation Science* 38(1):33–41, URL `http://dx.doi.org/10.1287/trsc.1030.0044`.

Adenso-Díaz B, González MO, González-Torre P, 1999 *On-line timetable re-scheduling in regional train services. Transportation Research Part B: Methodological* 33(6):387 – 398, URL `http://dx.doi.org/10.1016/S0191-2615(98)00041-1`.

Alfieri A, Groot R, Kroon L, Schrijver A, 2006 *Efficient circulation of railway rolling stock. Transportation Science* 40(3):378–391, URL http://dx.doi.org/10.1287/trsc.1060.0155.

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH, 1998 *Branch-and-price: Column generation for solving huge integer programs. Operations Research* 46(3):316–329, URL http://dx.doi.org/10.1287/opre.46.3.316.

Ben-Khedher N, Kintanar J, Queille C, Stripling W, 1998 *Schedule optimization at SNCF: From conception to day of departure. Interfaces* 28(1):6–23, URL http://dx.doi.org/10.1287/inte.28.1.6.

Borndörfer R, Reuther M, Schlechte T, Waas K, Weider S, 2016 *Integrated optimization of rolling stock rotations for intercity railways. Transportation Science* 50(3):863–877, URL http://dx.doi.org/10.1287/trsc.2015.0633.

Cacchiani V, Caprara A, Toth P, 2010 *Solving a real-world train-unit assignment problem. Mathematical Programming* 124(1-2):207–231, URL http://dx.doi.org/10.1007/s10107-010-0361-y.

Cacchiani V, Huisman D, Kidd M, Kroon L, Toth P, Veelenturf L, Wagenaar J, 2014 *An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological* 63:15–37, URL http://dx.doi.org/10.1016/j.trb.2014.01.009.

Cadarso L, Marín Á, Maróti G, 2013 *Recovery of disruptions in rapid transit networks. Transportation Research Part E: Logistics and Transportation Review* 53:15 – 33, URL http://dx.doi.org/10.1016/j.tre.2013.01.013.

Dollevoet T, Huisman D, Kroon LG, Veelenturf LP, Wagenaar JC, 2017 *Application of an iterative framework for real-time railway rescheduling. Computers & Operations Research* 78:203–217, URL http://dx.doi.org/10.1016/j.cor.2016.08.011.

Fioole PJ, Kroon L, Maróti G, Schrijver A, 2006 *A rolling stock circulation model for combining and splitting of passenger trains. European Journal of Operational Research* 174(2):1281–1297, URL http://dx.doi.org/10.1016/j.ejor.2005.03.032.

Haahr JT, Wagenaar JC, Veelenturf LP, Kroon LG, 2016 *A comparison of two exact methods for passenger railway rolling stock (re)scheduling. Transportation Research Part E: Logistics and Transportation Review* 91:15–32, URL http://dx.doi.org/10.1016/j.tre.2016.03.019.

Jespersen-Groth J, Potthoff D, Clausen J, Huisman D, Kroon L, Maróti G, Nielsen MN, 2009 *Disruption Management in Passenger Railway Transportation*, 399–421 (Berlin, Heidelberg: Springer Berlin Heidelberg), ISBN 978-3-642-05465-5, URL http://dx.doi.org/10.1007/978-3-642-05465-5_18.

Kennisinstituut voor Mobiliteitsbeleid, 2017 *Mobiliteitsbeeld 2017*. Technical report, URL https://www.rijksoverheid.nl/documenten/rapporten/2017/10/23/kim-publicatie-mobiliteitsbeeld-2017.

Kroon L, Huisman D, 2011 *Algorithmic Support for Railway Disruption Management*, 193–210 (Berlin, Heidelberg: Springer Berlin Heidelberg), ISBN 978-3-642-21192-8, URL http://dx.doi.org/10.1007/978-3-642-21192-8_11.

Lusby RM, Haahr JT, Larsen J, Pisinger D, 2017 *A branch-and-price algorithm for railway rolling stock rescheduling. Transportation Research Part B: Methodological* 99:228–250, URL http://dx.doi.org/10.1016/j.trb.2017.03.003.

Nielsen LK, 2011 *Rolling Stock Rescheduling in Passenger Railways: Applications in short-term planning and in disruption management*. Ph.D. thesis, Erasmus Research Institute of Management, URL http://hdl.handle.net/1765/22444.

Peeters M, Kroon L, 2008 *Circulation of railway rolling stock: a branch-and-price approach. Computers & Operations Research* 35(2):538–556, URL http://dx.doi.org/10.1016/j.cor.2006.03.019.

Schrijver A, 1993 *Minimum circulation of railway stock. CWI Quarterly* 6(3):205–217.

Veelenturf LP, Kidd MP, Cacchiani V, Kroon LG, Toth P, 2016 *A railway timetable rescheduling approach for handling large-scale disruptions. Transportation Science* 50(3):841–862, URL http://dx.doi.org/10.1287/trsc.2015.0618.

Veelenturf LP, Potthoff D, Huisman D, Kroon LG, 2012 *Railway crew rescheduling with retiming. Transportation Research Part C: Emerging Technologies* 20(1):95–110, URL http://dx.doi.org/10.1016/j.trc.2010.09.008.

A Delay Composition Model

$$
\begin{aligned}
&\min f(X, Z, I^{\infty}, Y) && \\
\text{s.t.}\quad & \sum_{p\in\eta(t)} X_{t,p} = 1 && \forall t \in \mathcal{T}, \\
& X_{t,p} = \sum_{q\in\varrho(\delta^{+}(t)):p_{q,t}=p} Z_{\delta^{+}(t),q} && \forall t \in \mathcal{T} \setminus \mathcal{T}_i^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q\in\varrho(\delta^{-}(t)):p'_{q,t}=p} Z_{\delta^{-}(t),q} && \forall t \in \mathcal{T} \setminus \mathcal{T}_o^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q\in\varrho(\delta^{+}(t)):p_{q,t}=p} Z_{\delta^{+}(t),q} + \sum_{p'\in \mathrm{UN}_{p,t}} Q_{t,p,p'} && \forall t \in \mathcal{T}_i^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q\in\varrho(\delta^{-}(t)):p'_{q,t}=p} Z_{\delta^{-}(t),q} + \sum_{p'\in \mathrm{CO}_{p,t}} W_{t,p,p'} && \forall t \in \mathcal{T}_o^f, p \in \eta(t), \\
& C_{c,m} = \sum_{q\in\varrho(c)} \gamma_{q,m} Z_{c,q} && \forall c \in \mathcal{C} \setminus \widetilde{\mathcal{C}}, m \in \mathcal{M}, \\
& U_{c,m} = \sum_{q\in\varrho(c)} \upsilon_{q,m} Z_{c,q} && \forall c \in \mathcal{C} \setminus \widetilde{\mathcal{C}}, m \in \mathcal{M}, \\
& C_{c,m} = \sum_{q\in\varrho(c)} \gamma_{q,m} Z_{c,q} + \sum_{t\in \mathrm{OUT}_c} \sum_{p\in\eta(t)} \sum_{p'\in \mathrm{CO}_{p,t}} (\mu_p^m - \mu_{p'}^m) W_{t,p,p'} && \forall c \in \widetilde{\mathcal{C}}, m \in \mathcal{M}, \\
& U_{c,m} = \sum_{q\in\varrho(c)} \upsilon_{q,m} Z_{c,q} + \sum_{t\in \mathrm{IN}_c} \sum_{p\in\eta(t)} \sum_{p'\in \mathrm{UN}_{p,t}} (\mu_p^m - \mu_{p'}^m) Q_{t,p,p'} && \forall c \in \widetilde{\mathcal{C}}, m \in \mathcal{M}, \\
& Y_t = d_t && \forall t \in \mathcal{T}_{\mathrm{init}}, \\
& Y_t \geq Y_{t'} - \Delta_{\delta^{-}(t)} - \Delta_t && \forall t \in \mathcal{T} \setminus \mathcal{T}_o^f, t' \in \mathrm{IN}_{\delta^{-}(t)}, \\
& Y_t \geq Y_{t'} - \Delta_{\delta^{-}(t)} - \Delta_t - d_u \left(\sum_{p\in\eta(t)} \sum_{p'\in \mathrm{CO}_{p,t}} W_{t,p,p'} \right) && \forall t \in \mathcal{T}_o^f, t' \in \mathrm{IN}_{\delta^{-}(t)}, \\
& \sum_{d\in\mathcal{D}} d D_{c,d}^i \geq Y_t && \forall c \in \mathcal{C}, t \in \mathrm{IN}_c, \\
& \sum_{d\in\mathcal{D}} D_{c,d}^i = 1 && \forall c \in \mathcal{C}, \\
& U_{c,m,d} \leq M_1 D_{c,d}^i && \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D}, \\
& \sum_{d\in\mathcal{D}} U_{c,m,d} = U_{c,m} && \forall c \in \mathcal{C}, m \in \mathcal{M}, \\
& Y_t \geq d D_{\delta^{-}(t),d}^o - \Delta_t && \forall t \in \mathcal{T}, d \in \mathcal{D}, \\
& C_{c,m,d} \leq M_2 D_{c,d}^o && \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D}, \\
& \sum_{d\in\mathcal{D}} C_{c,m,d} = C_{c,m} && \forall c \in \mathcal{C}, m \in \mathcal{M}, \\
& I_{c,m,d} = \iota_{s(c),m}^0 - \sum_{\substack{c'\in\mathcal{C}:\\ s(c')=s(c)}} \sum_{\substack{d'\in\mathcal{D}:\\ \tau^{+}(c')+d'\leq\tau^{+}(c)+d}} C_{c',m,d'} && \\
&\qquad + \sum_{\substack{c'\in\mathcal{C}:\\ s(c')=s(c)}} \sum_{\substack{d'\in\mathcal{D}:\\ \tau^{-}(c')+\left(d'-\Delta_{s(c')}\right)^{+}\leq\tau^{+}(c)+d}} U_{c',m,d'} && \forall c \in \mathcal{C}, m \in \mathcal{M}, d \in \mathcal{D},
\end{aligned}
$$

$$
\begin{aligned}
&I^{\infty}_{s,m} = \iota^0_{s,m} - \sum_{c\in\mathcal{C}:s(c)=s} C_{c,m} + \sum_{c\in\mathcal{C}:s(c)=s} U_{c,m} && \forall s \in \mathcal{S}\setminus\mathcal{S}_f, m\in\mathcal{M},\\
&I^{\infty}_{s,m} = \iota^0_{s,m} - \sum_{c\in\mathcal{C}:s(c)=s} C_{c,m} + \sum_{c\in\mathcal{C}:s(c)=s} U_{c,m} + \sum_{c\in\widetilde{\mathcal{C}}:s(c)=s}\sum_{t\in \mathrm{IN}_c}\sum_{p\in\eta(t)}\sum_{p'\in\mathrm{UN}_{p,t}} \mu^m_{p'} Q_{t,p,p'} && \\
&- \sum_{c\in\widetilde{\mathcal{C}}:s(c)=s}\sum_{t\in \mathrm{OUT}_c}\sum_{p\in\eta(t)}\sum_{p'\in\mathrm{CO}_{p,t}} \mu^m_{p'} W_{t,p,p'} && \forall s\in\mathcal{S}_f, m\in\mathcal{M},\\
&F^i_{t,d,p} \le D^i_{\delta^+(t),d} && \forall t\in\mathcal{T}^f_i, p\in\mathcal{P}, d\in\mathcal{D},\\
&\sum_{d\in\mathcal{D}} F^i_{t,d,p} = \sum_{p':p\in\mathrm{UN}_{p',t}} Q_{t,p',p} && \forall t\in\mathcal{T}^f_i, p\in\mathcal{P},\\
&F^o_{t,d,p} \le D^o_{\delta^-(t),d} && \forall t\in\mathcal{T}^f_o, p\in\mathcal{P}, d\in\mathcal{D},\\
&\sum_{d\in\mathcal{D}} F^o_{t,d,p} = \sum_{p':p\in\mathrm{CO}_{p',t}} W_{t,p',p} && \forall t\in\mathcal{T}^f_o, p\in\mathcal{P},\\
&P_{c,p,d} = \sum_{c'\in\widetilde{\mathcal{C}}:s(c')=s(c)} \sum_{\substack{d'\in\mathcal{D}:\\ \tau_a(c')+\left(d'-\Delta_{c'}\right)^+\le\tau_d(c)+d}} \sum_{t\in\mathrm{IN}_{c'}} F^i_{t,d',p} && \\
&\qquad - \sum_{c'\in\widetilde{\mathcal{C}}:s(c')=s(c)} \sum_{\substack{d'\in\mathcal{D}:\\ \tau_d(c')+d'\le\tau_d(c)+d}} \sum_{t\in\mathrm{OUT}_{c'}} F^o_{t,d',p} && \forall c\in\widetilde{\mathcal{C}}, p\in\mathcal{P}, d\in\mathcal{D},\\
&X_{t,p}\in\{0,1\} && \forall t\in\mathcal{T}, p\in\eta(t),\\
&Z_{c,q}\in\{0,1\} && \forall c\in\mathcal{C}, q\in\varrho(c),\\
&I^{\infty}_{s,m}\in\mathbb{Z}_+ && \forall s\in\mathcal{S}, m\in\mathcal{M},\\
&I_{c,m}, C_{c,m}, U_{c,m}\in\mathbb{Z}_+ && \forall c\in\mathcal{C}, m\in\mathcal{M},\\
&Y_t\in\mathcal{D} && \forall t\in\mathcal{T},\\
&C_{c,m,d}, U_{c,m,d}, I_{c,m,d}\in\mathbb{Z}_+ && \forall c\in\mathcal{C}, m\in\mathcal{M}, d\in\mathcal{D},\\
&D^i_{c,d}, D^o_{c,d}\in\{0,1\} && \forall c\in\mathcal{C}, d\in\mathcal{D},\\
&Q_{t,p,p'}\in\{0,1\} && \forall t\in\mathcal{T}^f_i, p\in\eta(t), p'\in\mathrm{UN}_{p,t},\\
&W_{t,p,p'}\in\{0,1\} && \forall t\in\mathcal{T}^f_o, p\in\eta(t), p'\in\mathrm{CO}_{p,t},\\
&F^i_{t,d,p}\in\{0,1\} && \forall t\in\mathcal{T}^f_i, d\in\mathcal{D}, p\in\mathcal{P},\\
&F^o_{t,d,p}\in\{0,1\} && \forall t\in\mathcal{T}^f_o, d\in\mathcal{D}, p\in\mathcal{P},\\
&P_{c,p,d}\in\mathbb{Z}_+ && c\in\widetilde{\mathcal{C}}, p\in\mathcal{P}, d\in\mathcal{D}.
\end{aligned}
$$

B Delay Path Model

$$
\begin{aligned}
& \min f(X, Z, I^{\infty}, Y) \\
\text{s.t.}\ & \sum_{p \in \eta(t)} X_{t,p} = 1 && \forall t \in \mathcal{T}, \\
& X_{t,p} = \sum_{q \in \varrho(\delta^+(t)) : p_{q,t} = p} Z_{\delta^+(t),q} && \forall t \in \mathcal{T} \setminus \mathcal{T}_i^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q \in \varrho(\delta^-(t)) : p'_{q,t} = p} Z_{\delta^-(t),q} && \forall t \in \mathcal{T} \setminus \mathcal{T}_o^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q \in \varrho(\delta^+(t)) : p_{q,t} = p} Z_{\delta^+(t),q} + \sum_{p' \in \mathrm{UN}_{p,t}} Q_{t,p,p'} && \forall t \in \mathcal{T}_i^f, p \in \eta(t), \\
& X_{t,p} = \sum_{q \in \varrho(\delta^-(t)) : p'_{q,t} = p} Z_{\delta^-(t),q} + \sum_{p' \in \mathrm{CO}_{p,t}} W_{t,p,p'} && \forall t \in \mathcal{T}_o^f, p \in \eta(t), \\
& \sum_{\pi \in \Pi_m} \omega_\pi^t L_\pi = \sum_{p \in \mathcal{P}} \mu_p^m X_{t,p} && \forall m \in \mathcal{M}, t \in \mathcal{T}, \\
& \sum_{\pi \in \Pi_m : b(\pi) = s} L_\pi = \iota_{s,m}^0 && \forall m \in \mathcal{M}, s \in \mathcal{S}, \\
& \sum_{\pi \in \Pi_m : e(\pi) = s} L_\pi = I_{s,m}^{\infty} && \forall m \in \mathcal{M}, s \in \mathcal{S}, \\
& Y_t = d_t && \forall t \in \mathcal{T}_{\mathrm{init}}, \\
& Y_t \geq Y_{t'} - \Delta_{\delta^-(t)} - \Delta_t && \forall t \in \mathcal{T} \setminus \mathcal{T}_o^f, t' \in \mathrm{IN}_{\delta^-(t)}, \\
& Y_t \geq Y_{t'} - \Delta_{\delta^-(t)} - \Delta_t - d_u \left(\sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} W_{t,p,p'} \right) && \forall t \in \mathcal{T}_o^f, t' \in \mathrm{IN}_{\delta^-(t)}, \\
& \sum_{\pi \in \Pi} \zeta_\pi^{t',t} L_\pi \leq M_3 S_{t',t} && \forall t \in \mathcal{T}, t' \in \mathrm{SS}_t, \\
& Y_t \geq Y_{t'} - d_u(1 - S_{t',t}) + \tau^-(\delta^+(t')) - \tau^+(\delta^-(t)) - \Delta_{s(\delta^-(t))} - \Delta_t && \forall t \in \mathcal{T}, t' \in \mathrm{SS}_t, \\
& P_{c,p} = \sum_{\substack{c' \in \widetilde{\mathcal{C}} : s(c') = s(c), \\ \tau_a(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{IN}_{c'}} \sum_{p' : p \in \mathrm{UN}_{p',t}} Q_{t,p',p} \\
& \qquad - \sum_{\substack{c' \in \widetilde{\mathcal{C}} : s(c') = s(c), \\ \tau_d(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{OUT}_{c'}} \sum_{p' : p \in \mathrm{CO}_{p',t}} W_{t,p',p} \\
& \qquad + \sum_{\substack{c' \in \widetilde{\mathcal{C}} : s(c) = s(c') \\ \tau_d(c') \leq \tau_d(c)}} \sum_{t \in \mathrm{OUT}_{c'}} \sum_{\substack{t' \in \mathrm{ST}_t : \\ \tau_a(\delta^+(t')) > \tau_d(c)}} B_{t',t,p} && \forall c \in \widetilde{\mathcal{C}}, p \in \mathcal{P}, \\
& \sum_{\pi \in \Pi_m} \phi_{t,\pi}^a L_\pi = \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{UN}_{p,t}} \mu_{p'}^m Q_{t,p,p'} && \forall t \in \mathcal{T}_i^f, m \in \mathcal{M}, \\
& \sum_{\pi \in \Pi_m} \phi_{t,\pi}^d L_\pi = \sum_{p \in \eta(t)} \sum_{p' \in \mathrm{CO}_{p,t}} \mu_{p'}^m W_{t,p,p'} && \forall t \in \mathcal{T}_o^f, m \in \mathcal{M}, \\
& \sum_{\pi \in \Pi_m} \xi_\pi^{t',t} L_\pi = \sum_{p \in \eta(t')} \sum_{p' \in \mathrm{UN}_{p,t'}} \mu_{p'}^m B_{t',t,p'} && \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, m \in \mathcal{M},
\end{aligned}
$$

$$Y_t \geq Y_{t'} - \Delta_t - M_4 \left(1 - \sum_{p \in \eta(t')} \sum_{p' \in \mathrm{UN}_{p,t'}} B_{t',t,p'}\right) - \Delta(t,t') \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t,$$

$$B_{t',t,p'} \leq \sum_{p^* \in \eta(t')} Q_{t',p^*,p'} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, p \in \eta(t'), p' \in \mathrm{UN}_{p,t'},$$

$$B_{t',t,p'} \leq \sum_{p^* \in \eta(t)} W_{t,p^*,p'} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, p \in \eta(t'), p' \in \mathrm{UN}_{p,t'},$$

$$L_\pi \in \{0,1\} \qquad \forall \pi \in \Pi,$$

$$X_{t,p} \in \{0,1\} \qquad \forall t \in \mathcal{T}, p \in \eta(t),$$

$$Z_{c,q} \in \{0,1\} \qquad \forall c \in \mathcal{C}, q \in \varrho(c),$$

$$I_{s,m}^\infty \in \mathbb{Z}_+ \qquad \forall s \in \mathcal{S}, m \in \mathcal{M},$$

$$Y_t \in \mathcal{D} \qquad \forall t \in \mathcal{T},$$

$$S_{t',t} \in \{0,1\} \qquad \forall t \in \mathcal{T}, t' \in \mathrm{SS}_t,$$

$$Q_{t,p,p'} \in \{0,1\} \qquad \forall t \in \mathcal{T}_i^f, p \in \eta(t), p' \in \mathrm{UN}_{p,t},$$

$$W_{t,p,p'} \in \{0,1\} \qquad \forall t \in \mathcal{T}_o^f, p \in \eta(t), p' \in \mathrm{CO}_{p,t},$$

$$P_{c,p} \in \mathbb{Z}_+ \qquad \forall c \in \widetilde{\mathcal{C}}, p \in \mathcal{P},$$

$$B_{t',t,p'} \in \{0,1\} \qquad \forall t \in \mathcal{T}_o^f, t' \in \mathrm{ST}_t, p \in \eta(t'), p' \in \mathrm{UN}_{p,t'}.$$