An alternative graph formulation for train service planning under ETCS Level 2, Moving Block and Virtual Coupling signalling systems with dynamic speed supervision and continuous infrastructure representation

By

William Busuttil

as an additional graduation project at the Delft University of Technology

Student number: 5522196

Track: Transport and Planning (T&P)

Project Duration: 5 September 2022 – 10 February 2023

Thesis Committee:	Dr. ir. E. Quaglietta	TU Delft, Supervisor
	K. Rigos	TU Delft, Supervisor

Date: 13 February 2023

Abstract

To help maximize utilization of its existing infrastructure, the railway industry requires train service optimization models with headway constraints for state-of-the-art distance-to-go signalling systems. To this end, this paper introduces an approach for assessing train path feasibility in ETCS Level 2, Moving Block, and Virtual Coupling, that could be used to construct alternative graphs for timetable optimization. The methodology is the first for distance-to-go signalling systems that accounts for the impact of train acceleration and braking on the braking curve, guaranteeing a feasible and optimal timetable. The alternative graph models are specifically adapted to the train and signalling systems used, with the models for Moving Block and Virtual Coupling being the first with continuous representation of open track. The methodology is used to assess the capacity of Virtual Coupling and Moving Block, respectively, on the South West Main Line in the United Kingdom. The results show that Virtual Coupling can increase railway capacity compared to plain Moving Block, but this may depend on network topology.

1. Introduction

Railway infrastructure managers are continuously evaluating options to expand the capacity of their existing infrastructure to meet increased demand spurred by population growth and mode shift to greener transport systems. As physical infrastructure is expensive to build and maintain, railways have been investigating the adoption of distance-to-go and train-centric signalling systems to safely shorten train separation. In contrast with traditional fixed-block multi-aspect signalling that uses fixed detection sections to ensure safe train separation based on worst-case braking performance of trains on the line, distance-to-go systems use route information, train-specific operating characteristics, and positional reporting of the front end to model the braking curve more accurately. By migrating the vital track-clear detection functions from track to train, Moving Block (MB) signalling systems, such as Communication-Based Train Control (CBTC) and ETCS Level 3 (Janssens, 2022), remove the need to divide the track into fixed detection sections. Instead, these signalling systems allow for reduced safe separation distances over open track to the absolute stopping distance plus a safety margin for position error, and communication and control delays. While MB signalling can function safely without track-based detection systems, most implementations have a secondary track-based detection system (track circuits or axle counters) to facilitate operation of non-equipped vehicles and to mitigate the impact of potential adverse events on system availability. Over open track, the train-based track-clear detection is more permissive than the secondary system, allowing for shorter headways. At switches, if a train has a different route than its predecessor, the secondary detection system will release the switch faster than the train-based system because it does not require an allowance for positional uncertainty.

By allowing trains to operate closer than absolute stopping distance, Virtual Coupling (VC) signalling provides an opportunity to further reduce headways over open track relative to MB. VC requires an additional vehicle-to-vehicle communication system to exchange information on position, speed, and acceleration between trains. This functionality allows trains to form *convoys,* where the following distance is less than absolute stopping distance, and *platoons,* whereby trains synchronize their acceleration and braking behaviour to maintain a short following distance. Virtual Coupling could improve capacity relative to plain Moving Block because platooning trains are effectively treated as one train at important interlockings, and the shorter following distance reduces occupation time.

While there is significant interest in using distance-to-go signalling systems to improve capacity, limited research is available on timetable optimization specifically tailored to these systems. One of the main research gaps relevant to all distance-to-go signalling systems relates to acceleration assumptions used to represent the evolution of the braking curve over time, which is not representative of platoon formation constraints in plain Moving Block or Virtual Coupling. In platoon formation, the follower train must approach its leader at a higher speed and decelerates to the leader's speed without violating safety constraints at any time. Existing models for Moving Block and Virtual Coupling also rely on discretization of open track into very short blocks, which could yield sub-optimal schedules. To address these limitations, this paper proposes novel methodologies for computing blocking times for ETCS Level 2, and for performing timetable compression for Moving Block and Virtual Coupling operations using the dynamic safety margin from Quaglietta et al. (2022). The main contributions of this paper are to:

(i) Propose a new microscopic and innovative methodology for optimally scheduling a line with a distance-to-go signalling system with detailed modelling of continuous transmission and brake supervision. In particular, the methodology and supporting model:
   - is the first for Moving Block and VC with continuous representation of open track; and
   - is practical and applicable to scheduling lines with ETCS Level 2, Moving Block, VC, as well as hybrid operations scenarios, and can accommodate future changes to the VC safety margin if required.

(ii) Apply and assess this new methodology by using the model to construct an alternative graph for scheduling trains, and perform a comparative capacity assessment between plain MB and VC for the Waterloo-Woking corridor on the South West Main Line in the United Kingdom (UK).

2. Literature Review

2.1 Review of Literature on Virtual Coupling (VC) Technologies

VC is the first signalling system that allows trains to operate closer than absolute stopping distance. Most existing research on VC focuses on its technical feasibility, system architecture, and potential use cases. Di Meo et al. (2020) performed a technical analysis to assess the feasibility of expanding the ERTMS/ETCS operating modes to include VC. The paper concluded that Virtual Coupling is viable within the ETCS standard, but that additional train-to-train communication equipment would be required to ensure safe convoy operations. Quaglietta et al. (2020) outlined operating principles of VC-over-ETCS, the architecture of the vital train-to-train communications system, and the form of the safety-critical messages communicated between trains (Figure 1). If trains have a common route, the leading train has TIM, and the V2V communication link has been established, the trains can transition from ETCS Level 3 Moving Block Running to the 'coupling' state whereby the following train approaches the leader at a higher speed to reduce the separation distance with the leader while respecting the Virtual Coupling safety distance requirement. The trains are said to be part of a *convoy* when the separation distance is less than the ETCS Level 3 Moving Block safety distance requirement. When the follower reduces the speed to that of the leader, the trains transition to the *coupled running* phase forming a *platoon*, synchronizing their acceleration and braking activities to maintain the short following distance. The coupled running phase ends when the trains stop synchronizing their control actions. Decoupling could occur because a train cannot match the actions of the leader over open track (referred to as *unintentional decoupling),* or the paths of the two trains diverge due to a diverging switch or scheduled stop (*intentional decoupling).* These principles were used to develop a train-following model implemented in the microscopic simulator EGTRAIN to assess capacity impacts. Quaglietta et al. (2022) proposed a dynamic safety margin that considers safety risks not addressed by a static safety margin. The proposed safety margin ensures that the following train has sufficient braking distance in the event of an emergency brake application by the leader. Virtual Coupling with the dynamic safety margin achieved safe headway reductions over plain Moving Block when simulated on the South West Main Line in the UK.

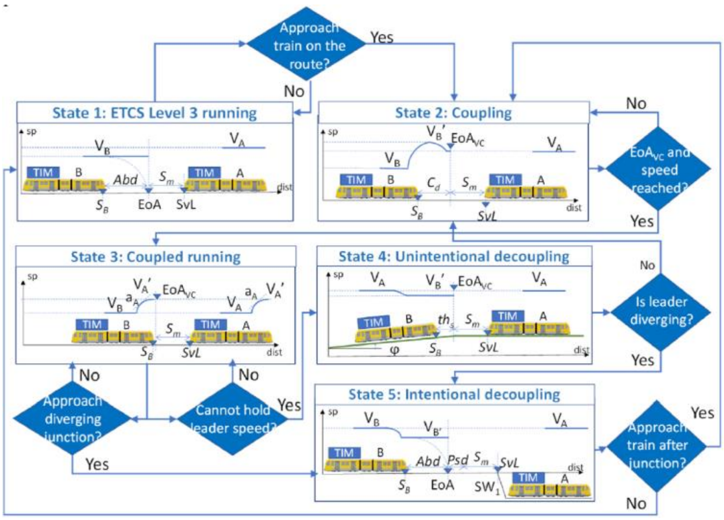

Figure 1: VC-over-ETCS operating states, and transition conditions (Quaglietta et al, 2020)

A property of the dynamic safety margin is that the VC safety distance requirement is a function of two braking distance terms: one representing the relative braking distance between leader and follower based on the service braking rate, and another for the emergency braking safety margin. In scenarios when the leader's speed is very low relative to the follower, the braking distance of the leader will approach zero, and the VC safety distance requirement could exceed the absolute stopping distance requirement for Moving Block running. The safety distance requirement for VC operations with dynamic safety margin the more permissive of relative braking distance plus the dynamic safety margin, and the Moving Block safety distance.

Some research has been performed to produce capacity-effective timetables or trajectories for Virtual Coupling with macroscopic infrastructure representation, or at the microscopic level for operations with a static safety margin. Gallo et al. (2020) developed a model to schedule virtually coupled carriages on a circular line, with the objective of minimizing empty seat milage. Gallo et al. (2021) expanded this model to schedule on-demand services in a branched metro network, where vehicles not needed immediately could be shunted into sidings at all stations. While able to improve rolling stock utilization, both models relied on macro representation of the infrastructure, with nodes and arcs representing stations and open track, respectively. Li et al. (2021) developed a macro level scheduling tool for metros that considers weighted vehicle utilization and generalized passenger costs, including the effect of scheduling decisions on station dwell times (and in turn trajectories). Cheng et al. (2021) developed a methodology for managing the coupling of two trains after merging from different routes that assumes a static safety margin. This approach could result in hazardous situations because it does not consider the safety risks that the dynamic safety margin addresses. Miao (2021) developed a mixed-integer quadratic programming model to optimize the speed profiles of trains on a high-speed rail network with two branches, given a predetermined train order at junctions and a static safety margin. The model, which represents train trajectories at the microscopic level, was able to deliver capacity gains compared to conventional moving block signalling. However, use of a static safety margin (as opposed to a dynamic one) may result in infeasible timetables under safe Virtual Coupling operations.

At present, nearly all microscopic optimization for Virtual Coupling has been performed with a static safety margin. While Moving Block running is a component of Virtual Coupling operations for any safety margin specification, the Moving Block safety distance over open track previously could be ignored because it is never more permissive than relative braking distance plus a static margin. The introduction of the dynamic safety margin necessitates development of control systems that can identify the more permissive safety distance at a given time.

2.2 Representation of signalling constraints in existing distance-to-go systems.

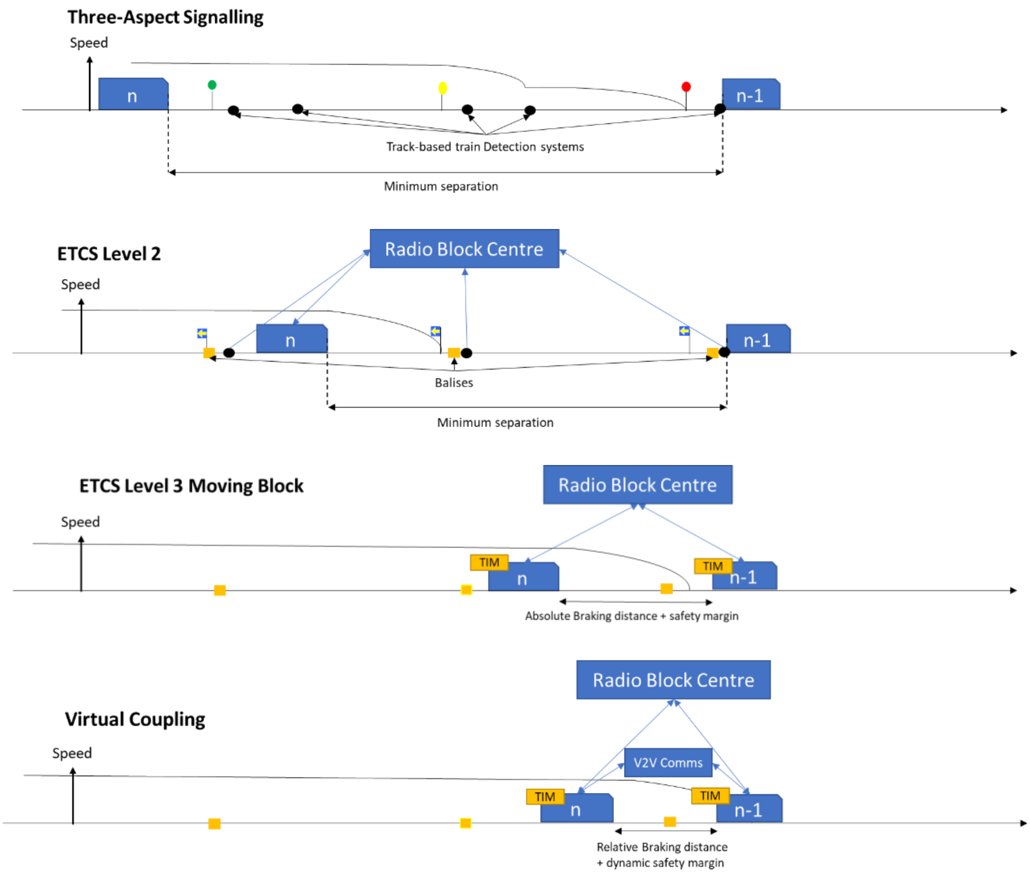

Figure 2: Architecture of Multi-aspect Signalling and different types of Distance-to-go Signalling

Some research has proposed mixed-integer linear programming (MILP) representations of ETCS Level 2, which relies entirely on track-based track-clear detection, for the purpose of real-time rescheduling. Xu et al. (2017) proposed a real-time rescheduling model for a high-speed line with CTCS-3 (functionally equivalent to ETCS Level 2) with signalling constraints for five different speed levels, where trains only comply with the constraint for their speed level. While capable of resolving conflicts, this model may not fully use available capacity because it assumes that the signalling system can only enforce the five defined speed levels, so headway constraints must be based on the worst-case scenario for that level. Luo et al. (2021) proposed a bi-level optimization rescheduling process with the headway constraint representation that requires fewer variables and constraints. This approach reduced computation times; however, modelling ETCS Level 2 operations with discrete speed levels limits its ability to take full advantage of available capacity. Long et al. (2019) proposed a cell-based rescheduling model for scenarios with a speed restriction on a high-speed line, where the number of cells occupied by a train can be varied depending on its speed. While this approach reduced delays compared to methods that do not consider the relationship between speed and headway at all, the methodology only checks headway constraints when trains enter cells. The intermittent checking of signalling constraints could result in sub-optimal capacity utilization.

To date, Moving Block signalling has mostly been used in metro systems, where stopping patterns are homogenous, and the minimum headway depends mostly on dwell times at stations. As infrastructure managers examine the potential benefits of implementing Moving Block signalling on lines with heterogenous service patterns and rolling stock, some work has been performed on developing timetables and rescheduling models for these scenarios. Schlechte et al. (2022) developed a microscopic approach to schedule lines with Moving Block signalling (referred to as *velocity expansion*) where minimum headway depends on both the leader's and follower's speed profiles. While it can produce feasible schedules, the calculation method for minimum headways depends partly on an 'oracle' to calculate the minimum headway between the two trains at the start of the section, the end of the section, and at the time that the following train enters the section. Janssens (2022) developed a real-time scheduling model for Moving Block where the track is divided into detection segments the size of the average train length plus a safety margin to approximate the route release process. The calculation of headway arc weights is based on blocking time theory, which requires precise specification of the 'block' it is being calculated for. The need to divide the infrastructure into a finite number of blocks limits the applicability of blocking time theory in continuous space.

The literature review did not return research on timetabling or real-time rescheduling for distance-to-go signalling using ETCS Level 2, plain Moving Block, or Virtual Coupling (with any safety margin specification) that modelled the continuous movement of the braking curve. All research papers found divided the infrastructure layout into detection sections with intermittent headway constraints based on events at times when trains enter or exit a section, or with headway constraints generated through simulation-based methods. The lack of representation of signalling constraints at other times requires use of a default constraint based on the worst-case scenario that costs additional capacity. Existing models create their speed-related constraints assuming trains are cruising with zero acceleration, limiting their applicability for modelling platooning formation Moving Block or Virtual Coupling. Most papers rely on blocking time theory for their constraints, which requires the infrastructure to be discretized into a finite number of track sections to be able to create a model in polynomial time. While blocking time theory is an appropriate method for ETCS Level 2, which relies entirely on track-based track-clear detection, it cannot be used to create constraints in continuous space because of the need to divide the line into an infinite number of zero-length blocks. For infrastructure managers to realize the capacity benefits of distance-to-go signalling systems, their processes (and associated constraints) must be modelled accurately. For modelling platooning operations, there is a need for a methodology for creating constraints for the platoon formation process, where a train must approach its leader at higher speed, then decelerate to the same speed while always complying with the safety distance requirement in continuous time and space.

Construction of a feasible timetable requires modelling of signalling constraints at interlockings and over open track for different train orders. This approach is modelled through a microscopic alternative graph with alternative arcs representing the minimum process times between events on different trains. In traditional fixed-block multi-aspect signalling systems, alternative arcs represent the minimum time between when the leading train releases a block section and a follower can occupy it. Since every block in a fixed block system is either completely occupied or not occupied at all, the representation of the finite number of occupation and release events is sufficient to construct a correct alternative graph. This method is not directly applicable to distance-to-go signalling systems because the area that a train needs to block is continuously changing.

A new methodology is needed for calculating the weights of the headway arcs with continuous movement of the brake curve/safety margin for trains operating in any distance-to-go signalling system, and the continuous release of the rear of trains with integrity monitoring in Moving Block and in virtually-coupled convoys.

3. Methodology

The proposed model is a novel alternative-graph (AG) formulation that can aid Infrastructure Managers in determining the order of trains for a set of requests in a periodic timetable that minimizes cycle time t_{cycle}. The objective function is:

$$\min t_{cycle}$$

Each train $n \in T$ has a predetermined route through the modelled area with scheduled station stops and no opportunities for local rerouting. The expansion of the model to allow for local rerouting possibilities is the subject of future work. Each train n's route is represented as a series of homogenous behavioural intervals B_n where acceleration is assumed constant. Within each interval, a speed arc $v_{n,i}(v_0, v_1) \in V_{n,i}$ is selected, representing train n's start speed v_0 and end speed v_1 for the interval, and associated acceleration. Every $n \in T$'s position $s_n(t)$ and speed $v_n(t)$ are both continuous when running in the modelled network.

The discretization of the train trajectory into homogenous behavioural intervals can be done without discretizing the infrastructure: the area blocked by train n is a function of its trajectory in continuous time and space. The continuous nature of $s_n(t)$ and $v_n(t)$ with respect to time means that the brake curve of the train, which is continuous with respect to these variables, can be expressed in as a function that is continuous with respect to time. For constructing fixed-block and interlocking constraints, the continuous movement of the brake curve over time means that it is possible to calculate the exact time that a train starts to occupy a block by Intermediate Value Theorem. The process for generating these constraints is shown in Section 3.2. For operations over open track in Moving Block or Virtual Coupling, the continuous movement of the leader's max safe rear end and the follower's brake curve (or safety distance in VC) over time means that the length of the unoccupied track between the two trains has a local minimum by Extreme Value Theory. If the local minimum is non-negative, the trajectories of the two trains are conflict-free. This forms the basis of the train-centric compression method outlined in Section 3.3.

The infrastructure representation requirements are as train-centric as the actual signalling system in use. For any pair of trains $n, n-1 \in T$, where $n-1$ is ordered before n, the set of well-defined blocks for the order is the set $FB_{n-1,n}$. This set contains the machines with clearly defined start and end locations, as in the traditional job-shop scheduling problem. Each machine in the set cannot be occupied by both trains at once. If train $n-1$ cannot use TIM for track release, the set contains all fixed blocks both trains need to occupy on their respective routes. If train $n-1$ can use TIM for track release, the set contains the fixed blocks with merging and diverging switches, and virtual blocks representing the area train $n-1$ occupies while dwelling at stations on the open track section (if there are any). In Moving Block or Virtual Coupling, it is necessary to model scheduled station stop constraints as fixed blocks because the methodology for creating continuous-space open-track constraints in Section 3.3 assumes that the arc process times are precisely defined, which is not the case for dwelling arcs. For leading train $n-1$, the only mandatory behavioural nodes are at location where train releases a fixed block $fb \in FB_{n-1} = \bigcup_{\substack{i \in T \\ i \neq n}} FB_{n-1,n}$, that it may have to release for another train using the track-clear detection, and arrival and departure nodes for scheduled station stops. For any fixed block fb, the location (in metres from the start of train n's route) of the Supervised Location (SvL) is represented by parameter $s_{SvL,n,fb}$, and the release point by parameter $s_{rel,n,fb}$. To generate desirable speed profiles, more nodes can be added at the modellers' discretion, though this is not strictly necessary to represent signalling constraints. The placement of additional behavioural nodes can be customized for each train: there is no requirement that node locations be the same for all trains. An example of the infrastructure representation showing only mandatory node locations for a pair of trains $n-1$ and n, where both trains have TIM, and train $n-1$ stops at a station on the open track section, is shown in Figure 4 below:

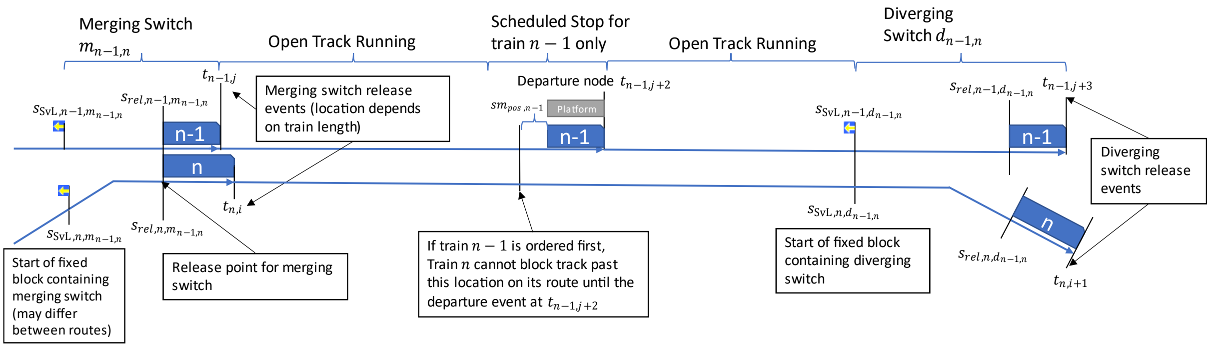

Figure 3: Moving Block and Virtual Coupling Infrastructure representation and mandatory node locations for trains n and $n-1$, where $n-1$ makes a stop on the open track section, and train n makes no stops

As shown in Figure 4, both trains have mandatory behavioural nodes for the events where they release the blocks with merging switch $m_{n-1,n}$ and diverging switch $d_{n-1,n}$. These nodes, which are used to identify applicable constraints during each arc running period, should be placed at the location of train's front end when the tail clears the relevant release point. Each train also has an associated SvL location that it cannot pass until the other train (if ordered first) releases the block. Train $n-1$ also has additional mandatory behavioural nodes representing the arrival and departure events at the station over open track. The SvL for train n (if ordered after) that is associated with the departure event at $t_{n-1,j+2}$ is the location of train $n-1$'s max safe rear end while dwelling, which also includes an allowance for position error because TIM is being used. The corresponding SvL for train n is:

$$s_{n-1,j+2} - l_{n-1} - sm_{pos,n-1} + s_{rel,n,m_{n-1,n}} - s_{rel,n-1,m_{n-1,n}}$$

where $s_{n-1,j+2}$ is the location (in metres from the start of train $n-1$'s route) of train $n-1$'s stop position in the station, l_{n-1} is the length (in metres) of train $n-1$, and $sm_{pos,n-1}$ is the (constant) worst-case position error of train $n-1$. The term $(s_{rel,n,m_{n-1,n}} - s_{rel,n-1,m_{n-1,n}})$ is a conversion factor to convert locations on train $n-1$'s route to locations on train n's route, as the release point for the merging switch is at the same location in the network for both trains.

The process of creating the running arcs occurs outside the optimization process, and the exact specification can be tailored to the use case. The created speed profiles should respect restrictions regarding line speed limits, track geometry, tractive effort and running resistances. The initial and final speed should be zero for behavioural intervals beginning with a departure event from a station and ending at a scheduled stop.

3.1 Run Time Constraints

If the position $s_{n,i}$ of behavioural node i (in metres from where train n's route starts) is the same as the position of node $i+1$, the interval between them is a dwelling arc. A minimum dwell time constraint is added:

$$t_{n,i+1} \geq t_{n,i} + \tau_{dwell,n,i} \;\; \forall i \in B_n \forall n \in T\colon s_{n,i} = s_{n,i+1} \tag{1}$$

where $t_{n,i}$ is the event time for behavioural node i, and $\tau_{dwell,n,i}$ is the minimum dwell time (in seconds). For other homogenous behavioural intervals where trains are running through the network, the following constraints are added:

$$t_{n,i+1} = t_{n,i} + \sum_{\{v_0,v_1\} \in V_{n,i}} \tau_{r,n,i}(v_0, v_1) * z_{n,i,v_0,v_1} \;\forall i \in B_n \forall n \in T\colon s_{n,i} < s_{n,i+1} \tag{2}$$

$$\sum_{\{v_0,v_1\} \in V_{n,i}} z_{n,i,v_0,v_1} = 1 \;\; \forall i \in B_n \forall i \in B_n \forall n \in T\colon s_{n,i} < s_{n,i+1} \tag{3}$$

$$\sum_{\{v_0,v_1\} \in V_{n,i}} z_{n,i,v_0,v_1} = \sum_{\{v_1,v_2\} \in V_{n,i+1}} z_{n,i+1,v_1,v_2} \;\; \forall v_1 \; \forall i \in B_n \; \forall n \in T\colon s_{n,i} < s_{n,i+1} < s_{n,i+2} \tag{4}$$

Where z_{n,i,v_0,v_1} is a binary variable equal to 1 is the running arc $v_{n,i}(v_0, v_1) \in V_{n,i}$ between nodes i and $i+1$ is chosen, or 0 otherwise. Constraint (2) establishes the run time between nodes i and $i+1$ equal to the run time $\tau_{r,n,i}(v_0, v_1)$ of running arc $v_{n,i}(v_0, v_1) \in V_{n,i}$, if it is selected. Constraint (3) requires that exactly one running arc be selected for each non-dwelling homogenous behavioural interval. Constraint (4) requires that at any running interval starting at node i that is followed by another running interval starting at node $i+1$, the ending speed of the running arc chosen must be the same as the starting speed for the running arc chosen interval starting at node $i+1$. These linking constraints ensure that the train's speed $v_n(t)$ is continuous at each behavioural node, which guarantees that speed is continuous over the entire interval that train n is within the modelled network.

To model represent the requirement to order trains, the binary variable $z_{n-1,n}$ equal to 1 if train $n-1 \in T$ is ordered before train $n \in T$, is introduced. To ensure that only one order can be selected between a pair of trains, constraint (5) is added.

$$z_{n-1,n} + z_{n,n-1} = 1 \;\; \forall n-1, n \in T \tag{5}$$

The fixed- and moving-block headway constraints use a multiple-Big M formulation. To ensure that this does not distort the solution space, all event times are bounded in the time interval $[0, \tau_{max}]$, where τ_{max} is sufficiently high for all events to be scheduled:

$$0 \leq t_{n,i} \leq \tau_{max} \; \forall i \in B_n \forall n \in T \tag{6}$$

The value τ_{max} is used as the Big-M value for the headway constraints. Constraint (6) ensures that these constraints will be non-binding unless all conditions are met, and that if a headway constraint is non-binding (because at least one condition for it is not met), the number of conditions not met does not affect the size of the solution space.

3.2 Constraints for Interlockings, Fixed Blocks, or when the Leading Train is dwelling in Moving Block.

To represent the continuous occupation process that characterizes all distance-to-go signalling systems, the *job-starting frontier* $JS_n(t)$ is defined as the furthest location blocked by train n at time t based on the absolute stopping distance: all locations past this point on the route can be occupied by preceding trains. The equation for the frontier is:

$$JS_n(t) = s_n(t) + \frac{\left(v_n(t)\right)^2}{2b_{s,n}} + sm_{pos,n} + sm_0$$

where $b_{s,n}$ is the train's service braking rate, $sm_{pos,n}$ is the allowance for position error (assumed to be a constant worst-case value), and sm_0 is a constant safety margin for other factors. The continuous properties of $s_n(t)$ and $v_n(t)$ means that $JS_n(t)$ is also continuous on the time interval $[t_{n,0}, t_{n,exit}]$. In other words, for any fixed block $fb \in FB_{n-1,n}$ that leading train $n-1$ must release for n using the trackside track-clear detection, or through TIM when departing a scheduled station stop, and $JS_n\left(t_{n,0}\right) < s_{SvL,n,fb}$, it is possible to calculate the time that train n first occupies the supervised location (i.e. when $JS_n(t) = s_{SvL,n,fb}$) by Intermediate Value Theory.

This calculation is done by searching all running arcs $v_{n,i}(v_0, v_1) \in V_{n,i}, i \in B_n$ comprising the path where occupation of fb starts during the arc's run time. $JS_n(t)$ is continuous on any arc running interval $[t_{n,i}, t_{n,i+1}]$, so by Extreme Value Theory, there exists a maximum and minimum blocked location for each running arc that can be used to determine if fb is blocked during that arc. Pseudocode for determining if a binding constraint exists, and calculating occupation lag time $\tau_{lag,n,i,fb}(v_0, v_1)$ for train n to the SvL is shown below:

Algorithm 1: *Generation of Headway Arc Weight for Train n with Running Arc $v_{n,i}(v_0, v_1)$ for Block fb*	
Input	*Train position $s_n(t)$, speed $v_n(t)$, block SvL location on train $n's$ route $s_{SvL,n,fb}$, Job-starting frontier $JS_n(t)$ of train n, arc run time $\tau_{r,n,i}(v_0, v_1)$*
	Set $z_{n,i,v_0,v_1} = 1$
	If $JS_n\left(t_{n,i}\right) > s_{SvL,n,fb}$ ***or*** $\max\limits_{[t_{n,i},t_{n,i+1}]} JS_n(t) < s_{SvL,n,fb}$ ***then***
	Return *'No Constraint'*
	Else
	Return $min(t)\; s.t.$
	$JS_n\left(t + t_{n,i}\right) = s_{SvL,n,fb}$
	$0 \leq t \leq \tau_{r,n,i}(v_0, v_1)$
Output	*Confirmation that no constraint is needed, or reservation lag time $\tau_{lag,n,i,fb}(v_0, v_1)$ from $t_{n,i}$*

If running arc $v_{n,i}(v_0, v_1) \in V_{n,i}, i \in B_n$ is one where occupation of fixed block fb by train n starts during its running interval, and $t_{n-1,j}$ is the time that train $n-1$ clears the block, the constraint is added as:

$$t_{n,i} + \tau_{lag,n,i,fb}(v_0, v_1) - \tau_{comm} - \tau_{ctrl,n} \geq t_{n-1,j} + \tau_{setup,fb,n-1,n} - \tau_{max}(2 - z_{n,i,v_0,v_1} - z_{n-1,n}) \tag{7}$$

The left side represents the time that train n begins occupation of fb if the running arc $v_{n,i}(v_0, v_1)$ is selected. $\tau_{lag,n,i,fb}(v_0, v_1)$ is the time from when the train passes behavioural node $B_{n,i}$ to when its braking curve enters the block. τ_{comm} is the signalling system communication delay. $\tau_{ctrl,n}$ is the train control delay for train n. The right side is the earliest time that the route can be set up if train $n-1$ is ordered first at the block, which is the release time $t_{n-1,j}$ plus the time $\tau_{setup,fb,n-1,n}$ required to release $n-1$'s route and set up train n's route at the block. This value is dependent on whether the two train's routes differ, and whether the route release is through trackside track-clear detection, or though TIM (in the case where the block modelled is associated with train $n-1$ serving a station over open track). A diagram of the derivation of the occupation lag time is shown in Figure 4.

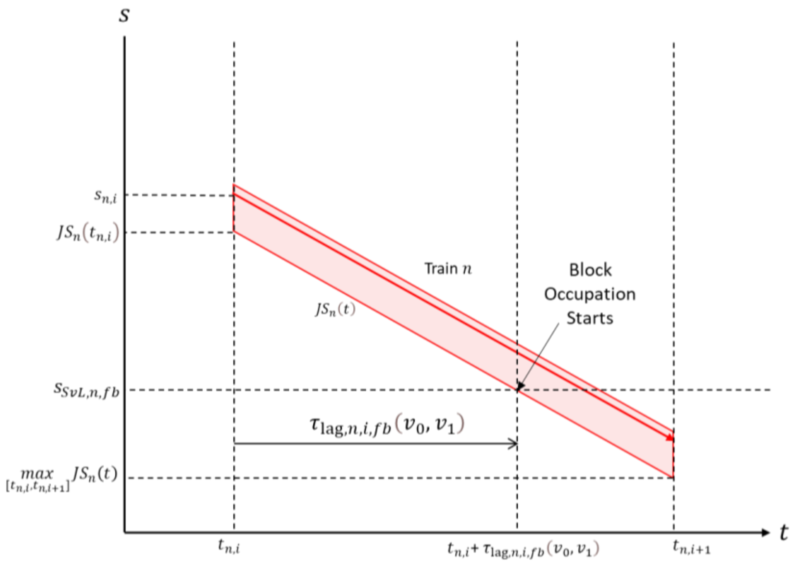

Figure 4: Derivation of occupation lag for arc $v_{n,i}(v_0, v_1)$ for fixed block fb with SvL at $s_{SvL,n,fb}$

3.3 Moving Block and Virtual Coupling

The addition of TIM plus position monitoring of the tail allows leading train $n-1$ to continuously release the track between the merging and diverging switch. Trains n and $n-1$ are in open track running during the time interval $[t_{n-1,j}, t_{n-1,k}]$, where $j \in B_{n-1}$ is the behavioural node where train $n-1$ releases the block with the merging switch, and $k \in B_{n-1}$ is the behavioural node where the diverging switch block is released. To model the continuous release of open track in this interval, the *job-ending frontier* $JE_{n-1,n}^{n-1}(t)$ of train $n-1$ from the perspective of train n is defined as:

$$JE_{n-1,n}^{n-1}(t) = s_{n-1}(t) - l_{n-1} - sm_{pos,n-1} + \left(s_{rel,n,m_{n-1,n}} - s_{rel,n-1,m_{n-1,n}}\right)$$

where $s_{n-1}(t)$ is the position of the front end (in metres from the start of train $n-1$'s route), l_{n-1} is the train length, $sm_{pos,n-1}$ is the allowance for position error, and $s_{rel,n,m_{n-1,n}} - s_{rel,n-1,m_{n-1,n}}$ is the conversion factor to express the position of train $n-1$ over open track in terms of train n's route. This function is continuous in the open track running interval, so $JE_{n-1,n}^{n-1}(t) - JS_n(t)$ is also continuous on $[t_{n-1,j}, t_{n-1,k}]$, and has a local minimum by Extreme Value Theory. If one can show:

$$\min_{[t_{n-1,j}, t_{n-1,k}]} JE_{n-1,n}^{n-1}(t) - JS_n(t) \geq 0$$

Then the paths of trains $n-1$ and n are feasible in plain Moving Block over open track.

Virtual Coupling allows trains to operate at relative braking distance plus the dynamic safety margin. This safety distance can be represented as the job-starting frontier:

$$JS_{n-1,n}^{n,VC}(t) = s_n(t) + \frac{\left(v_n(t)\right)^2 - \left(v_{n-1}(t)\right)^2}{2b_{s,n}} + \max\left(\frac{\left(v_n(t)\right)^2}{2b_{s,n}} - \frac{\left(v_{n-1}(t)\right)^2}{2b_{e,n-1}}, 0\right) + sm_{pos,n} + sm_0$$

Where $b_{e,n-1}$ is the emergency braking rate of leading train $n-1$. The frontier is equal to the position of train n plus the relative braking distance with the leader, the emergency braking safety margin, position error of train n, and the static safety margin. To remove the max term, the job is split into two jobs $JS_{n-1,n}^{n,RBD}(t)$, representing the relative braking distance requirement, and $JS_{n-1,n}^{n,Emerg}(t)$, representing the emergency braking requirement. These are:

$$JS_{n-1,n}^{n,RBD}(t) = s_n(t) + \frac{\left(v_n(t)\right)^2 - \left(v_{n-1}(t)\right)^2}{2b_{s,n}} + sm_{pos,n} + sm_0$$

$$JS_{n-1,n}^{n,Emerg}(t) = s_n(t) + \frac{\left(v_n(t)\right)^2 - \left(v_{n-1}(t)\right)^2}{2b_{s,n}} + \frac{\left(v_n(t)\right)^2}{2b_{s,n}} - \frac{\left(v_{n-1}(t)\right)^2}{2b_{e,n-1}} + sm_{pos,n} + sm_0$$

The Virtual Coupling job-starting frontiers are both continuous, so it follows that $JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t)$ and $JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,Emerg}(t)$ are also continuous on the open-track running interval, and take on local minimums by Extreme Value Theory. If one can show:

$$\min_{[t_{n-1,j}, t_{n-1,k}]} JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t) \geq 0, \text{and}$$

$$\min_{[t_{n-1,j}, t_{n-1,k}]} JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,Emerg}(t) \geq 0$$

Then the paths of trains $n-1$ and n are conflict-free in Virtual Coupling over the entire open-track running section without any Moving Block running. If there exists $t \in \left[t_{n-1,j}, t_{n-1,k}\right]$ where:

$$JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t) < 0,$$

It is not possible to use Moving Block running (whose safety distance is never lower than RBD) to achieve feasibility in this section, so the paths are infeasible. If the relative braking distance is never violated, but there exists an interval $[t_s, t_e] \subseteq \left[t_{n-1,j}, t_{n-1,k}\right]$ where:

$$JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t) < 0 \; \forall t \in [t_s, t_e]$$

The paths could still be feasible if the trains use the MB safety distance in $[t_s, t_e]$. If:

$$\min_{[t_s, t_e]} JE_{n-1,n}^{n-1}(t) - JS_n\left(t + \tau_{ctrl,n} - \max\left(\tau_{ctrl,n} - \tau_{ctrl,n-1}, 0\right)\right) \geq 0$$

Then the paths are feasible over through a combination of Moving Block running and Virtual Coupling. The term $\tau_{ctrl,n} - \max(\tau_{ctrl,n} - \tau_{ctrl,n-1}, 0)$ is the additional train control delay allowance for the Moving Block running compared to Virtual Coupling.

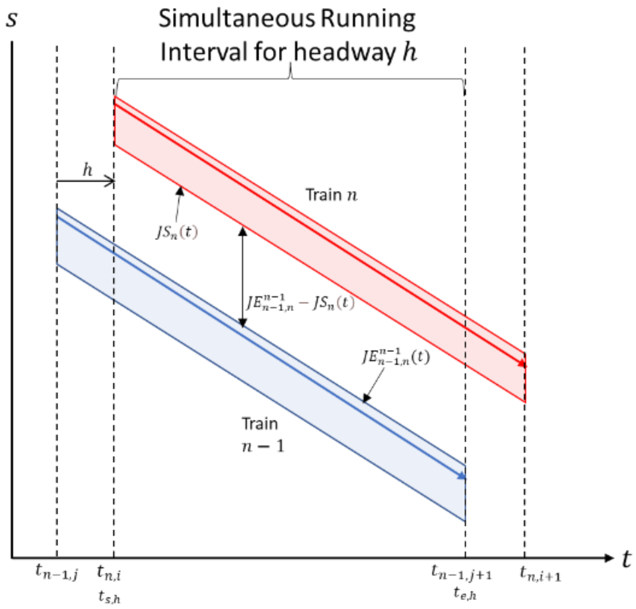

Figure 5: Definition of the simultaneous running interval for headway h, and job-starting and job-ending functions for the two trains in the feasibility-checking process for two arcs for plain Moving Block

The process of creating open-track constraints for plain Moving Block and Virtual Coupling is performed at the level of the running arcs, which form the train trajectories. The headway calculation must be performed for every pair of nodes $i \in B_n$ and $j \in B_{n-1}$ for every pair of running arcs $v_{n,i}(v_0, v_1) \in V_{n,i}$ and $v_{n-1,j}(v_s, v_e) \in V_{n-1,j}$ that start at the respective nodes. In practice, most of the running arc pairs will not have a binding constraint, but this cannot be established without first checking for possible conflicts. An open track constraint cannot exist if leading train $n - 1$ has not yet cleared the merging switch ($s_{n-1,j} < s_{rel,n-1,m_{n-1,n}} + l_{n-1}$) or has already cleared the diverging switch ($s_{n-1,j} \geq s_{rel,n-1,d_{n-1,n}} + l_{n-1}$). For any pair of arcs $v_{n,i}(v_0, v_1) \in V_{n,i}$ and $v_{n-1,j}(v_s, v_e) \in V_{n-1,j}$, a given headway h: $t_{n,i} = t_{n-1,j} + h$ is considered feasible if there is no time t where both trains claim a given piece of infrastructure. The common running period starts at the later of the two arc start times $t_{s,h} = \max(t_{n-1,j}, t_{n-1,j} + h)$ and ends at the earliest of the two arc finish times $t_{e,h} = \min(t_{n-1,j} + \tau_{r,n-1,j}(v_s, v_e), t_{n-1,j} + h + \tau_{r,n,i}(v_0, v_1))$. The interval $[t_{s,h}, t_{e,h}]$ is when both arcs are occupying infrastructure, and thus the only time interval where a conflict could occur. If the trajectories are feasible in $[t_{s,h}, t_{e,h}]$, headway h is feasible. Figure 5 shows the components of the feasibility-checking process for plain Moving Block.

If $h \notin [-\tau_{r,n,i}(v_0, v_1), \tau_{r,n-1,j}(v_s, v_e)]$, the simultaneous running period is empty ($t_{s,h} > t_{e,h}$) so h would be feasible by default. The minimum headway (if it exists) must therefore be in the interval $[-\tau_{r,n,i}(v_0, v_1), \tau_{r,n-1,j}(v_s, v_e)]$. It is possible for this minimum headway to be negative: this means train n can enter its arc at $t_{n,i}$ before behavioural event $t_{n-1,j}$, but if $t_{n-1,j}$ does not occur before the end of running arc $v_{n,i}(v_0, v_1)$ at $t_{n,i+1}$, a conflict will occur in the running interval. If $h = -\tau_{r,n,i}(v_0, v_1)$ is feasible, there exists no constraint because every headway is feasible. If $h = \tau_{r,n-1,j}(v_s, v_e)$ is infeasible, then it is not possible for running arc $v_{n,i}(v_0, v_1)$ of the follower to start until after the leader has exited running arc $v_{n-1,j}(v_s, v_e)$ (i.e. $t_{n,i} > t_{n-1,j+1}$). Since behavioural event $t_{n,i}$ is also contained in the running interval for train n starting at node $i-1$, there is no need to add an additional constraint, as the previous headway constraint between $t_{n-1,j}$ and $t_{n,i-1}$ for the running arc $v_{n-1,j}(v_s, v_e)$ and the running arc $v_{n,i-1}(v_{-1}, v_0) \in V_{n,i-1}$ of train n ensures that $t_{n,i} > t_{n-1,j+1}$. The process for checking if a minimum headway exists for a generic train-centric signalling system c, and searching for it using a binary search, is shown below:

Algorithm 2: *Generation of Headway Arc Weight $h_{min,c}\left(v_{n-1,j}(v_s, v_e), v_{n,i}(v_0, v_1)\right)$ between Running Arc $v_{n,i}(v_0, v_1)$ of following train n and $v_{n-1,j}(v_s, v_e)$ running arc of leading train $n-1$ for signalling system c*

Input *Both trains' positions $s_n(t)$, $s_{n-1}(t)$ and speeds $v_n(t)$, $v_{n-1}(t)$ in their arcs, all relevant train characteristics, Criteria for feasibility for signalling system c, Release point locations $s_{rel,n-1,m_{n-1,n}}$ and $s_{rel,n-1,d_{n-1,n}}$ for the merging and diverging switches for leader $n-1$'s route*

Set $z_{n,i,v_0,v_1} = 1$ *and* $z_{n-1,j,v_s,v_e} = 1$
If $s_{n-1,j} \geq s_{rel,n-1,d_{n-1,n}} + l_{n-1}$ ***or*** $s_{n-1,j} < s_{rel,n-1,m_{n-1,n}} + l_{n-1}$ ***then***
    Return *'No Constraint'*
Else If $h = -\tau_{r,n,i}(v_0, v_1)\ feasible$ ***or*** $h = \tau_{r,n-1,j}(v_s, v_e)\ infeasible$ ***then***
    Return *'No Constraint'*
Else
    $high = \tau_{r,n-1,j}(v_s, v_e)$
    $low = -\tau_{r,n,i}(v_0, v_1) + 1$
    While $low \leq high$ ***do***
        $h = \left\lfloor \dfrac{high + low}{2} \right\rfloor$
        If $h\ infeasible$ ***then***
            If $h + 1\ feasible$ ***then***
                Return $h + 1$
            Else
                $low = h + 1$
        Else if $h = -\tau_{r,n,i}(v_0, v_1) + 1$ ***then***
            Return h
        Else
            $high = h - 1$
    End While

Output *Arc weight $h_{min,c}\left(v_{n-1,j}(v_s, v_e), v_{n,i}(v_0, v_1)\right)$ between, or confirmation there is no constraint*

This headway calculation process can be used for both plain Moving Block and Virtual Coupling, with feasibility being checked using the processes described earlier in the section. This search method is flexible enough to be applicable for VC operations even if future research identifies new requirements for the safety margin.

If plain Moving Block operations are being modelled, and there exists a binding minimum headway $h_{min,MB}\left(v_{n-1,j}(v_s, v_e), v_{n,i}(v_0, v_1)\right)$, the constraint with communication and control delay is:

$$t_{n,i} \geq t_{n-1,j} + h_{min,MB}\left(v_{n-1,j}(v_s, v_e), v_{n,i}(v_0, v_1)\right) + \tau_{comm} + \tau_{ctrl,n} - \tau_{max}\left(3 - z_{n,i,v_0,v_1} - z_{n-1,j,v_s,v_e} - z_{n-1,n}\right) \qquad (8.MB)$$

If Virtual Coupling operations are being modelled, and there exists a binding minimum headway $h_{min,VC}\left(v_{n-1,j}(v_s,v_e),v_{n,i}(v_0,v_1)\right)$, the constraint with communication and control delay is:

$$t_{n,i} \geq t_{n-1,j} + h_{min,VC}\left(v_{n-1,j}(v_s,v_e),v_{n,i}(v_0,v_1)\right) + \tau_{comm} + \max(0,\tau_{ctrl,n} - \tau_{ctrl,n-1}) - \tau_{max}(3 - z_{n,i,v_0,v_1} - z_{n-1,j,v_s,v_e} - z_{n-1,n}) \quad (8.VC)$$

The minimum headway constraints are formulated as Big-M constraints that are binding only if train $n-1$ is ordered before train n at the junction ($z_{n-1,n} = 1$), and the two running arcs are both chosen ($z_{n,i,v_0,v_1} = 1$ and $z_{n-1,j,v_s,v_e} = 1$). This multiple Big-M formulation is permissible because of constraint (6), which require all event times to be nonnegative and less than τ_{max}. This ensures that having all three binary variables equal to zero produces the same solution space as when only one or two are equal to zero. This multiple Big-M formulation is preferred to conditional programming because it requires fewer binary variables and constraints.

3.4 Constraints Establishing the Cyclic Timetable

The model of the cyclic timetable is created by selecting a single train $n \in T$, which has a copy n_{last} created to represent the first train of the next cycle. The run time constraints and headway constraints with other trains are created for n_{last} as for any other train $m \in T$. Constraint (9) established train n as the first in the order, and constraint (10) establishes n_{last} as the last in the order. Constraint (11) establishes that trains n and n_{last} have the same trajectory, with all event times offset by the cycle time t_{cycle}.

$$z_{n,m} = 1\ \forall m \in T, \mathrm{m} \neq n: n\ is\ first\ train \quad (9)$$

$$z_{m,n_{last}} = 1\ \forall m \in T \quad (10)$$

$$t_{n_{last},i} = t_{n,i} + t_{cycle}\ \forall i \in B_n \quad (11)$$

3.5 Computational Refinements for Moving Block and Virtual Coupling

The model formulation explained thus far is sufficient to construct a feasible timetabling model for Moving Block or Virtual Coupling; however, contains many Big-M constraints that will be non-binding in the optimal solution. The three Big-M terms per constraint results in a weak relaxation, increasing the difficulty of proving optimality. To improve computational performance, the model is refined by removing redundant constraints, which forms the basis of a row generation strategy.

The expression of the moving block separation distance between leader $n-1$ and follower n is:

$$JE_{n-1,n}^{n-1}(t) - JS_n(t) = s_{n-1}(t) - \frac{\left(v_n(t)\right)^2}{2b_{s,n}} - s_n(t) + C$$

where C is the total of all the constant terms in $JE_{n-1,n}^{n-1}(t) - JS_n(t)$, which are not explicitly shown here to improve readability. The derivative with respect to time is:

$$\frac{\partial(JE_{n-1,n}^{n-1}(t) - JS_n(t))}{\partial t} = v_{n-1}(t) - \frac{v_n(t)a_n(t)}{b_{s,n}} - v_n(t)$$

During the period of simultaneous running of arcs $v_{n,i}(v_0,v_1) \in V_{n,i}$ and $v_{n-1,j}(v_s,v_e) \in V_{n-1,j}$, the acceleration of both trains is constant with $a_n(t) = a_{n,i}(v_0,v_1)$ and $a_{n-1}(t) = a_{n-1,j}(v_s,v_e)$, so the only variables in the equation are $v_{n-1}(t)$ and $v_n(t)$, which are constrained between their starting and ending speeds. If both trains are cruising, the derivative will be the same throughout the common running interval. If one of the two trains is accelerating or decelerating, the combination of speeds of both trains at a given time t_0 is a function headway between $t_{n,i}$ and $t_{n-1,j}$, as speed $v_n(t_0)$ or $v_{n-1}(t_0)$ can only be attained once during train's arc. If the solution to:

$$\min_{v_{n-1}(t),v_n(t)} \frac{\partial(JE_{n-1,n}^{n-1}(t) - JS_n(t))}{\partial t} \ s.t.$$

$$\min(v_0, v_1) \le v_n(t) \le \max(v_0, v_1) \qquad (A)$$

$$\min(v_s, v_e) \le v_{n-1}(t) \le \max(v_s, v_e) \qquad (B)$$

is non-negative, then the separation distance more than the safety distance requirement for moving block will never be decreasing in the simultaneous running interval. Thus, the local minimum will always be at the start of the simultaneous running interval, which is already checked for feasibility as part of the constraint for the previous arc pair. As a result, for modelling Moving Block operations, the constraint between arcs $v_{n,i}(v_0, v_1), i > 0$ and $v_{n-1,j}(v_s, v_e), j > 0$ is redundant and not needed to guarantee feasibility. In the case where one of the two trains is entering the network ($i = 0$ or $j = 0$), there are no constraints guaranteeing feasibility at the start of the simultaneous running period, so the constraint (if there is one) must be included.

For Virtual Coupling operations with the dynamic safety margin, Moving Block running is sometimes more permissive, so it is necessary to show the Moving Block identity above in addition to the relative braking distance requirement with derivative:

$$\frac{\partial(JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t)\ )}{\partial t} = v_{n-1}(t) - \frac{v_n(t)a_{n,i}(v_0,v_1)}{b_{s,n}} + \frac{v_{n-1}(t)a_{n-1,j}(v_s,v_e),}{b_{s,n}} - v_n(t)$$

and the emergency braking requirement with derivative:

$$\frac{\partial\left(JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t)\ \right)}{\partial t} = v_{n-1}(t) - 2\frac{v_n(t)a_{n,i}(v_0,v_1)}{b_{s,n}} + \left(\frac{v_{n-1}(t)a_{n-1,j}(v_s,v_e)}{b_{s,n}} + \frac{v_{n-1}(t)a_{n-1,j}(v_s,v_e)}{b_{e,n-1}}\right) - v_n(t)$$

If the Moving Block derivative is always non-negative, and one can also show:

$$\min_{v_{n-1}(t),v_n(t)} \frac{\partial(JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,RBD}(t)\ )}{\partial t} \ \text{s.t.}\ (A), (B) \ge 0, \text{ and}$$

$$\min_{v_{n-1}(t),v_n(t)} \frac{\partial(JE_{n-1,n}^{n-1}(t) - JS_{n-1,n}^{n,Emerg}(t)\ )}{\partial t} \ \text{s.t.}\ (A), (B) \ge 0,$$

Thus, the constraint between arcs $v_{n,i}(v_0, v_1), i > 0$ and $v_{n-1,j}(v_s, v_e), j > 0$ is redundant and not needed to guarantee feasibility under VC operations. The implication of these identities is that headway constraints in the decoupling phase (other than the absolute stopping distance requirement for the diverging interlocking), and in periods of coupled running where both trains are cruising at the same speed are never more restrictive than those for the coupling phase or periods of coupled running where both trains are accelerating or decelerating.

The scheduling problem for Moving Block and Virtual Coupling is solved through a row-generation process that prioritizes feasibility in interlocking and station areas. The initial relaxed problem is constructed with all running arcs, merging & diverging interlocking constraints, and stopped train constraints (for stations served by the leader over open track), and open track running constraints only for arc pairs where the leading train is passing through an interlocking area (where the route of a third train merges or diverges). All open track constraints for the intervals where the leading train is passing through an interlocking area are included at this stage, even if they are redundant based on the identity shown above. The inclusion of these redundant constraint prevents trains from achieving unreasonably low headways at bottlenecks, thereby increasing the likelihood of finding a good train order. Macroscopic headway constraints (12) and (13) are also added to prevent unrealistically low headways at locations where multiple trains enter ($t_{n,0}$ for train $n \in T$) or exit the model ($t_{n,ex}$ for train $n \in T$). The macroscopic minimum headways $h_{m,en}(n-1,n)$ and $h_{m,ex}(n-1,n)$ should be set high enough to prevent unrealistically low headways, but low enough that they are never more restrictive than the microscopic minimum headway for the location.

$$t_{n,0} \geq t_{n-1,0} + h_{m,en}(n-1,n) - \tau_{max}(1 - z_{n-1,n}) \; \forall n-1, n \in T \; s_{n-1,0} \; same \; location \; as \; s_{n,0} \tag{12}$$

$$t_{n,ex} \geq t_{n-1,ex} + h_{m,ex}(n-1,n) - \tau_{max}(1 - z_{n-1,n}) \; \forall n-1, n \in T \; s_{n-1,ex} \; same \; location \; as \; s_{n,ex} \tag{13}$$

Once a set of feasible routes has been found for the interlocking areas, the open track sections for each pair of trains n and $n-1$, (where $n-1$ was ordered first) are checked for feasibility in the areas immediately after the merging interlocking for the two trains, and after any station served by $n-1$ in open track. These are the locations where coupling could take place, and thus the most likely locations for a violation of a non-redundant open track constraint. If there is a violated non-redundant headway constraint between chosen arcs $v_{n,i}(v_0, v_1)$ and $v_{n-1,j}(v_s, v_e)$, then all non-redundant open track constraints with running arcs of train $n-1$ that start at node j are added to the model. If there are more than three arcs of leading train $n-1$ with a violated headway constraint with arcs of follower n, only the violated constraints for the first three arcs after the most recent interlocking/station area are added in. This limits the number of added constraints in each iteration, allowing for adjustment to occur incrementally. The solving process for each iteration is limited to 10 seconds, given that the weak relaxation of the headway constraints or existence of symmetries could increase the time it takes to prove optimality. The process is repeated until a feasible timetable is produced.

4. Case Study

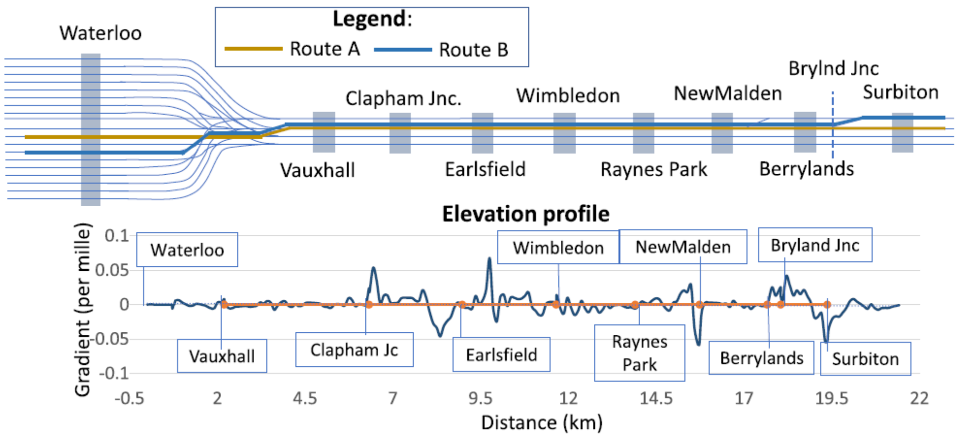

Figure 6: Case Study Route from Waterloo to Surbiton (Quaglietta et. al, 2020)

The Optimization model is used to determine a schedule for services on the Down Main Fast (DMF) line of the South West Main Line between London Waterloo and Woking. This line carries fast services from Waterloo toward Portsmouth and Southampton and has a zone-based service plan: trains generally operate non-stop from Waterloo or Clapham Junction to an outlying station, after which they make all stops. The analysis focuses on the bottleneck at Berrylands Junction (Brylnd Jnc), where some fast services switch to the Down Main Slow (DMS, Route B) to access the New Guildford Line past Surbiton, or to make stops on the SWML before Woking. The ability to form convoys of trains with a common route at this location could increase capacity versus plain MB.

The assessment of capacity gains of Virtual Coupling over plain Moving Block is performed for a scenario with a heterogenous stopping pattern, where speed and interlocking restrictions around Waterloo have less impact on capacity. The 15mph speed limit at Waterloo makes it is unlikely that Virtual Coupling will deliver capacity benefits on the line if this location is the bottleneck limiting throughput. Past Clapham Junction, the ability to operate at relative braking distance is expected to deliver capacity improvement.

The assessment is performed for a cycle containing nine services originating out of Waterloo. All services are pathed for a 12-car composition of the British Rail (BR) Class 450 EMU. A summary of the services, including their route and scheduled station stops, is shown in Table. For this analysis, one of the Route B services that calls at all stations between Surbiton and Woking is chosen to be the first train of the cycle.

Table 1: Summary of Train Path Requests for each cycle

Route at Berrylands	Scheduled Stops	Number of Services
A	Waterloo, Clapham Junction, Woking	2
	Waterloo, Clapham Junction	2
	Waterloo, Woking	2
B	Waterloo, Surbiton	1
	Waterloo, Surbiton, all stops from Surbiton to Woking	2

The paths of each train are modelled with behavioural nodes at the clear events for all interlockings, at scheduled station stops, and at locations not greater than 500m apart over open track. In the Waterloo station area, all trains operate at 20km/h until they reach the location where the speed limit increases (around kilometre 0.80). From there to kilometre 6, trains are allowed to choose speeds in increments of 10km/h, between 50km/h and 110km/h. From Kilometre 6 to 8, trains can operate at 100km/h, 110/km/h, or 115km/h to enable coupling with trains departing Clapham Junction. Past kilometre 8, trains can cruise at 110km/h, or at 115km/h to reduce separation with their leader. In this area, it is not possible to accelerate from 110 to 115 km/h: the provision approach speed of 115km/h is included only to allow trains to reduce their following distance before Berrylands junction. For trains taking diverging route B at Berrylands Junction, where there is a 60km/h speed restriction, it is possible to decelerate to this speed after kilometre 16. Past kilometre 18, trains can accelerate to 120km/h or 130km/h.

The optimization problem, including route and rolling stock information is constructed from the EGTrain model of the South West Main Line. This is the network used in Quaglietta et. al (2020) and

Quaglietta et. al (2022) to simulate VC operations. The minimum headway calculator for both Moving Block and Virtual Coupling was implemented in Python and the commercial optimization tool Gurobi is used to solve the MILP problems. The macroscopic minimum headway used for constraints (12) and (13) is 20 seconds, which is 3 seconds less than the lowest average headway reported for VC in Quaglietta et. al (2022). The optimization process was performed on a computer with Windows 11, with a 2.30GHz Intel Core i7 processor, and 32GB of RAM. The computation times reported include the times required to construct all running arc constraints, compute minimum headways to check feasibility and to create needed constraints, and MILP solving time.

5. Results

The line plans produced for the nine-train cycle in plain Moving Block operations and Virtual Coupling with the dynamic safety margin are shown below. Trajectories with solid lines indicate that the train is on the Down Main Fast Line. Dashed-line trajectories indicate the train is on another line.

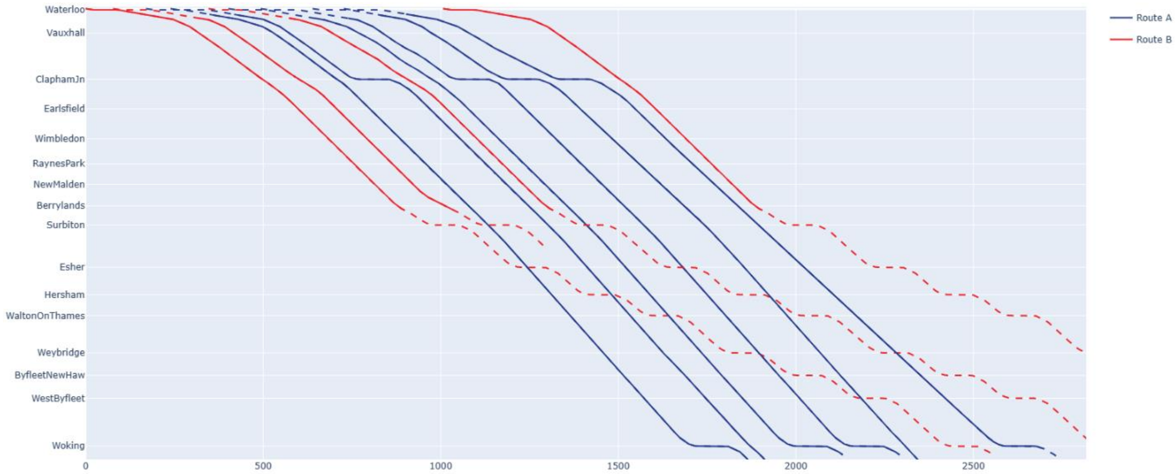

Figure 7: Optimal line plan of the nine-train cycle in Moving Block (Cycle Time 1007s)

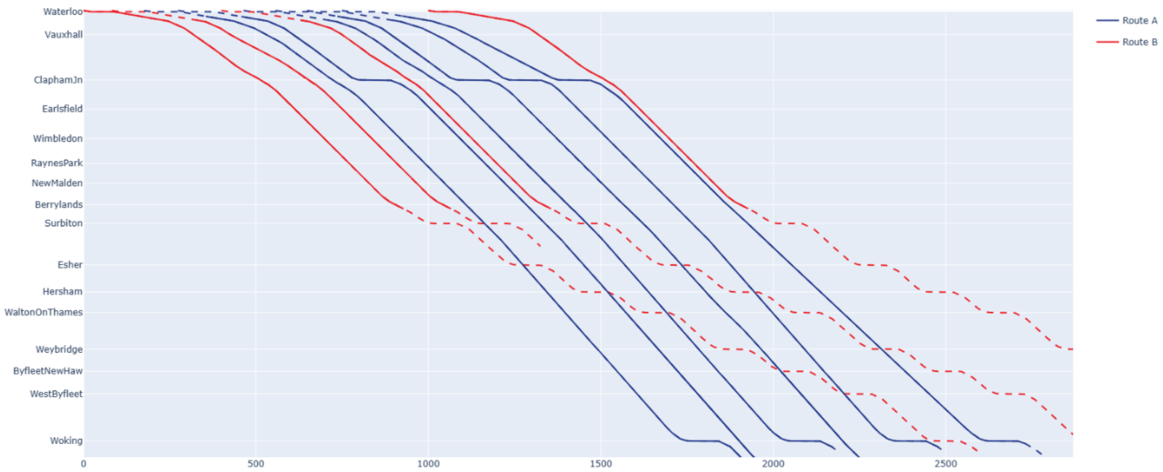

Figure 8: Optimal line plan of the nine-train cycle with Virtual Coupling (Cycle Time 1000s)

The ability to operate at relative braking distance plus a dynamic safety margin yields a small decrease in the cycle time of 7 seconds (1000s for Virtual Coupling vs 1007s for Moving Block), with three convoys scheduled. Two convoys were formed between a leader on Route A, and a follower on Route B. The First convoy, which is formed between the third and fourth train in the cycle, has a headway of 34s. The other convoy is formed between the last train in the cycle and the first train in the next cycle (the last two trains shown in Figure 8). This convoy has a headway of 25s. The need for Route B trains to decelerate to the 60km/h speed limit over the diverging switch may mitigate the capacity loss from having to brake to decouple.

All convoys are formed between a leader accelerating from standstill at Clapham Junction and a follower that does not call at the station. For these convoys, whose formation starts with the leader at standstill and the follower at line speed, the trains must remain in the Moving Block running state until the leader attains a high enough speed for the Virtual Coupling safety distance to be shorter than the Moving Block safety distance. Once this state is reached, the trains can form a capacity-effective convoy. The ability to form convoys at higher speeds reduces the cycle time, but the delay in transitioning to the coupling state (due to the presence of two braking distance terms in the VC safety distance) is a limiting factor.

The speed and interlocking constraints in the Waterloo station area prevent the formation of capacity-effective convoys of trains that have just left the terminal. In Quaglietta et al. (2022), which sought to evaluate the safety of VC with the dynamic safety margin, convoys were formed by limiting the speed of the leading train to 65km/h until it reached the Clapham Junction area. In practice, these types of trajectories will not improve capacity unless they delay the time the train starts to occupy a critical location downstream. A relevant downstream bottleneck is not present on the SWML, which has 4 tracks and no level junctions from Waterloo to Woking, but may be present in other networks.

In this case study, the zonal nature of the line plan limits the ability of Virtual Coupling to improve capacity versus plain Moving Block. In the SWML timetable, most fast services operate nonstop from Waterloo or Clapham Junction to an outlying zone, where they are then routed onto the Slow line to make all stops. For any train order in this line plan, the optimal train trajectory is one where trains operate as fast as possible to the interlocking where they switch to the Slow line. From Waterloo to Woking Station, the line does not have any level junctions, timed overtakes or single-track running, where closer running of trains with the same route could reduce occupation time at critical interlockings. More research is needed to assess the impact of network topology on Virtual Coupling capacity.

Computational results are shown in the table below for the row-generation method proposed in section 3.5. Computational time for both plain Moving Block and Virtual Coupling are at reasonable levels, but the VC method has higher computation time and requires more iterations than moving block. The Virtual Coupling model had more difficulty proving optimality. Out of ten trials, the VC model converged to a zero gap only three times, whereas the plain Moving Block model proved optimality in most of its trials. These differences are likely attributable to two factors: the VC solution space is larger than for MB, which could increase the time and number of iterations required to prove optimality, and the process of calculating a VC minimum headway has more steps than the comparable MB calculation. In any case, computation times are low enough that the method can be used for scheduling in real-world applications, and both models can reliably find good solutions.

Table 2: Results for Optimization with plain Moving Block Signalling

Trial	Cycle Time	Best Dual Objective	Computation Time	Number of Iterations
1	1007s	993s	446s	9
2	1007s	1007s	405s	7
3	1007s	1007s	372s	7
4	1007s	1007s	371s	7
5	1007s	1007s	367s	7
6	1007s	1007s	356s	7
7	1007s	1007s	401s	8
8	1007s	1007s	401s	9
9	1007s	993s	372s	6
10	1007s	1007s	673s	8

Table 3: Results for Optimization with Virtual Coupling Signalling

Trial	Cycle Time	Best Dual Objective	Computation Time	Number of Iterations
1	1000s	997s	437s	10
2	1000s	998s	423s	9
3	1000s	1000s	447s	9
4	1000s	997s	445s	10
5	1000s	997s	434s	10
6	1000s	997s	441s	10
7	1000s	1000s	413s	10
8	1000s	1000s	442s	12
9	1000s	992s	443s	11
10	1000s	997s	507s	14

6. Conclusion

In this paper, novel methods are proposed for translating the speed-dependent safety constraints present in distance-to-go signalling systems into MILP constraints that guarantee conflict-free operations. The proposed methodology for plain Moving Block and Virtual Coupling is also the first methodology that can translate the continuous track release process present in both systems into MILP constraints without having to discretize the open track into blocks or detection sections. The methodology for Virtual Coupling is flexible enough to accommodate possible future changes to the safety margin specification if future research identifies additional requirements. This methodology is used to construct alternative graph models for Moving Block and Virtual Coupling to minimize the cycle time of the Down Fast Line of the South West Main Line in the UK, by selecting the order of departures from the terminal at Waterloo. The results show that Virtual Coupling has potential to improve capacity compared to plain Moving Block at the planning level, but that the amount of improvement may depend on the network topology.

Additional research is needed on the impact of network topology and service plans on the capacity of Virtual Coupling at the planning level, with the aim of identifying real-life use cases. Future research should also explore application of the continuous-space AG formulation to real-time rescheduling, which is necessary to assess the performance of Virtual Coupling in disturbed operational scenarios.

7. References

Cheng, F., Su, S., Tang, T., & Wang, W. (2021). Optimization of the coupling operation for train platoon in the junction area. Paper presented at the *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, , 2021-September* 2405-2410. doi:10.1109/ITSC48978.2021.9565065

Di Meo, C., Di Vaio, M., Flammini, F., Nardone, R., Santini, S., & Vittorini, V. (2020). ERTMS/ETCS virtual coupling: Proof of concept and numerical analysis. *IEEE Transactions on Intelligent Transportation Systems, 21*(6), 2545-2556. doi:10.1109/TITS.2019.2920290

Felez, J., Kim, Y., & Borrelli, F. (2019). A model predictive control approach for virtual coupling in railways. *IEEE Transactions on Intelligent Transportation Systems, 20*(7), 2728-2739. doi:10.1109/TITS.2019.2914910

Gallo, F., Di Febbraro, A., Giglio, D., & Sacco, N. (2020). A mathematical programming model for the management of carriages in virtually coupled trains. Paper presented at the *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020,* doi:10.1109/ITSC45102.2020.9294277

Gallo, F., Di Febbraro, A., Giglio, D., & Sacco, N. (2021). Planning and optimization of passenger railway services with virtually coupled trains. Paper presented at the *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2021,* doi:10.1109/MT-ITS49943.2021.9529318

M. Janssens (2022). *Multi machine approaches for conflict resolution under moving block signalling*

Li, Y., Zhu, L., Wang, X., Zhao, H., & Gao, X. (2021). Metro train scheduling optimization for virtual coupling based train control. Paper presented at the *Proceeding – 2021 China Automation Congress, CAC 2021,* 1560-1565. Doi:10.1109/CAC53003.2021.9728666

F. Liu, J. Xun, H. Dong, Z. Zhang and Y. Li, "A Bi-level Real-time Rescheduling Approach for Train Operation in High-speed Railways," *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 2417-2422, doi: 10.1109/ITSC48978.2021.9564832.

Luo, X., Liu, H., Zhang, L., & Xun, J. (2021). A model predictive control based inter-station driving strategy for virtual coupling trains in railway system. Paper presented at the *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, , 2021-September* 3927-3932. doi:10.1109/ITSC48978.2021.956489

S. Long, L. Meng, Y. Wang, J. Miao, T. Li and F. Corman, "A Train Rescheduling Optimization model with considering the Train Control for A High-Speed Railway Line under Temporary Speed Restriction," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 2809-2816, doi: 10.1109/ITSC.2019.8917173.

R. Miao (2021). *A capacity-effective scheduling approach for Virtual Coupling operations.*

Quaglietta, E., Wang, M., & Goverde, R. M. P. (2020). A multi-state train-following model for the analysis of virtual coupling railway operations. *Journal of Rail Transport Planning and Management, 15* doi:10.1016/j.jrtpm.2020.100195

Quaglietta, E., Spartalis, P., Wang, M., Goverde, R. M. P., & van Koningsbruggen, P. (2022). Modelling and analysis of virtual coupling with dynamic safety margin considering risk factors in railway operations. *Journal of Rail Transport Planning and Management, 22* doi:10.1016/j.jrtpm.2022.100313

Schlechte, T., Borndörfer, R., Denißen, J., Heller, S., Klug, T., Küpper, M., Lindner, N., Reuther, M., Söhlke, A., & Steadman, W. (2022). Timetable optimization for a moving block system. *Journal of Rail Transport Planning & Management*, *22*, 100315. https://doi.org/10.1016/j.jrtpm.2022.100315

Su, H., Chai, M., Chen, L., & Lv, J. (2021). Deep learning-based model predictive control for virtual coupling railways operation. Paper presented at the *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2021-September* 3490-3495. doi:10.1109/ITSC48978.2021.9564411

Xu, P., Corman, F., Peng, Q., & Luan, X. (2017). A train rescheduling model integrating speed management during disruptions of high-speed traffic under a quasi-moving block system. *Transportation Research. Part B: Methodological*, *104*, 638-666. https://doi.org/10.1016/j.trb.2017.05.008