CASPT 2018

Exploring the dynamic impact zone for conflict prevention in real-time railway traffic management

Sofie Van Thielen · Francesco Corman · Pieter Vansteenwegen

Abstract In real-time, dispatchers have the challenging task to perform railway traffic management. The goal is to update the timetable in order to reduce delays in case of unexpected events, by means of rescheduling or rerouting some of the trains. In order to support them in making decisions, a Traffic Management System (TMS) is required that is capable of predicting train movements, detecting potential conflicts and deciding on how to prevent these conflicts. A conflict occurs when two trains require the same part of the infrastructure at the same time. However, for dispatchers it is almost impossible to anticipate the impact of their preventing actions on the entire network. This paper therefore examines the dynamic impact zone created and used in a retiming/reordering heuristic. This heuristic is part of a conflict prevention strategy (CPS) aimed at assisting dispatchers by considering the relevant part of the network and the traffic when preventing conflicts. The CPS is extended such that also passenger delays and cancellations of passenger trains can be considered. This CPS is validated on a large study area of the Belgian network. The proposed CPS outperforms FCFS by, on average, 41 % in secondary delays and 33 % in passenger delays while delivering solutions in 2.1 seconds on average and in 12.3 seconds at maximum.

Keywords Conflict prevention · Dispatching · Traffic Management System · Real-time scheduling

Sofie Van Thielen · Pieter Vansteenwegen
KU Leuven, Centre for Industrial Management, Traffic and Infrasture, Celestijnenlaan 300, box 2422, 3001 Leuven
E-mail: Sofie.VanThielen; Pieter.Vansteenwegen@kuleuven.be
Francesco Corman
ETH Zürich, Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland
E-mail: Francesco.Corman@ivt.baug.ethz.ch

1 Introduction

The growing need for sustainable transport entails a growing interest in public transport. Therefore, the quality of train services should be enhanced such that railway companies attract more passengers. This requires a robust timetable and as few delays as possible in real-time. A timetable ensures that two trains should not conflict with one another, so no two trains need the same part of the infrastructure at the same time. However, during daily operations, trains can suffer delays due to, e.g., mechanical failures. Once trains start deviating from their original schedule, conflicts can start appearing.

During the past years, many academic methods have been developed to improve real-time railway traffic management (e.g., Corman and Meng, 2013; Cacchiani et al., 2014; Pellegrini et al., 2016). Recently, some methods have been implemented in practice on limited networks (e.g., Borndörfer et al., 2017). However, a large gap still exists between the state-of-the-art traffic management in academic research and the state-of-the-art in practice. This shows that many challenges arise when putting academic research in practice.

Some infrastructure managers start using advanced Traffic Management Systems (TMS) as a decision support system for their dispatchers (Dolder et al., 2009). Such TMS is only capable of predicting train movements and detecting conflicts. There is no decision support available to prevent or resolve these conflicts. Therefore, a Conflict Prevention Module (CPM), including a Conflict Prevention Strategy (CPS), is required, capable of resolving conflicts such that it can be easily integrated into a current TMS.

This paper proposes a CPS capable of preventing, in a fast way, detected conflicts in large and complex networks. The contributions of this paper are threefold. First, the dynamic impact zone is created more efficiently to make it applicable for larger networks. Second, the very fast Conflict Prevention Strategy is adapted to include passenger numbers. Third, the options for resolving conflicts are extended by also allowing trains to be canceled.

Section 2 starts with explaining some important definitions required for the remainder of the paper. A short literature review is included in Section 3 (for a more extended review, we refer the interested reader to Van Thielen et al., 2018). Section 4 explains the conflict resolution methodology and discusses all extensions included in the CPM. Section 5 shows the results of these extensions on a very large and complex network. The paper is concluded in Section 6.

2 Definitions

In this section, a number of concepts are discussed in order to define the scope of this research. A railway network can be considered either on a microscopic or a macroscopic level. The microscopic level includes every detail concerning switches, tracks, signals, etc. This level is important for both train drivers and dispatchers. The macroscopic level is a simplification of the microscopic level and is often closely related to what passengers experience.

Every signal indicates the beginning or the end of a *block section*. The part of the infrastructure between two subsequent, similarly directed signals thus determines a block section. These two signals fix the size of the block section, which is typically around 1000 meters in Belgium. A *switch* divides a track into two, and enables a train to be guided from one track to another track.

The network can be further decomposed into two types of areas: a station area and a non-station area. A station area is located around a platform area where passengers can debark or embark a train. There are switch areas before and after the platforms such that many possibilities are available for a train when entering or departing the station area. A station area then typically consists of a switch area, a platform area and a switch area. This allows to reroute trains in station areas. All areas that are not station areas are assumed to be non-station areas.

In this research, trains are assumed to drive from signal to signal (and block section to block section). The signal gives information on the next section. A green signal indicates that the next two sections are available. A double yellow signal implies that the following section is available, but the one following that section is occupied by another train. A red signal indicates that the next block section is occupied by another train. In practice a train has to slow down after a double yellow signal and stop at a red signal. In this research trains can only move on if the two following block sections are available.

According to Hansen and Pachl (2008), a block section is exclusively occupied by one train during a time interval. This time interval starts at the *reservation time* and ends at the *release time*. A conflict occurs whenever two blocks overlap in the time-distance diagram. The timetable determines the *route* of a train, i.e. the sequence of block sections.

3 Literature review

Recently, many research has been devoted to the Conflict Detection and Resolution problem (CDR). However, only few approaches have been implemented in practice implying that there is still a large gap between state-of-the-art in practice and in academics (Caimi et al., 2012), especially for dealing with large and complex networks.

Macroscopic approaches have been developed, but these approaches do not deliver a clear and direct resolution to a particular conflict (Meng and Zhou, 2011). Therefore, a feasibility check needs to be included to assure that these resolutions can also be applied to the microscopic network. On the contrary, microscopic approaches deliver resolutions that are on the precision level required for dispatchers. These solutions can be immediately implemented in TMS. In this research a microscopic detail in infrastructure and timetable is used.

There are several approaches found in the literature to solve the CDR problem. The CDR problem can be modeled as a blocking (or no-store), no-wait job shop scheduling problem (Mascis and Pacciarelli, 2002). Some approaches use an alternative graph to model this problem (D'Ariano et al., 2007). A state-of-the-art TMS has been developed in this stream, called ROMA, capable of rescheduling and iteratively rerouting trains in a limited network. Different approaches for rescheduling are included in this TMS (D'Ariano et al., 2008, Corman et al., 2010a, Corman et al., 2010b, Corman et al., 2012, Quaglietta et al., 2013).

Another stream of research solves the CDR problem by creating a mixed integer linear program (MILP) (Törnquist and Persson, 2007; Pellegrini et al., 2016). The academic decision support tool RECIFE uses RECIFE-MILP for solving the CDR problem for a time horizon of 20 minutes (Pellegrini et al., 2016). This optimization model considers all possible routes for all trains, which is very time consuming. Therefore, Samà et al. (2017) introduce a meta-heuristic algorithm to deal with the real-time train routing selection problem.

Borndörfer et al. (2017) summarize some successfully implemented approaches that are or were in operations. For instance, a method based on the approaches presented in Lamorgese and Mannino (2015) and Lamorgese et al. (2016) are currently in operations in several systems in Italy and Latvia.

All the previous mentioned methods have not been proven to be effective on large networks. Therefore, Van Thielen et al. (2018) proposes a CPS that is capable of delivering fast and good resolutions in a real-size network, when a conflict is detected by an advanced TMS. For every detected conflict, a *dynamic impact zone* is constructed and considered when preventing the conflict. This dynamic impact zone includes all trains and conflicts that are potentially affected by the resolution of the detected conflict. This paper proposes and evaluates several extensions to the dynamic impact zone that should assure the applicability on larger networks. Additionally, the CPS is extended such that passenger numbers can be considered and trains can be canceled.

4 Conflict resolution methodology

When a conflict is detected by TMS, it will be sent to the Conflict Prevention Module. If the conflict is located in the station area, the rerouting optimization is started to try to minimize the secondary delays by giving alternative routes to (some of the) trains. As soon as an optimal solution is found, this is implemented. Otherwise, this optimization is limited by a maximum computation time. When this time limit is reached, the best found feasible solution is implemented. If the conflict is located in a non-station area or if the conflict could not be resolved by rerouting, the retiming heuristic is started. This retiming heuristic starts by creating a dynamic impact zone including all other conflicts and trains that might be affected by the resolution of the conflict. Then, a simplified progress examination is started with only the trains present in the dynamic impact zone. This progress examination is executed for every possible conflict resolution. During this examination, the total delay of all considered trains is summed up. Then, the resolution attaining the least total delay is chosen and returned by the CPM.

In the next subsection, we will discuss the update in the rerouting optimization for dealing with passenger numbers and connections between trains. Then, the creation of the dynamic impact zone is looked upon in more detail, and adapted to keep the computation time low when enlarging the network. Subsequently, the adaption to including passenger numbers in the DIZ heuristic is discussed. Lastly, the option of canceling trains is added to the DIZ heuristic.

4.1 Update of the rerouting optimization

The rerouting optimization is based on a flexible job shop problem and is built in Cplex. During the optimization, only changes inside the station area are considered, not in the rest of the network. This means that the lines on which all trains will enter and leave the station area and also the current position of the different trains already in the station area, are fixed. The rerouting starts from the situation in the station area, one minute after the moment of detecting the conflict (the control delay), whereas the end time is the latest departure time of one of the conflicting trains from the station area. The rerouting optimization considers all trains passing this station within this time period. As long as a train has not started its route in the station area, its planned route can be changed.

The complete mathematical model used for rerouting is available in the appendix, together with an explanation of all parameters and variables. It is discussed in detail in our previous paper (Van Thielen et al., 2018). This section only describes and motivates the changes made to the previous mathematical model. First an alternative objective function is discussed which is used when the objective of the CPM is changed to minimizing passenger delays. Let PD_t be the passenger delay of train t. The passenger delay is measured as the delay of a train times the number of passengers affected. The number of passengers affected is determined by taking the average number of passengers right before and after the stop at the current station, which is denoted by PN_t. The objective function then becomes

$$\text{minimize} \quad \sum_{t=1}^{|T|} \Big(PD_t - \frac{x_{t,0}}{100} \Big), \tag{1}$$

with $PD_t = PN_t \times DT_{t,\text{nbBS}_r} - \Big(\delta_t + mr_{t,1} + \sum_{b=1}^{\text{nbBS}_r} mt_{t,b} \Big)$.

The second term of the objective function, with $x_{t,0}$, assures that the original route from the train is preferred if the passenger delay cannot be decreased. The rerouting optimization is also extended to take *connections* between trains into account. Three different types of connections are considered: coupling, decoupling and making a turnaround. Dependent on the type of connection, some additional constraints need to be imposed on the mathematical model.

If trains couple at a station platform, both trains should be allowed on the platform and they will departure from the platform as one train. Denote this train by t_k. The coupled train t_k cannot depart before the two trains t_i and t_j, that are coupled, both have arrived at the platform plus an additional service time ma_{t_i,t_j}. The block section including the platform is denoted by b_p. The block sections before and after the platform are denoted by $b_p - 1$ and $b_p + 1$ respectively.

$$DT_{t_k,b_p} \geq DT_{t_i,b_p-1} + ma_{t_i,t_k}, \forall\, b_p \in B_p,\ t_i, t_k \in T,\ \big(t_i, t_k, \text{`coupling'}\big) \in A.$$
$$DT_{t_k,b_p} \geq DT_{t_j,b_p-1} + ma_{t_j,t_k}, \forall\, b_p \in B_p,\ t_j, t_k \in T,\ \big(t_j, t_k, \text{`coupling'}\big) \in A,$$

with DT_{t_k,b_p} the departure time of train t_k from block section b_p and A the set of all connections.

If a train t_i decouples at a station platform, two trains t_j and t_k will depart from the station. The order in which trains depart is fixed based on how the trains are physically coupled. Assume that train t_j departs first from the platform. In that case, the additional constraints are:

$$DT_{t_j,b_p} \geq DT_{t_i,b_p-1} + ma_{t_i,t_j}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`decoupling'}\big) \in A.$$
$$DT_{t_k,b_p} > DT_{t_j,b_p}, \forall\, b_p \in B_p,\ t_j, t_k \in T,\ \big(t_i, t_k, \text{`decoupling'}\big) \in A.$$

In case a train t_i has a turnaround at its terminal station, it is necessary that t_i remains occupying the platform until the connected train t_j takes over. Therefore train t_j should start reserving the block section immediately when t_i releases the block section.

$$AR_{t_i,b_p} = SR_{t_j,b_p}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`turnaround'}\big) \in A,$$
$$DT_{t_j,b_p} \geq SR_{t_j,b_p} + mt_{t_j,b_p} + ma_{t_i,t_j}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`turnaround'}\big) \in A,$$

with AR_{t_i,b_p} the release time of train t_i on block section b_p, SR_{t_j,b_p} the reservation time of train t_j on the same block section and mt_{t_j,b_p} the minimum travel time of train t_j for block section b_p.

4.2 Creation of the *dynamic impact zone*

If the rerouting optimization does not resolve the conflict or the conflict is not located in a station area, then a resolution based on retiming and/or reordering a train needs to be proposed. The retiming heuristic introduced in Van Thielen et al. (2018) is based on a *dynamic impact zone*, and therefore called the Dynamic Impact Zone (DIZ) heuristic. This dynamic impact zone is based on offline calculations determining which conflicts are most likely to take place and on a set of *potentially conflicting trains*, which normally use the same infrastructure as the conflicting trains within a time range of 15 minutes.

The *initial conflict* is the conflict that is detected by TMS and sent to the CPM. A *first order conflict* is a conflict that contains (at least) one of the trains in the initial conflict. In general, a nth order conflict is a conflict that contains (at least) one of the trains in a $(n-1)$th order conflict, but it is not a $(n-1)$th order conflict itself. Figure 1 shows an example network with an initial conflict, first, second and third order conflicts.

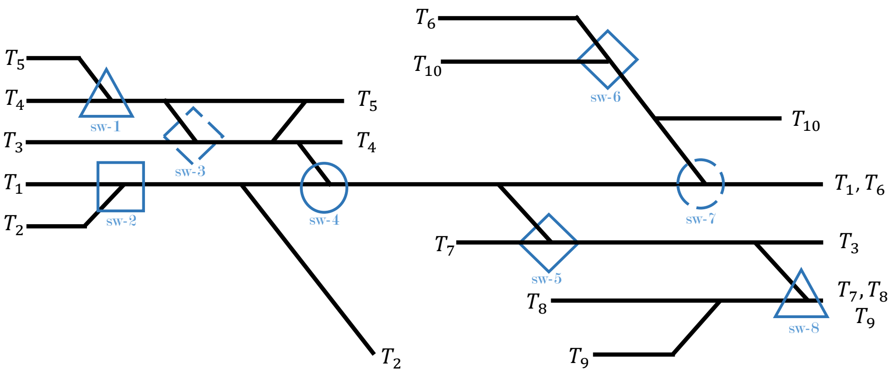

Fig. 1 Small network with different conflicts indicated by a geometric shape. The square indicates the initial conflict. A circle marks a first order conflict, a diamond a second order conflict and a triangle a third order conflict. Full line shapes indicate the *most likely conflicts*. The names of the relevant switches are indicated below the shape.

The first train in the conflict is the train requiring the block section where the conflict takes place, first. The second train is then the train that wants to occupy the same block section second. A conflict can be resolved in two ways: delaying the first train or delaying the second train. The proposed DIZ heuristic examines both resolutions to the conflict by looking at a simplified progress of the next half hour during which all conflicts are solved by FCFS. Naturally, this examination should be limited in time and space because the computation time is restricted, for usage in practice. Therefore, only relevant parts of the network/traffic that are assumed to be impacted by the solution of the current conflict should be taken into account. These relevant parts are captured in the dynamic impact zone. The dynamic impact zone includes all potentially conflicting trains in first order conflicts and second order most likely conflicts. The remainder of this section focuses on improvements of this standard DIZ heuristic.

4.3 Limiting the dynamic impact zone

The dynamic impact zone includes all first order conflicts and second order most likely conflicts. However, previous research (see Van Thielen et al., 2018) showed that only about 10 % of all conflicts present in the dynamic impact zone are actually detected as a new conflict during the examination progress. Moreover, the computation time of the heuristic is mostly related to the number of conflicts included in the dynamic impact zone and much less related to resolving these conflicts (with FCFS). Therefore, it is important to try to reduce the number of conflicts considered to only those that are likely to happen with the current delays of the trains. So we try to limit further the conflicts and trains considered in the dynamic impact zone to those actually affected by the resolution of the initial conflict. This subsection includes two measures taken to further limit the number of conflicts considered in the dynamic impact zone.

4.3.1 Updating the potentially conflicting trains

The dynamic impact zone is based on the offline-determined set of potentially conflicting trains, which are, on average, frequently conflicting. However, at a particular time, some potentially conflicting trains can still be on schedule or significantly delayed, and therefore not be potentially conflicting. Therefore, their current timing, including all known delays up to the time instant of detection, are considered when determining whether the offline-determined potentially conflicting trains should still be considered as such. This feature is illustrated with an example.

Example

Figure 1 shows a small network where ten trains are driving from left to right. The square indicates the initial conflict detected by TMS. All other conflicts are *potential conflicts*, i.e. conflicts between potentially conflicting trains. This implies that *on average* these conflicts take place frequently, but also that these conflicts might not take place in real-time when a certain initial conflict is under evaluation. The corresponding blocking times of all trains on the switches indicated are stated in Table 1. These blocking times include all delays known up to the current time instant, i.e. the time instant that the initial conflict is detected.

The initial conflict is detected at 06:55:00. All potential conflicts are situated in the future. Looking at the current situation, only one conflict in addition to the initial conflict will actually take place in the future: T_1 and T_3 will conflict on sw_4.

Previous research did not include a real-time update of the potentially conflicting trains. The limitation was made by the time limit of the progress examination and the order of the conflicts. When looking at Figure 1, all conflicts with a circle and diamond with a full line were considered in the dynamic impact zone previously. Trains T_1, T_2, T_3, T_6, T_7 and T_{10} were taken into account during the progress examination. However, due to previous delays it is possible that some trains are not potentially conflicting anymore at the current time instant. Including less trains in the progress examination will reduce the computation time for the DIZ heuristic. Therefore, the information is now updated and it is checked which trains still have a potential conflict within a time range of γ minutes. Assume that $\gamma = 5$. First, the first order conflicts are looked upon. Clearly, the conflict between T_1 and T_6 is not a potential conflict anymore. Therefore, T_6 will not be included in the dynamic impact zone when assuming $\gamma = 5$ minutes. Also, the conflict between T_3 and T_7 is not considered. The set of trains in the dynamic impact zone is now limited to T_1, T_2 and T_3. So, we invest a little more computation time in updating the set of potential conflicts and this should save us more computation time when considering the conflicts during the progress examination and lead to (at least) the same quality of results.

T_a	T_b	switch	Start-end of blocking time T_a	Start-end of blocking time T_b
T_1	T_2	sw_2	07:01:00 - 07:02:30	07:00:30 - 07:01:15
T_1	T_3	sw_4	07:04:00 - 07:05:00	07:04:30 - 07:05:45
T_1	T_6	sw_7	07:10:30 - 07:12:00	07:20:30 - 07:22:15
T_3	T_4	sw_3	07:02:15 - 07:03:45	07:07:45 - 07:08:45
T_3	T_7	sw_5	07:15:00 - 07:16:45	07:08:00 - 07:09:45
T_4	T_5	sw_1	07:06:15 - 07:08:00	07:04:00 - 07:05:00
T_6	T_{10}	sw_6	07:15:45 - 07:17:45	07:01:15 - 07:03:00
T_7	T_8	sw_8	07:13:45 - 07:15:00	07:21:45 - 07:23:00

Table 1 Blocking times on the switches where a potential conflict is situated on the example network in Figure 1.

4.3.2 Adding a maximum distance from the initial conflict

The dynamic impact zone was only limited by the time in the progress examination. When enlarging the study area further, this limitation could be insufficient to keep the computation time low. Therefore, the dynamic impact zone is also limited in distance from the initial conflict. This extension restricts the search further by only considering the next ϵ km of the initial conflicting trains, starting at the location of the initial conflict.

Figure 2 shows the small network from Section 4.3.1. There is an initial conflict between T_1 and T_2. All conflicts in the corresponding dynamic impact zone within $\epsilon = 25$ km are indicated in red. Compared to the standard DIZ heuristic, there are less trains and conflicts considered in the new dynamic impact zone. With $\epsilon = 25$ km, only trains T_1, T_2, T_3 and T_7 are taken into account during the progress examination of the heuristic. In general, this should speed up the calculations without reducing the quality of the solution too much.

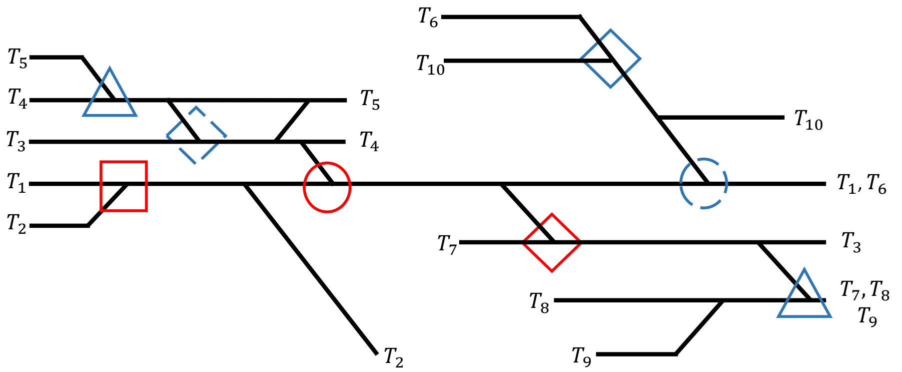

Fig. 2 Small network with different conflicts indicated by a geometric shape. Only the first and second order conflicts within a distance of $\epsilon = 25$ km are considered in the dynamic impact zone (indicated in red).

4.4 Passenger vs secondary delays

Operations managers are typically interested most in minimizing the secondary delays, whereas train operating companies might want to minimize the passenger delays. Previously, the CPM only considered secondary delays. Therefore the CPM is adapted in such a way that it can consider passenger delays instead of secondary train delays.

Instead of the secondary delay, the passenger delay is considered. The passenger delay equals the amount of delay a train is given due to a conflict times the number of passengers affected at the current block section. Evidently, freight trains do not receive any passenger delay. When examining the progress, the total passenger delay is calculated and used to determine the best conflict resolution.

4.5 Canceling passenger trains

If delays become longer, it can be beneficial to cancel a train because the train will cause too many more conflicts in the remainder of its route. Therefore, the DIZ heuristic is extended such that (at maximum) four possibilities are regarded: delay the first train, delay the second train, cancel the first train, cancel the second train. In this paper, only passenger trains are allowed to be canceled. Trains can only be canceled at a station platform. Therefore, a train first needs to be delayed to resolve the conflict if the conflict is not taking place at a station platform. The train can then be canceled at the next station where all passengers can debark, and the train can be put on a side track.

Whenever a train is canceled, its (remaining) connections need to be checked. In case of decoupling, the two, new trains are also canceled because the first train was canceled. In case of a turnaround at a terminal station, it is examined whether the second train should pass the station where the first train was canceled. If this is the case, then the second train is only canceled up to this station. The second train is then assumed to depart again, at its scheduled time in this station, if this is feasible.

Canceling a train certainly causes annoyance to passengers, and is a measure that is considered worse than delaying a train. Therefore, canceling a train should correspond to a certain amount of delay, denoted by δ minutes. In Section 5, the effect of altering this parameter from 15 to 30 minutes is examined.

5 Case Study

The CPS is extensively tested by the use of simulation. This simulation consists of three main compoments that model a complete TMS interacting with real life. The three components are a simulator module, a predictor module and a CPM. Figure 3 shows how these components interact. The discrete event simulation resembles the real-time driving of all trains. The prediction and conflict detection resembles an advanced TMS. The CPM includes the rerouting optimization and the DIZ heuristic. For more information on this closed loop setup, we refer the interested reader to Van Thielen et al. (2018).

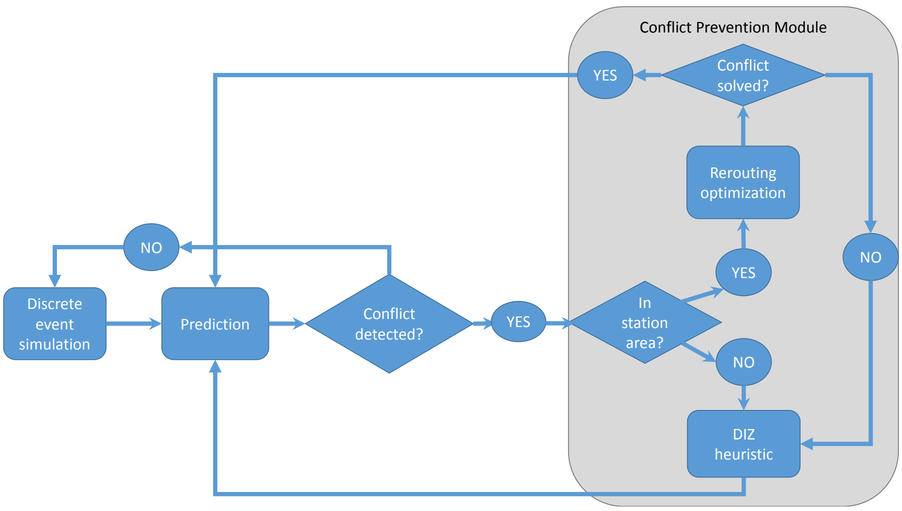

Fig. 3 A flowchart representing the three components within the simulation and their interaction.

The CPM with new extensions is tested on a large and complex study area that includes the provinces of East and West Flanders in Belgium. This microscopic network contains 130 stations and 11766 block sections. Both freight and passenger trains are taken from the microscopic timetable of 17/03/2017. The simulation horizon is set from 7 a.m. to 8 a.m. During this time horizon and in this study area, there are 240 trains considered. All 51 connections taking place between these trains in practice, i.e. (de)coupling and turnarounds, are taken into account.

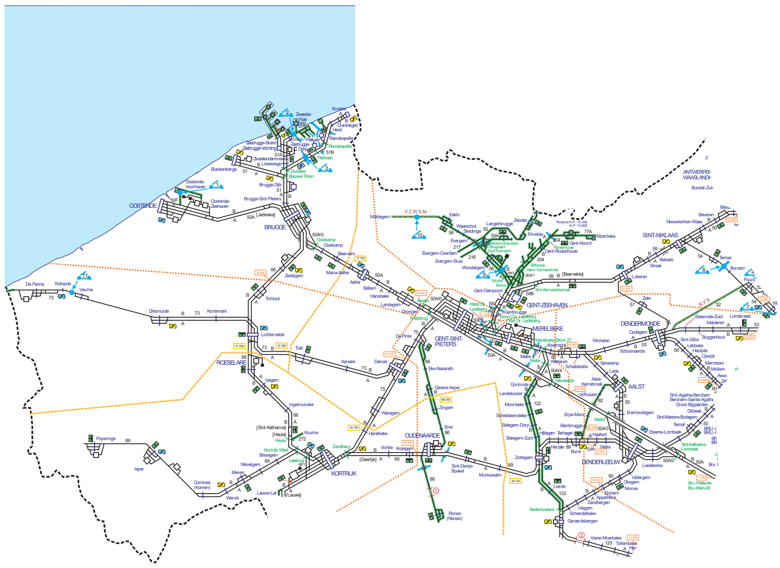

Fig. 4 Study area: the provinces of West and East Flanders in Belgium.

As a standard, the prediction horizon is set to 5 minutes, the heuristic horizon to 30 minutes and the control delay to 60 seconds (see Van Thielen et al. (2018) for more information). The offline calculations are based on 350 runs from a delay scenario with α % of the trains delayed. The value α is randomly taken from the interval $[20\%, 80\%]$. The computation time of the rerouting optimization is limited to 10 seconds. In order to evaluate the CPM, different runs from a delay scenario α are taken. In one run, approximately α% of all trains, which are randomly chosen, are given a random delay from an exponential distribution with an average of 3 minutes and a maximum of 15 minutes. The CPM is evaluated based on the total secondary delay, the total passenger delay and the computation time of the CPM. The total secondary delay sums up the secondary delays of all trains at the end of the simulation horizon or when leaving the study area. The total passenger delay sums up the multiplication of the delay and the passengers affected at the block section where the delay is given. The computation time of the CPS for a single initial conflict includes the creation and execution of the rerouting optimization, the creation of the *dynamic impact zone* and the execution of the DIZ heuristic itself.

Due to a lack of realistic data, simple assumptions are made on the passenger flows. Trains can either be busy or quiet. A busy train has approximately four times the number of passengers aboard compared to a quiet train. The passengers function is a stepwise function that changes randomly in every station. It is assumed that the number of passengers increases until the middle point of the trains full route. After the middle point of its route, the number of passengers decreases. Figure 5 depicts the passenger pattern of a quiet train.

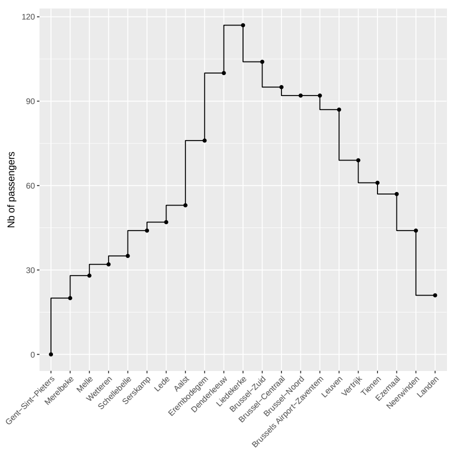

Fig. 5 The passenger pattern of a quiet train for its full route.

5.1 Results

The CPM is tested and compared to some basic dispatching strategies. The most basic dispatching strategy is First Come, First Served (FCFS), where the train wanting to reserve the block section where the conflict occurs first, is allowed to drive first. This strategy resembles an inexperienced dispatcher. As a second comparison, the first order heuristic is considered. This heuristic is also based on a progress examination during the next half hour, but the impact zone is limited to first order conflicts. This strategy thus makes no use of offline calculations or any extension discussed in Section 4. When returning to the example network in Figure 1, only trains T_1, T_2, T_3 and T_6 are considered during the progress examination. Table 2 shows the results based on five runs from random delay scenarios. This number of runs is limited, for now, due to the long simulation times required when performing experiments. It should be noted that the objective function used in CPS and the first order heuristic is secondary train delays and that rerouting is also included in the CPS. Clearly, our CPS outperforms FCFS by on average 41 % based on secondary delays and 17 % on passenger delays. CPS also outperforms the first order heuristic by on average 9.3 %.

Strategy	SD (in min)	PD (in passenger min)
FCFS	584	20141
First order + rerouting	376	18642
CPS	341	16681

Table 2 The average secondary and passenger delays of CPS compared to basic dispatching strategies.

Table 3 shows the computation time of the different strategies. Clearly, both first order and our CPS are slower than the FCFS strategy, though delivering better solutions based on both secondary and passenger delays.However, in real-time, the computation time of our CPS is low. And with a little extra computation time, significantly higher quality solutions are obtained.

Strategy	Average computation time (in s)	Maximum computation time (in s)
FCFS	0.02	0.2
First order + rerouting	1.5	11.3
CPS	1.7	11.7

Table 3 The average and maximum computation time of CPS compared to basic dispatching strategies.

5.1.1 Canceling trains

Allowing trains to be canceled can decrease the overall delays. Canceling a train equals delaying a train with δ minutes. The results are shown in Table 4. The average secondary delay over all trains can clearly be decreased when considering the additional option of canceling trains. However, when the value of canceling a train equals 30 minutes, not very often a train is canceled, limiting the inconvenience for passengers. Decreasing the value δ to 15 minutes further decreases the secondary delays.

For the remainder of this paper, it is assumed that canceling a train corresponds to giving a delay of 30 minutes. In terms of passenger delays, canceling a train is then equal to 30 minutes times the number of passengers affected by the cancellation.

Strategy	SD (in min)	# of canceled trains
CPS - no canceling	341	0
CPS - $\delta = 30$ min	335	0.8
CPS - $\delta = 15$ min	312	2.8

Table 4 Comparison of the effects of the parameter δ on the overall delays.

Including the option of canceling a train, implies including (up to) two more options per heuristic calculation. Table 5 shows the effect on the average and maximum computation time. Clearly, including more options tends to increase the computation time. However, the computation time still remains low.

Strategy	Average computation time (in s)	Maximum computation time (in s)
CPS - no canceling	1.7	11.7
CPS - $\delta = 30$ min	2.1	12.3
CPS - $\delta = 15$ min	2.2	12.2

Table 5 Comparison of the effects of the parameter δ on the computation time.

5.1.2 Passenger delays

Taking passenger numbers into account can lead to different results than looking at the secondary delays. Therefore, the objective function of the CPM is altered to deal with passenger or secondary delays. Table 6 shows the total passenger and secondary delays when changing the objective function in the CPM. Though the results are based on simplified passenger assumptions, they show that taking passengers into account can deliver a completely different result. The secondary delays are clearly lower in case of a secondary delay objective, but with regards to the passengers it performs worse. The passenger objective also results in more canceled trains, mostly when trains do not have a lot of passengers on board.

Strategy	SD (in min)	PD (in passenger min)	# of canceled trains
CPS - secondary	335	15001	8
CPS - passenger	591	13561	15

Table 6 Comparison of changing the objective function in the CPS.

5.1.3 Effects on the DIZ

In this section, the effect of the measures presented in Section 4.3 is analyzed. Table 7 shows the effect of updating the information during the creation of the DIZ. The computation time decreases slightly when an update is included. This decrease is limited due to the fact that including the update also requires some computation time.

Strategy	SD (in min)	Average computation time (in s)	Maximum computation time (in s)
CPS - no update	341	2.3	14.3
CPS - $\gamma = 10$ min	335	2.1	12.3

Table 7 Comparison of including updated information for creating the *dynamic impact zone* based on secondary delays and computation time.

The limitation in distance is examined by comparing values of $\epsilon = 10, 50$ and 100 km. Table 8 shows the effects on the secondary delays and the computation time. Clearly, the computation time is lowest for the lowest value of ϵ. However, the secondary delays decrease when the distance is enlarged. Based on these first experiments, the value of 50 km seems a good trade-off between the quality of the results and the required computation time.

Strategy	SD (in min)	Average computation time (in s)	Maximum computation time (in s)
CPS - $\epsilon = 10$ km	373	1.6	11.8
CPS - $\epsilon = 50$ km	343	2.0	12.4
CPS - $\epsilon = 100$ km	341	2.3	14.3

Table 8 Comparing the effects of using different distances from the initial conflict when creating the *dynamic impact zone* based on secondary delays and computation time.

6 Conclusion

This paper proposes an improved Conflict Prevention Strategy (CPS) capable of resolving conflicts in real-time for large and complex networks. Our CPS consists of a rerouting optimization for station areas and a retiming heuristic, called the Dynamic Impact Zone (DIZ) heuristic. This heuristic can choose to either delay or cancel trains. Its objective function can easily be altered to the necessary requirements. In this paper a comparison is made between an objective based on secondary train delays and on passenger delays.

The DIZ heuristic is extended by a speed improvement in the implementation and several other features. Passengers are taken into account when making decisions. Passenger trains are allowed to be canceled at a certain cost in terms of delay. Also, the creation of the *dynamic impact zone* is examined. These extensions assure that the computation time of the DIZ heuristic remains low such that the heuristic is applicable to a large and complex network.

In order to evaluate the DIZ heuristic, the CPS is tested on a large part of the Belgian network. Results show that our CPS can outperform FCFS by 41 % when considering secondary delays and by 33 % when considering passenger delays. The computation times remain low on this large network. Conflict resolutions are delivered on average within 2.1 seconds and in a maximum time of 12.3 seconds. Without considering the additional option of canceling trains, resolutions are delivered within 2.3 seconds.

References

Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M. and Schlechte, T, Recent Succes Stories on integrated optimization of railway systems. Transportation Research Part C, 74, 196–211 (2017)

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L. and Wagenaar, J, An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B, 63, 15–37 (2014)

Caimi, G., Fuchsberger, M., Laumanns, M. and Lüthi, M, A model predictive control approach for discrete-time rescheduling in complex central railway station areas. Computers & Operations Research, 39, 2578–2593 (2012)

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M., A tabu search algorithm for rerouting trains during rail operations. Transportation Research Part B, 44, 175–192 (2010a)

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M., Centralized versus distributed systems to reschedule trains in two dispatching areas. Public Transport, 2 (3), 219–247 (2010b)

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M., Optimal inter-area coordination of train rescheduling decisions. Transportation Research Part E, 48, 71–88 (2012)

Corman, F. and Meng, L. A review of online dynamic models and algorithms for railway traffic control. IEEE International Conference on Intelligent Rail Transportation (ICIRT), 2013, Beijing, 128–133 (2013)

Corman, F. and Quaglietta, E., Closing the loop in railway traffic control: framework design and impacts on operations. Transportation Research Part C, 54, 15–39 (2015)

Corman, F., D'Ariano, A., Marra, A., Pacciarelli, D. and Samà, M, Integrating train scheduling and delay management in real-time railway traffic management. Transportation Research Part E, 105, 213–239 (2017)

D'Ariano, A., Pranzo, M. and Hansen, I.A., Conflict resolution and train speed coordination for solving real-time timetable perturbations. IEEE Transactions on Intelligent Transportation Systems, 8(2), 208–222 (2007)

D'Ariano, A., Corman, F., Pacciarelli, D. and Pranzo, M., Reordering and local rerouting strategies to manage train traffic in real time. Transportation Science, 42 (4), 405–419 (2008)

Dolder, U., Krista, M. and Voelcker, M. RCS - rail control system, realtime train run simulation and conflict detection on a netwide scale based on updated train positions. In Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis, RailZurich (2009)

Hansen, I. and Pachl, J, Railway timetable and traffic: analysis, modelling, simulation. Eurail Press (2008)

Lamorgese, L. and Mannino, C., An exact decomposition approach for the real-time train dispatching problem. Operations Research, 63 (1), 48–64 (2015)

Lamorgese, L., Mannino, C. and Piacentini, M., Optimal train dispatching by Benders'-like reformulation. Transportation Science, 50(3), 910–925 (2016)

Mascis, A. and Pacciarelli, D., Job-shop scheduling with blocking and no-wait constraints. European Journal of Operations Research, 143, 598–517 (2002)

Meng, L. and Zhou, X., Robust single-track dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. Transportation Research Part B, 45, 1080–1102 (2011)

Pellegrini, P., Marlière, G. and Rodriguez, J., A detailed analysis of the actual impact of real-time railway traffic management optimization. Journal of Rail Transport Planning & Management, 6, 13–31 (2016)

Quaglietta, E., Corman, F. and Goverde, R.M.P., Stability analysis of railway dispatching plans in a stochastic and dynamic environment. Journal of Rail Transport Planning & Management, 3, 137–149 (2013)

Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giaroli, M. and Nicholson, G., The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardized data communication architecture. Transportation Research Part C, 63, 23–50 (2016)

Samà, M., D'Ariano, A., Corman, F. and Pacciarelli, D., A variable neighbourhood search for fast train scheduling and routing during disturbed railway traffic situations. Computers & Operations Research, 78, 480–499 (2017)

Törnquist, J. and Persson, J.A., N-tracked railway traffic re-scheduling during disturbances. Transportation Research Part B, 41, 342–362 (2007)

Van Thielen, S., Corman, F. and Vansteenwegen, P., Considering a dynamic impact zone for real-time railway traffic management. Transportation Research Part B, 111, 39–59 (2018)

A Appendix

This appendix describes shortly the full mathematical model of the rerouting optimization. A previous version of the model is extensively described in Van Thielen et al. (2018).

The parameters of the model describe the real-time situation of all trains in the station areas.

$$T = \{t_1, \ldots, t_{|T|}\}, \text{ set of trains with } |T| \text{ the number of trains.}$$
$$R = \{r_1, \ldots, r_{|R|}\}, \text{ set of routes with } |R| \text{ the number of routes.}$$
$$R_t \subset R = \text{set of routes train } t \text{ can take.}$$
$$C = \{\big((b_i, t_i, r_i), (b_j, t_j, r_j)\big) \mid b_i \text{ and } b_j \text{ are block sections, that share a part of the infrastructure or that are identical, with } b_i \text{ belonging to route } r_i \text{ from train } t_i \text{ and } b_j \text{ to route } r_j \text{ from train } t_j\}$$
$$A = \{\big(t_i, t_j, y\big) \mid t_i \text{ and } t_j \text{ have an association of type } y \text{ in this station area}\}$$
$$B_p = \{b_{p_1}, \ldots, b_{p_n}\}, \text{ set of all block sections having a platform, with } n \text{ the number of platforms in the station area.}$$
$$\delta_t = \text{earliest time instant at which train } t \text{ can start reserving the first block section in the station area.}$$
$$mr_{t,b} = \text{minimum reservation time required for train } t \text{ before entering block section } b$$
$$mt_{t,b} = \text{minimum travel time required for train } t \text{ to travel through block section } b$$
$$mc_{t,b} = \text{minimum clearing time required for train } t \text{ to leave block section } b$$
$$ma_{t_i,t_j} = \text{minimum association time between trains } t_i \text{ and } t_j$$
$$nbBS_r = \text{total amount of block sections in route } r \text{ of this station area.}$$

The subset R_t includes routes that start and end at the predetermined signals of train t when, respectively, entering and leaving the station area. Note that δ_t is the start reservation time instant determined by the timetable and the imposed delay on train t at the current time. This model also uses block section 0, indicating the block section just before the station area. The sum of $mr_{t,b}$, $mt_{t,b}$ and $mc_{t,b}$ equals a block in the blocking time theory.

The decision variables are now introduced.

$$x_{t,r} = \begin{cases} 1 & \text{if train } t \text{ takes route } r \in R_t, \\ 0 & \text{otherwise.} \end{cases}$$
$$SR_{t,b} = \text{actual start reservation time of train } t \text{ on its } b\text{-th block section in the station area.}$$
$$DT_{t,b} = \text{actual departure time of train } t \text{ from its } b\text{-th block section in the station area.}$$
$$AR_{t,b} = \text{actual release time of train } t \text{ on its } b\text{-th block section in the station area.}$$
$$SD_t = \text{secondary delay of train } t \text{ measured in this station area.}$$
$$PD_t = \text{passenger delay of train } t \text{ measured in this station area.}$$

Note that $SR_{t,1}$ differs from δ_t in the fact that $SR_{t,1}$ can include additional delays due to conflicts in the station area. The index 0 in variable $x_{t,0}$ corresponds to the current route of train t. The non-binary decision variables are expressed in seconds.

As an objective function, a choice needs to be made between minimizing the secondary delays or the passenger delays. In case the secondary delays are minimized, the objective function becomes

$$\text{minimize} \quad \sum_{t=1}^{|T|} \Big(SD_t - \frac{x_{t,0}}{K}\Big), \tag{5}$$

with $SD_t = DT_{t,\text{nbBS}_r} - \Big(\delta_t + mr_{t,1} + \sum_{b=1}^{\text{nbBS}_r} mt_{t,b}\Big)$.

In case the passenger delays are minimized, the objective function equals

$$\text{minimize} \quad \sum_{t=1}^{|T|} \Big(PD_t - \frac{x_{t,0}}{K}\Big), \tag{6}$$

with $PD_t = PN_t \times DT_{t,\mathrm{nbBS}_r} - \Big(\delta_t + mr_{t,1} + \sum_{b=1}^{\mathrm{nbBS}_r} mt_{t,b}\Big)$.

The mathematical model also requires constraints to assure that any solution is feasible. The first constraints assure the minimum duration according to the blocking time theory.

$$
\begin{aligned}
SR_{t,1} &\geq \delta_t, \forall\, t \in T \\
\text{If } x_{t,r} == 1, \text{ then } DT_{t,b-1} &\geq SR_{t,b} + mr_{t,b}, \forall\, t \in T,\ b = 1, \ldots, \mathrm{nbBS}_r \\
\text{If } x_{t,r} == 1, \text{ then } DT_{t,b} &\geq DT_{t,b-1} + mt_{t,b}, \forall\, t \in T,\ b = 1, \ldots, \mathrm{nbBS}_r \\
\text{If } x_{t,r} == 1, \text{ then } AR_{t,b} &\geq DT_{t,b} + mc_{t,b}, \forall\, t \in T,\ b = 1, \ldots, \mathrm{nbBS}_r.
\end{aligned}
$$

Every train in the model should use exactly one route:

$$\sum_{r \in R_t} x_{t,r} = 1, \forall\, t \in T.$$

Two trains sharing an connection at the given station area, should always use the same station platform.

$$\text{If } b_p \in r_i \wedge b_p \in r_j, \text{ then } x_{t_i,r_i} = x_{t_j,r_j}, \ \forall\, \big(t_i, t_j, y\big) \in A.$$

Dependent on the type of connection, different constraints need to be imposed. These constraints are already described in Section 4.1.

$$
\begin{aligned}
DT_{t_k,b_p} &\geq DT_{t_i,b_p-1} + ma_{t_i,t_k}, \forall\, b_p \in B_p,\ t_i, t_k \in T,\ \big(t_i, t_k, \text{`coupling'}\big) \in A. \\
DT_{t_k,b_p} &\geq DT_{t_j,b_p-1} + ma_{t_j,t_k}, \forall\, b_p \in B_p,\ t_j, t_k \in T,\ \big(t_j, t_k, \text{`coupling'}\big) \in A, \\
DT_{t_j,b_p} &\geq DT_{t_i,b_p-1} + ma_{t_i,t_j}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`decoupling'}\big) \in A. \\
DT_{t_k,b_p} &> DT_{t_j,b_p}, \forall\, b_p \in B_p,\ t_j, t_k \in T,\ \big(t_i, t_k, \text{`decoupling'}\big) \in A. \\
AR_{t_i,b_p} &= SR_{t_j,b_p}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`turnaround'}\big) \in A, \\
DT_{t_j,b_p} &\geq SR_{t_j,b_p} + mt_{t_j,b_p} + ma_{t_i,t_j}, \forall\, b_p \in B_p,\ t_i, t_j \in T,\ \big(t_i, t_j, \text{`turnaround'}\big) \in A,
\end{aligned}
$$