Università degli Studi di Roma Tre
Dipartimento di Informatica e Automazione
Via della Vasca Navale, 79 – 00146 Roma, Italy

Rescheduling models for network-wide railway traffic management

Pavle Kecman[1], Francesco Corman[2], Andrea D'Ariano[3], Rob M.P. Goverde[1]

RT-DIA-194-2012 **Giugno 2012**

(1) Department of Transport and Planning, Delft University of Technology
Stevinweg, 1
2628 CN Delft, The Netherlands.

(2) Centre for Industrial Management, Katholieke Universiteit Leuven,
Celestijnenlaan, 300A
3001 Heverlee, Belgium.

(3) Università degli Studi Roma Tre,
Via della Vasca Navale, 79
00146 Roma, Italy.

This work is partially funded by the Dutch Technology Foundation STW, research project: Model-Predictive Railway Traffic Management (project no. 11025).

ABSTRACT

In the last decades of railway operations research, microscopic models have been intensively studied to support traffic operators in managing their dispatching areas. However, those models result in long computation times for large and highly utilized networks. The problem of controlling country-wide traffic is still open since the coordination of local areas is hard to tackle in short time and there are multiple interdependencies between trains across the whole network. This work is dedicated to the development of new macroscopic models that are able to incorporate traffic management decisions. Objective of this paper is to investigate how different level of detail and number of operational constraints may affect the applicability of models for network-wide rescheduling in terms of quality of solutions and computation time. We present four different macroscopic models and test them on the Dutch national timetable. The macroscopic models are compared with a state-of-the-art microscopic model. Trade-off between computation time and solution quality is discussed on various disturbed traffic conditions.

Keywords: *Alternative Graph, Delay Propagation, Macroscopic Modeling, Railway Traffic Management, Timed Event Graph*

1 Introduction

Railway traffic usually operates according to a timetable. Disturbances originating from external factors (weather, number of passengers and their behavior, etc.) as well as from internal entities from within a railway system (reliability of infrastructure and vehicle equipment, behavior of personnel, etc.) create primary delays, deviations from the operational plans defined in the timetable. In heavily utilized networks, a deviation of one train from its schedule can affect other trains in the network and create secondary delay chains in a domino effect [7, 9]. Therefore, decisions are made in order to minimize the possible effect of those deviations on the system, both in the stage of timetable construction and in real-time during railway operation.

In the process of timetabling it is crucial to focus on the robustness of the timetable, i.e., its ability to resist and adapt to minor disturbances. For that reason, running time supplements and buffer times are introduced in order to enable trains to make up for their delay and to avoid creating secondary delays. However, both running time supplements and buffer times have to be limited by capacity consumption constraints. Finally, neither of them is meant to compensate for major disruptions such as accidents, infrastructure or vehicle equipment failures, etc. Therefore, good timetabling can only to certain extent contribute to punctuality of the railway traffic.

Dynamic traffic management is necessary as a complementary real-time direction to maintain the punctuality of railway operations [5]. The concept of dynamic traffic management has so far been widely understood as a reactive set of actions with the purpose of minimizing the consequences of actual delays. It is performed by traffic control centers in two levels. The tactical level (regional or network controllers) comprises the supervision of the state of traffic on a network level, detection of deviations from the timetable, resolution of conflicts affecting the overall network performance, handling failures and events that may have big impact on performance indicators, etc. The operational level consists of local traffic controllers (in major stations with a complex topology of interlocking areas) or centers for remote control (for multiple small stations with a simple topology and possible points of conflict between major stations such as junctions, movable bridges, level crossings etc.) with the task to perform all safety related actions, set routes for trains, predict and solve conflicts on a local level and control processes that take place on the part of infrastructure under their supervision.

Current practice in operational control of disruptions and delays still relies predominantly on predetermined rules and the experience and skills of personnel. Neither local nor network traffic controllers have a reliable supporting tool to predict the effect of their decisions and evaluate them, which often leads to creating new conflicts and suboptimal effects on the network level. We therefore aim at developing a global, network scale optimization tool that optimizes the actual state over the overall network and controls the traffic from a global perspective with adjustments to the timetable.

In this paper, we examine the applicability of macroscopic models for rescheduling railway traffic at a network-wide level. Railway traffic is represented by a timed event graph that allows computing delay propagation in large and strongly interconnected networks in a short time [9]. The timed event graph is then converted to four alternative graph models with different number of included operational constraints. An efficient solution algorithm [6] is applied on the alternative graph models to optimize the rescheduling actions.

All presented models have been tested on a series of delay scenarios, compared to each other and evaluated by comparison to the microscopic model of D'Ariano [5], that takes into account detailed infrastructure data and train dynamics. All comparisons have been performed both in terms of resulting secondary delay and dispatching decisions on a case study of the corridor between Utrecht and Den Bosch in the Netherlands. Furthermore, the macroscopic models have been applied on a test case of one peak hour of the Dutch national timetable in order to test their applicability on large and busy networks.

An important objective of this work is to analyze the compromise between precise modeling of railway capacity constraints and a reasonable time to compute the alternative solutions for the large scale railway traffic management instances. A suitable choice of the granularity of the macroscopic model is crucial in order to find the balance between aggregating constraints and limiting the problem complexity.

The next section gives an overview of existing models for railway operations. Section 3 defines the problem tackled by the models presented in this paper. Section 4 describes the general approach to macroscopic modeling of railway operations as well as the procedure used to solve the rescheduling problem and obtain the new schedule. Section 5 gives a specific description of presented models. Sections 6 and 7 report on a comparison of the models on a railway corridor and on the whole Dutch network, respectively. Finally, we discuss the performance of each model and give directions for future research.

2 Literature review

Recent contributions in the field of railway operations research and modeling follow mainly two categories: (i) off-line analysis including timetabling, timetable evaluation and performance analysis, (ii) real-time traffic management and rescheduling. In this review we will concentrate on models from the latter group and those from the former that capture the complex interdependencies between train services and are thus able to accurately compute delay propagation and predict the consequences of disruptions.

Another major criterion for partitioning the area of modeling railway traffic is the level of detail considered in the models. We distinguish between microscopic models, which consider detailed description of infrastructure and train dynamics, and macroscopic models with a higher level of aggregation.

A model for delay propagation should, as a supporting tool for traffic controllers, give accurate forecasts of conflicts (on both global and local level) resulting from the detected deviations and disruptions in traffic and infrastructure. Moreover, such model has the task to estimate the effect and evaluate the quality of the potential control decisions. Nash & Huerlimann [16] and Siefer & Radtke [19] presented advanced microscopic simulation tools, able to accurately simulate railway operations based on a detailed modeling of infrastructure, signaling, rolling stock characteristics, train dynamics and the timetable. However, using microscopic models to capture the structure and processes on large, complex and heavily utilized railway networks can result in long computation times. Therefore, such models are not appropriate for real-time applications on the level of multiple dispatching areas.

Goverde [9, 10] introduced a macroscopic model of train delay propagation based on timed event graphs and max-plus algebra. Railway traffic is modeled by events and processes that specify precedence relations between them. The model allows fast computation of performance indicators in a short time even for large networks.

The limitation of conventional max-plus models is the assumption of a fixed structure, i.e., fixed train orders, sequences, and routes. They can therefore not be used to model dispatching actions which may decrease and prevent delay propagation, such as changing the order of trains, canceling a train or a connection. Van den Boom & De Schutter [23] proposed an approach called switching max-plus linear systems that can be used to incorporate discrete dispatching actions into the max-plus framework. In their approach, the structure of the timed event graph can be changed. Every change corresponds to a dispatching decision and results in a new structure (mode) which represents a railway traffic model with the specified order of events and synchronization constraints. The system is managed by switching between different modes, thus allowing the inclusion of discrete decisions into the model. They recast the optimal switching problem as a mixed integer linear programming (MILP) problem and propose commercial software or meta heuristic algorithms to obtain solutions.

Berger et al. [1] incorporated stochasticity in their graph-based macroscopic model for delay prediction. By using the set of waiting policies for passenger connections and assuming discrete distributions of process times, they are able to estimate delay propagation over the network.

The two major issues for developing decision support systems for traffic controllers and dispatchers are: (i) combinatorial complexity of the rescheduling problem, (ii) requirement for short computation time because new schedules should not be outdated by the time they are produced. Therefore, the majority of approaches to solve the rescheduling problem rely on macroscopic modeling of railway operations and often employ (meta) heuristic methods to tackle the problem.

Tomii et al. [20] used simulated annealing to solve the train rescheduling problem posed as a project network PERT. Schöbel [18] used a similar graph based interpretation of railway operations to optimize the solution to delay management problem by minimizing passenger delay using integer programming. The model was further extended by including headways and capacity constraints and testing multiple preprocessing heuristics in order to fix integer variables and speed up the computation [17].

Törnquist & Persson [21] formulated the macroscopic dispatching (rescheduling and rerouting) problem as MILP and applied different heuristics to reduce computation time depending on the size of the instance. Törnquist recently applied the greedy heuristic approach on the same formulation of the problem [22]. The idea was to obtain reasonably good feasible solutions in a very short time and use the rest of the predefined computation time to try to improve it by backtracking and reversing decisions made in the first stage.

Min et al. [15] describe application of a column generation algorithm for train conflict resolution. They define a macroscopic model of railway operations and introduce a number of restrictive assumptions, thus enabling the separation of the problem. They prove that solving small separate problems in topological order yields the global solution.

Jacobs [13] investigated application of microscopic simulation models for rescheduling. Asynchronous simulation was coupled with heuristics for conflict resolution which relies on hierarchical ordering of trains into categories and always giving priority to trains in higher category. The drawback of this approach is that the resulting schedule can be far from global optimum.

Recently developed railway traffic management tools based on microscopic models use decomposition of large networks to local areas and corridors [5]. Real-time computation time requirements prevent straightforward application of these models to large-scale networks of strongly interconnected lines.

Corman et al. [3] focused on enlarging the applicability of microscopic models to large networks by spatial decomposition into local areas and global coordination of dispatching actions. However, the problem of controlling country-wide traffic is still open since the coordination of local areas is hard to tackle in short time and there are multiple interdependencies between trains across the whole network.

Table 1 summarizes and classifies the mentioned publications according to the level of detail in the model (microscopic or macroscopic) and ability to model dispatching actions (delay propagation or rescheduling).

Table 1: Summary of railway scheduling models

	Delay propagation	Rescheduling
Macroscopic	[1, 9]	[15, 17, 18, 20, 21, 22, 23]
Microscopic	[16, 19]	[3, 5, 13]

The mentioned macroscopic models for rescheduling were tested mostly on subnetworks of a national network [23, 21, 22, 18, 17] or large urban networks [20, 15]. Therefore, their applicability for network-wide traffic control has not been tested.

In this paper we address the problem of real-time rescheduling on the level of national networks. We develop macroscopic models with different levels of abstraction and apply the solution procedure [6] that aims at finding the globally optimal schedule with minimum secondary delays.

3 Problem definition

We first define the terminology and operational constraints that will be considered in this paper:

- Timetable points - places on the railway network where: (i) some or all trains are scheduled to stop and operations with passengers or goods can be performed and (ii) no trains are scheduled to stop but for traffic safety reasons passing time of trains over these points is included in the timetable e.g. junctions, movable bridges
- Stations - timetable points that have enough tracks to facilitate overtaking/simultaneous dwelling of more than one train, in both directions;
- Stops - timetable points that do not have enough tracks to facilitate overtaking/simultaneous dwelling of more than one train, in any direction;
- Junctions - timetable points where two or more railway lines intersect or merge and no trains are scheduled to stop;
- Open track segment - track that connects two timetable points (timetable points can be connected by multiple open track segments, e.g. double or multi-track lines).

3.1 Macroscopic operational constraints

Capacity of stations can be defined by the number of platform tracks and capacity of stops and junctions is equal to the number of tracks on the line. Open track segments are divided into fixed block sections, which can contain only one train at a time. Unhindered running of two successive trains on the same open track segment is ensured by two blocks separation between trains [11]. For that reason, minimum headway times are introduced between successive departures (and arrivals) of trains that use the same open track segment. In real operations, trains cannot overtake each other on open track segments (FIFO property).

Single track segments or open track segments with bidirectional traffic are attributed by a constraint that all trains operating in one direction need to leave the segment before the train in the opposite direction can depart.

We make a distinction between two types of route conflicts and adopt terminology from Min et al. [15]. Intra-track conflicts arise between trains that use the same open track segment when the two block separation property is violated. Inter-track conflicts arise between trains that use different open track segments. Those conflicts can occur in block sections that contain switches where different open track segments are merging or intersecting (in stations or junctions).

The minimum connection times between the arrival of feeder trains and the departure of connecting trains in stations are defined by the timetable to enable passengers to transfer between different services and to ensure realization of planned rolling-stock and staff circulations. If the connection planned by the timetable is maintained, the connecting train cannot depart from a station until all feeder trains have arrived and the minimum connection times have passed. A connection can be canceled in order to reduce delay propagation from the connecting train to the feeder train. In this paper we consider all connections to be fixed as planned by the timetable.

Planned departure and arrival times are scheduled in the timetable. Trains can not depart before their scheduled departure time and are considered late if they arrive after their scheduled arrival time.

In case of primary delays, a planned timetable becomes infeasible, i.e. trains can not operate according to the times scheduled in the timetable. In busy networks delays propagate easily to other trains and can have consequences in a wide area of the railway network, due to operational constraints listed above. A new feasible schedule on the network-wide level with minimum deviation from the original needs to be produced in a reasonably short time. In this paper, we define a family of macroscopic models and investigate how increase in the number of considered operational constraints affects the computation time of the solution algorithms and consequently their applicability on heavily utilized networks.

4 Macroscopic modeling of railway operations

4.1 Timed event graphs

We model railway operations at the macroscopic level by means of timed event graphs (TEG), as formally defined in Goverde [9, 10]. A TEG is a representation of a discrete-event dynamic system which consists of events, connected by processes that are attributed by the minimum process times. In a timed event graph events are modeled by nodes and processes by arcs.

A TEG can be represented by a max-plus linear system. Efficient analytic methods and graph algorithms exist for analysis and computing various performance indicators of max-plus linear systems [9, 10].

An individual train run is modeled as a series of events and processes that connect them. Every node is an event, defined by the train number, the timetable point where it occurs, type (departure, arrival or through) and the scheduled event time. An arrival event is the moment when a train comes to a standstill at the platform track and a departure event is the moment when a train starts accelerating from the platform track after a scheduled stop. A through event is the moment when a train passes the center axis of a timetable point without stopping.

Every arc is a process, defined by the train number, type (run or dwell), starting and completion event, and the minimum process time.

Interactions between trains are modeled with headway and connection processes. Headway processes separate (by a minimum headway time) events of different trains that have identical, intersecting or merging routes. The minimum headway time between two trains is computed according to the blocking time theory [11]. Scheduled headway time is defined in the timetable and represents the minimum headway time extended by the buffer time (rounded up to the full minute if scheduled event times are in full minutes). All events in a TEG take place on the platform tracks or center axis of a timetable point. Since route conflicts occur at signals, that prevent a train from entering the occupied or reserved block, a minimum headway time needs to be computed between events in the station where conflicting outbound routes start and inbound routes end.

The connection processes separate the departure event of a connecting train and arrival events of each feeder train by a minimum connection time. The minimum connection time has to be long enough to allow passengers or staff that arrived by a feeder train to board the connecting train, or shunting movements to be performed in order to maintain a rolling-stock circulation plan.

An event in TEG can occur only after all processes represented by incoming arcs of the corresponding node have been completed. Events in a TEG occur in a fixed sequence determined by the topology of the graph.

The fixed structure of a timed event graph is a major obstacle for application in the field of real-time rescheduling since many dispatching decisions imply changes of relative order of occurrence among events. In this paper, we overcome this limitation by converting a TEG to an alternative graph [14].

4.2 Alternative graphs

An alternative graph (AG) is a representation of a job-shop scheduling model with additional operational constraints. On a microscopic level, the train rescheduling problem posed as a job-shop scheduling problem [5] is to schedule a finite set of jobs (trains), defined by fixed sequences of operations (train runs and dwellings) which cannot be interrupted, on a finite set of resources (block sections or platform tracks) that can perform one operation at a time (no-store or blocking constraint). The objective is to schedule all operations on corresponding resources and to minimize the waiting time, i.e. the time difference between the moment when an operation can start and the actual starting time, due to conflicting jobs on some resources.

We extend this model to a macroscopic scale by aggregating multiple block sections into open track segments and platform tracks into timetable points and use them as resources in macroscopic models. Therefore, the number of operations that can simultaneously be handled by one resource, depends on the capacity of that resource.

Alternative graphs consist of nodes N, fixed arcs F, and pairs of alternative arcs A. We add the connection arcs C to this generic formulation as in Corman et al. [2]. We use the following notation: M, O, T are sets of resources (machines), operations and trains (jobs), respectively; i, j are indices of resources $m_i, m_j \in M$; r, s, are indices of trains $t_r, t_s \in T$. We denote by x_i^r the starting time and by p_i^r the processing time of operation $o_i^r \in O$ of train t_r on resource m_i. The headway time between the starting times of operations of trains t_r and t_s on resource m_i is denoted by $h_i^{r,s}$. T_i is a set of trains that use resource m_i. L is a sufficiently large number (larger than the latest completion time of the latest operation).

A node in the graph represents a single operation $o_i^r \in O$ of job $t_r \in T$, that is performed on resource $m_i \in M$. Every node is attributed with the starting time x_i^r of the corresponding operation. Since one job consists of the predetermined sequence of operations, node x_i^r at the same time represents the completion time of the previous operation.

An arc $(x_i^r, x_j^s) \in \{F \cup A \cup C\}$ with weight p_i^r represents the precedence relation between operations o_i^r and o_j^s given by the following equation.

$$x_j^s \geq x_i^r + p_i^r, \ \forall i, j, r, s : m_i, m_j \in M, t_r, t_s \in T \tag{1}$$

Fixed arcs are used to model fixed precedence relations between operations that have to be performed in a fixed relative order.

Alternative arcs are decision variables used to determine the relative order of operations scheduled to be performed on the same resource. If operations o_i^r and o_i^s are scheduled to be performed on the same resource m_i, than the relative order of operations can be determined by selecting the appropriate alternative arc. The concept of alternative arcs can be modeled with the binary control variable $k_i^{r,s}$ such that:

$$k_i^{r,s} = \begin{cases} 1 & \text{if } x_i^r < x_i^s, \\ 0 & \text{otherwise} \end{cases} \ \forall i : m_i \in M, \forall r, s : t_r, t_s \in T_i \tag{2}$$

with the constraint that exactly one arc from each pair has to be selected:

$$k_i^{r,s} + k_i^{s,r} = 1, \ \ \forall i, r, s : m_i \in M, \ t_s, t_r \in T_i \tag{3}$$

Selection of exactly one arc from each pair is called a complete selection. The objective is to select alternative arcs in a way that would minimize the waiting time of all operations. A valid solution determines the precedence relations between each two operations that are scheduled on the same resource. Two basic properties need to be respected: (i) completeness (exactly one arc from each alternative pair is selected), (ii) consistency (it must not contain positive length cycles). A complete and consistent selection yields the full schedule of all jobs. A possible interpretation of a complete selection of an alternative graph is a mode in a switching max-plus linear system [23].

Connections can be represented by a constraint between events of different trains. We can define them independently of resource types where they occur. If there is a scheduled connection between a feeder train t_s and the connecting train t_r then:

$$x_i^r \geq x_j^s + c^{r,s}, \tag{4}$$

where x_i^r is the departure time of train t_r, x_j^s is the arrival time of train t_s, the arc $(x_j^s, x_i^r) \in C$ and $c^{r,s}$ is the minimum connection time.

Scheduled starting and completion times of operations (timetable constraints) are incorporated in an AG by means of two dummy operations (nodes) 0 and n, with starting times x_0 and x_n. If operation o_i^r is scheduled to start at time d_i^r, the fixed arc (x_0, x_i^r) with weight d_i^r (release time) is added to the graph to ensure that the operation can not start before its scheduled starting time.

In real operation, scheduled completion time α_i^r (due date) of operation o_i^r may become infeasible due to disturbances that can cause extension to the planned processing time p_i^r (primary delays). This delay can propagate over successive operations of the train t_r, thus making their due dates also infeasible. A modified due date can therefore be defined by $-\max(\alpha_i^r, \tau_i^r)$ where τ_i^r is the earliest possible completion time of operation o_i^r, considered isolated from interactions with all other operations not belonging to job t_r, that can not be improved by any rescheduling action. A fixed arc with the weight equal to the modified due date is added to the graph from the node that represents the completion time of the operation to node n. We define an unavoidable delay by $\max(0, \tau_i^r - \alpha_i^r)$, a secondary delay by $\max(0, x_{i+1}^r - \max(\alpha_i^r, \tau_i^r))$ and a total delay as the sum of the unavoidable and secondary delays.

D'Ariano [5] showed that minimization of the critical path between nodes 0 and n is equivalent to minimizing the maximum secondary delay over all operations. Thus the objective function can be formally expressed with:

$$\min \quad x_n - x_0 \tag{5}$$

The solution procedure determines the starting times x_i^r for every operation o_i^r, and values of binary variables $k_i^{r,s}$ representing the orders within each pair of trains $t_r, t_s \in T$ scheduled to use the same resource m_i. An exact search is performed in the solution space by means of a branch and bound algorithm, as follows. A good starting solution is found by a set of heuristics (*first come first served*, *first leave first served*, *avoid most critical completion time*). The solution procedure is truncated after a time limit is reached. A time limit of 5 minutes is used in this work. We refer to [8] for additional information on the solution procedure.

4.3 Conversion of TEG to AG

The macroscopic modeling of railway traffic by means of alternative graphs is explained with terminology introduced in Section 3. Each timetable point and each open track segment is modeled by a resource. An individual train run is modeled as a sequence of operations. Every operation is attributed by the starting time, duration and the resource traversed by the train. A train run is represented with nodes and fixed arcs.

In order to describe how a TEG is converted to an AG, the difference in meanings of nodes and arcs in these two graphs needs to be resolved. We will keep the interpretation of nodes and arcs as in TEG and convert it to AG in the following manner. Fixed arcs represent operations (run or dwell). Weight of each fixed arc is equal to the minimum processing time of the corresponding operation. Every node is an event - arrival or departure (through event is included by fixing the dwell time to 0), representing the start of the operation denoted by the outgoing fixed arc and completion of the operation denoted by the incoming fixed arc. Every departure after a scheduled stop is connected to node 0 and every arrival to station with a scheduled stop to node n, as explained in the previous section.

The advantage of using AG for macroscopic models is in modeling discrete decisions that manage interactions between trains. If two trains have operations that can not be simultaneously performed on the same resource with constrained capacity, at least one pair of alternative arcs weighted by the minimum headway time between two operations is added in order to specify the precedence relation between operations. The number of arc pairs and their starting and end nodes depend on the resource type. The order of operations is determined by selecting the appropriate arc. This way, resolution of intra-track conflicts (conflicts between trains using the same resource) can be appropriately modeled.

However, since generic alternative graphs can only model conflicts between operations taking place on the same resource, inter-track conflicts on a macroscopic level had to be modeled in a different manner since they represent conflicts between operations taking place on different resources (Section 5.3).

Connections in the macroscopic AG models are fixed and modeled in the same way as in timed event graphs. A connection arc is added between the node that models arrival event of the feeder train and the node that models departure event of the connecting train with the weight equal to the minimum connection time.

4.4 Resources as building blocks of alternative graphs

Infrastructure elements can be modeled by using resources with different properties in terms of capacity. This results in multiple models with different complexity and number of operational constraints. Before presenting a detailed description of the macroscopic models (Section 5), we first describe the essential meaning of each type of resource.

4.4.1 Infinite capacity resources (IC)

The simplest way of specifying a resource is by considering only the temporal duration of the scheduled operation. This means that no further restriction is posed on train orders and headways between trains. Therefore, these resources do not model interactions between trains and all trains can use them independently from each other. This resource type is used under assumption that capacity is sufficient to accommodate demand at all times, thus no conflict can occur and the only binding constraint is the processing time.

Figure 1 shows how two trains t_r and t_s are modeled on IC type resource m_i. Each node represents the starting time of an operation on the resource (x_i^r is a starting time of operation o_i^r of train t_r on resource m_i). Arcs represent the operation that started at their parent node and their weight is equal to the minimum processing time of operation (p_i^r is the minimum processing time of operation o_i^r). There are no arcs between operations associated with different trains in Figure 1, thus operations of both trains can be performed independently from each other. The only constraint that has to be respected when scheduling operations on this resource type is given by equation (1).

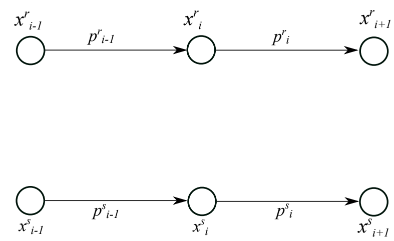

Figure 1: Graph representation of resources with infinite capacity

4.4.2 Infinite capacity resource with headway (IC+H)

If a resource is modeled as infinite capacity with headway, starting times of two consecutive operations o_i^r and o_i^s on the same resource m_i are separated by a time interval defined with headway $h_i^{r,s}$. Trains are thus prevented to occupy the same infrastructure element within a predefined headway time.

Introducing a minimum time separation between the starting times of two operations on a resource does not constrain capacity. Moreover, completion times of operations are not constrained by any headway. They can therefore occur simultaneously and not necessarily in the same relative order in which they started. Figure 2 depicts the alternative graph that can be used to compute the starting time of both operations o_i^r and o_i^s on resource (machine) m_i. Alternative arcs are shown with dashed lines with weights equal to the minimum headway time between two operations of different trains on the same resource. Note that in order to independently observe the properties of this type of resources, neighboring resources m_{i-1} and m_{i+1} are modeled as non-constrained infinite capacity resources.

If we define by $M_2 \subset M$ a set of machines of type IC+H, the starting time of an operation scheduled on this resource can fully be defined by equations (1)-(3) and the additional constraint:

$$x_i^r \geq x_i^s + h_i^{s,r} - \mathrm{L} \cdot k_i^{r,s}, \quad \forall i, r, s : m_i \in M_2,\ t_r, t_s \in T_i \tag{6}$$

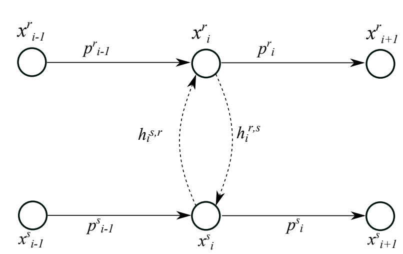

Figure 2: Graph representation of resources with infinite capacity and headway constraint

4.4.3 Infinite capacity resources with FIFO property (IC+FIFO)

This type of resource is an extension of the type infinite capacity with headway in the sense that an additional headway constraint is imposed on the completion times of operations. Note that capacity is still not restricted in this resource type. The graph depicting two operations on resource m_i of type IC+FIFO, is shown in Figure 3 (adjacent resources m_{i-1} and m_{i+1} are modeled as non-constrained infinite capacity resources).

In contrast to other resource types, two operations, performed on the same resource, are separated with two pairs of alternative arcs. Namely, alternative arc (x_i^r, x_i^s) that assigns precedence to start of operation o_i^r is paired with arc (x_{i+1}^s, x_{i+1}^r) that gives precedence to completion of operation o_i^s. In the same way, alternative arc (x_i^s, x_i^r) is paired with arc (x_{i+1}^r, x_{i+1}^s) (arcs belonging to one pair are shown in the same color). That way, both starting times and completion times of two operations are separated.

However, the increase in the number of arcs does not directly contribute to the increase in complexity since the selection of an arc from one pair implies the selection of an arc from the second pair, i.e. two pairs of alternative arcs represent only one decision variable (e.g. selection of arc (x_i^r, x_i^s) from the red pair implies the selection of arc (x_{i+1}^r, x_{i+1}^s) from the blue pair). Selection of arcs (one from each pair) that would violate the FIFO constraint would result in a positive length cycle, which is not permitted neither in TEG [10] nor in AG [8].

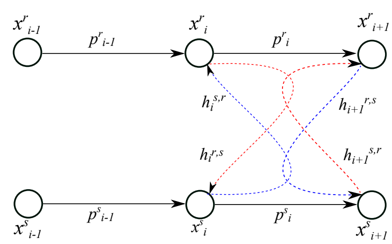

Figure 3: Graph representation of resources with infinite capacity and FIFO constraint

If $M_3 \subset M$ is a set of machines of type IC+FIFO, the starting time of an operation scheduled on those machines can fully be defined by constraints (1)-(3), (6) and the additional constraints:

$$x_{i+1}^r \geq x_{i+1}^s + h_{i+1}^{s,r} - \mathrm{L} \cdot k_{i+1}^{r,s}, \quad \forall i, r, s : m_i \in M_3, \; t_r, t_s \in T_i \tag{7}$$

$$k_i^{r,s} = k_{i+1}^{r,s}, \quad \forall i, r, s : m_i \in M_3, \; t_r, t_s \in T_i \tag{8}$$

4.4.4 Finite capacity resources (B)

The most restrictive resource type allows only one operation to be processed at the same time. Figure 4 shows that an operation that does not have precedence on the resource, can be initiated only after the preceding operation has been completed and the required headway time has passed.

If we define by $M_4 \subset M$ a set of machines of type IC+B, the starting time of an operation scheduled on those resources can fully be defined by constraints (1)–(3) and the additional constraint:

$$x_i^r \geq x_{i+1}^s + h_i^{s,r} - \mathrm{L} \cdot k_i^{r,s}, \quad \forall i, r, s : m_i \in M_4, \; t_r, t_s \in T_i \tag{9}$$

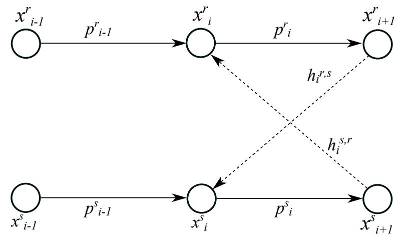

Figure 4: Graph representation of resources with finite capacity

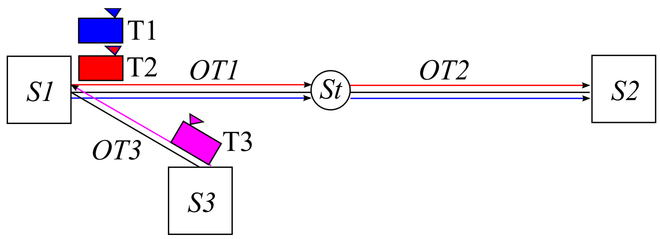

Figure 5: Layout of the illustrative example

5 Models examined

Description of the four rescheduling models for network-wide traffic management will be given in this section. The resources presented in the previous section will be used to model different infrastructure elements. All macroscopic models assume unidirectional traffic on double track lines. Bidirectional open track segments (single track line segments) are modeled with resource type B under assumption of low traffic volumes over such segments. That approach is *conservative* because it limits the capacity of the line segment to one train at a time which in reality is not the case for successive trains running in the same direction.

Macroscopic models will be described on an illustrative example shown in Figure 5. Infrastructure elements in the example are stations S1, S2 and S3, stop St and open track segments OT1, OT2 and OT3. Trains T1 and T2 run from S1 to S2 on open track segments OT1 and OT2. Train T1 has a scheduled stop at St. Train T3 runs from S3 to S1 on open track segment OT3. Routes of the three trains T1, T2 and T3 are presented in Figure 5 with arrows of corresponding colors.

Since trains T1 and T2 use the same open track segments, all potential conflicts between them can be characterized as intra-track conflicts. However, conflicts between the inbound route of train T3 and the outbound routes of trains T1 and T2 at station S1 are an example of inter-track conflicts.

Figures 6, 7, 9 and 10 present the alternative graphs for each described model. Every node is an operation of a train (defined by the color) on the specified resource. Dummy nodes (0 and n) incorporate timetable constraints in the model.

Fixed arcs are presented in colors that correspond to train colors from Figure 5. They are marked by type of operation: *run* or *dwell*. Train departure is modeled as a start of operation on the open track resource and train arrival as a start of operation on a timetable point resource. The outgoing fixed arcs from node 0 are weighted by the scheduled departure times (*SDT*). The incoming fixed arcs to node n are weighted by

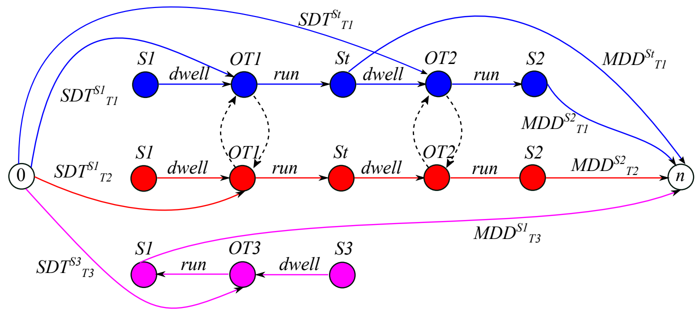

Figure 6: Illustrative example - Model 1

modified due dates (*MDD*) as explained is Section 4.2.

Alternative arcs are shown in dashed lines. For the sake of clarity their weights (minimum headway times) are not shown in the figures.

5.1 Model 1

This is the simplest macroscopic model considered in this paper. The AG of the illustrative example modeled by Model 1 is shown in Figure 6.

All timetable points are modeled as resources with infinite capacity and no constraints (resource type IC described in Section 4.4.1). This black-box approach to modeling stations relies on assumption that capacity of each station is at all times sufficient to satisfy demand.

Open track segments that connect stations are modeled with resource IC+H (Section 4.4.2). A pair of alternative arcs is added to ensure the time separation between starting times of two successive operations on the same open track resource (departures). However, headways between arrivals are not considered in this model and the order of arrivals is not implied by the order of departures.

Moreover, inter-track conflicts are not included as a constraint in this model which has a great level of idealization and its use can only be justified with low complexity and short computation time. Operational constraints considered in this model satisfy the requirements for modeling homogeneous traffic (all trains have equal speeds) on the line. In that case, trains are separated in time at the departure points and the model assumes fixed running times, thus arrival headways become redundant if trains have the same running time.

5.2 Model 2

We extend the previous model by considering arrival headway time and sequence of arrivals to a timetable point from the same open track segment. That is achieved by modeling open track segments with resource type IC+FIFO (Figure 7). This ability to model intra-track conflicts between trains with different speeds on the line, results in the increased size of alternative graph since the number of alternative arcs used to model train interactions on open track segments has doubled when compared with Model 1. However, the complexity of this model is not directly influenced by the increase of the size of the graph as shown in Section 4.4.3.

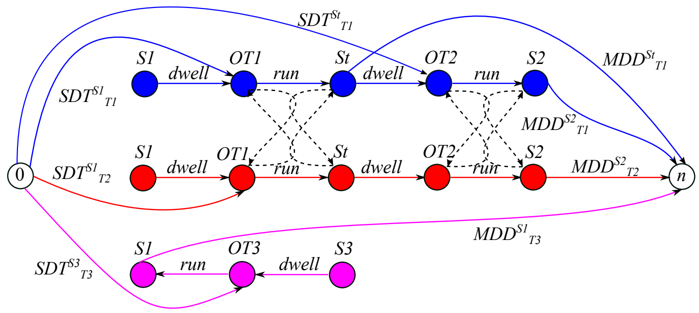

Figure 7: Illustrative example - Model 2

5.3 Model 3

None of the previously presented macroscopic models is able to capture potential inter-track conflicts. Since the two potentially conflicting operations are performed on different resources (different open track segments), capacity constraints associated with each resource are not able to model these conflicts.

In order to overcome this, we introduce an additional finite capacity resource with processing time 0. This resource does not have any physical interpretation (we therefore refer to it as a virtual resource) and its purpose is to separate in time events leading to inter-track conflicts.

If two trains with conflicting routes through a timetable point arrive to (depart from) the timetable point using different open track segments, we add the virtual resource to the path of each train. An inbound route is represented by an arrival event (the resource is added between the open track and the timetable point) and an outbound route with a departure event (the resource is added between the timetable point and the open track resource).

Having an additional resource results in the additional operation (therefore also a node in the AG) with processing time 0. A pair of alternative arcs is then added between every two nodes that represent events leading to inter-track conflict, in order to regulate the precedence relation between the two events.

Figure 5 shows an example of potential inter-track conflicts between train T3 and trains T1 and T2 at station S1. Figure 8 shows the resulting incompatibility graph. Events that can lead to conflict and can thus not occur within a specified headway time are connected by undirected arcs (red for inter-track and black for intra-track conflicts).

Since the time separation of trains running on the same open track segment is ensured by the selection of the alternative arcs related to the open track resource, there is no need to separate virtual resources representing D1 and D2 by additional alternative arcs. That way we avoid redundancy in modeling. Savings in computation time and graph size, achieved by avoiding redundancies will be presented in Section 7.

The alternative graph for this illustrative example is shown in Figure 9. Alternative arcs between virtual resources D1, D2 and A3 are added according to the incompatibility graph (Figure 8), where red pairs represent inter-track conflicts and black pairs represent intra-track conflicts. For example, a possible inter-track conflict in station S1 between T2 and T3 is modeled in the following way. If T2 departs first, T3 can arrive (operation A3 can start) only after T2 has departed (operation D2 has been completed, i.e. operation at resource OT2 of train T2 has started) and corresponding headway time has passed. Similarly, if T3 arrives first, T2 can depart (operation D2 can start) only after operation at station S1 of train T3 has started and the minimum headway time has passed.

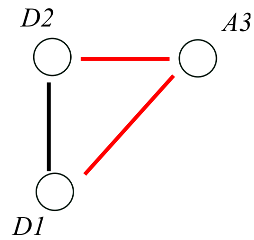

Figure 8: Incompatibility graph of illustrative example

5.4 Model 4

In this model, we partition the set of timetable points to stations, where overtaking is possible and stops on open tracks (or other timetable points), with no additional tracks to accommodate overtaking. The important property of the latter group is that their capacity allows only one operation (dwelling or through ride) at a time per direction.

Stations are modeled with resources type IC like in the previously described models. Stops are modeled with two resources of type B, one per direction. That way, due to the properties of this resource type (Section 4.4.4), timetable points where overtaking is not possible can not be occupied by more than one train per direction at the same time. Overtaking is in this model enabled only in stations with sufficient number of tracks and appropriate layout.

Alternative graph of the illustrative example is presented in Figure 10.

5.5 Microscopic model

The microscopic model of D'Ariano [5] is used to evaluate the performance of each macroscopic model studied here. This model has been validated and tested on numerous case studies. The model incorporates all operational constraints of railway traffic and provides accurate estimations of train movements at the level of block sections and signals.

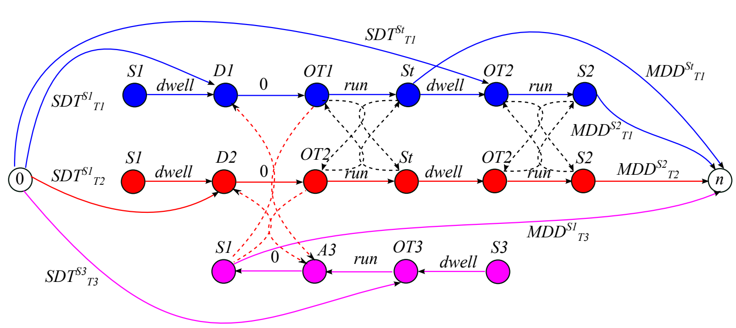

Figure 9: Illustrative example - Model 3

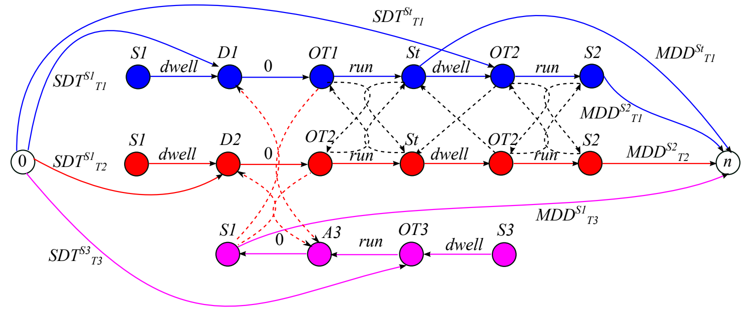

Figure 10: Illustrative example - Model 4

5.6 Overview of the five models

Table 2 summarizes operational constraints which are taken into account in the presented models. A gradual increase in number of considered operational constraints in the presented sequence from Models 1 to the microscopic model (Micro) is visible.

Depending on the network and traffic properties such as: capacity of stations, possibilities for occurrence of inter-track conflicts and heterogeneity of traffic, the appropriate modeling approach can be applied.

Another important criterion for selecting the most appropriate model is the size of the resulting graph and the computation time needed to obtain a solution of good quality. The performance of each model in terms of this criterion depends mainly on the network size and the number of trains. Further analysis will be presented in the following sections, which are focused on the computational experiments.

Table 2: Operational constraints in models

Model	Stations capacity	Stops capacity	Inter tr. conf.	Intra tr. conf.	Departure headway
Model1	-	-	-	-	✓
Model2	-	-	-	✓	✓
Model3	-	-	✓	✓	✓
Model4	-	✓	✓	✓	✓
Micro	✓	✓	✓	✓	✓

6 Test case A: corridor Utrecht - Den Bosch

Comprehensive evaluation of the macroscopic models relies on comparison with the microscopic model, which requires detailed infrastructure data on the level of block sections, signals and valid rolling-stock dynamics. The selection of the case study was thus related to the availability of data for generating the instances for the microscopic model.

All models have been applied to one hour of a timetable for the busy double-track line between Utrecht (Ut) and Den Bosch (Ht) in the Netherlands. Track layout of the corridor is presented in Figure 11.

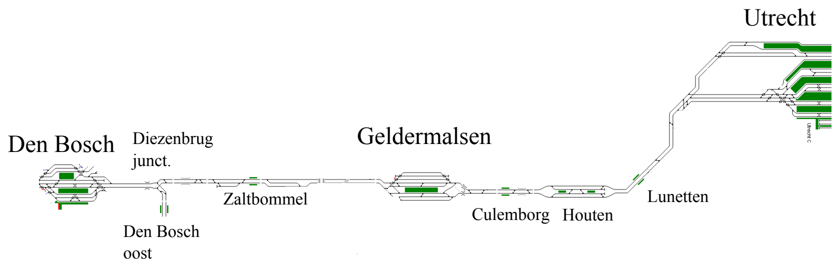

Figure 11: Layout of infrastructure and main stations [24]

The macroscopic infrastructure layout with all timetable points (stations, stops and junctions) is presented in Figure 12. In order to test the performance of the models in conditions where conflicts between intersecting or merging routes are possible, a branch that leads to Nijmegen (until station Den Bosch Oost, Hto) and merges with the main corridor in Diezebrug junction (Htda) just outside Den Bosch, is included. Big circles represent large stations where overtaking is possible (since Ht and Ut are area limits in this study, overtaking can be performed only in Geldermalsen), small circles represent stops on open track and the red circle in Htda specifies that inter-track route conflicts are possible.

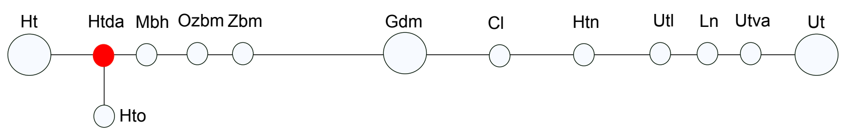

Figure 12: Macroscopic track layout

According to the periodic hourly timetable (Figure 13) there are four pairs of intercity trains that run between Utrecht and Den Bosch without stopping in intermediate stations (one pair of trains means one train in each direction). There are also two pairs of regional trains that stop in Zaltbommel (Zbm), Geldermalsen (Gdm), Culemborg (Cl), Houten (Htn), Utrecht Lunetten (Utl) and two pairs between Ut and Gdm (also stop in Cl, Htn, Utl). Trains that are shown to operate between Den Bosch and Htda (junction with a branch toward Nijmegen) in Figure 13 are two pairs of intercity trains and two pairs of regional trains running on the service between Nijmegen and Den Bosch. No trains are scheduled to stop at Lunetten - freight (Ln), Oud Zaltbommel (Ozbm) and Hedel (Mbh).

The scheduled departure and arrival times are given in the timetable for each station. The minimum dwell time is 120 s in large stations Ut, Gdm and Ht and 60 s in stops.

The minimum running times over an open track segment between two timetable points in the macroscopic models were obtained by summing up minimum running times over the corresponding block sections comprised by the open track segment.

The minimum headway times are in the microscopic model computed according to so-called 'departure on yellow' concept of blocking time theory [11], i.e., trains can depart as soon as the first block section on the open track segment has been cleared and released.

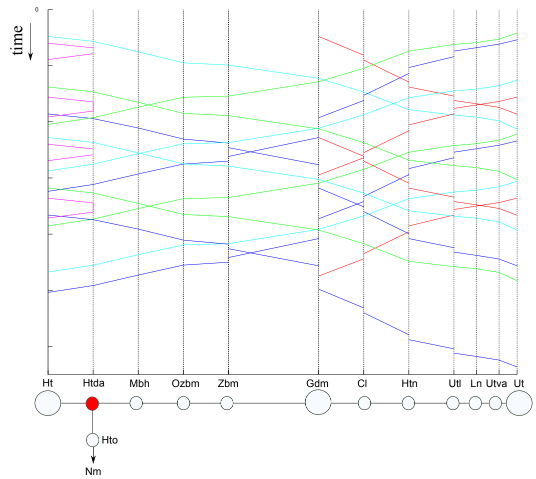

Figure 13: Timetable

This reflects the behavior of local traffic controllers in disturbed conditions. A train is allowed to depart as soon as the previous train has released the first block section. intra-track conflicts within the open track segment can occur only if a slower train is being followed by a faster train. The logic of blocking time theory is implemented in the microscopic model. Therefore, interactions between trains along the open track segments are regulated with high precision (i.e., a block section can never be occupied by more than one train).

On the other hand, macroscopic rescheduling models need to mimic the behavior of network traffic controllers with the aim to produce a new operational and conflict-free timetable with minimum deviation from the published timetable. Moreover, the only regulation of interaction between successive trains that run on the same open track segment is performed in departure station (in all macroscopic models) and arrival station (Model 2, 3 and 4).

Having this in mind, a different approach has been adopted in computing the minimum headway times between successive trains in the macroscopic models. The minimum headway time between departures (and arrivals in Models 2, 3 and 4) of two successive trains is equal to the maximum of running time of the first train over two adjacent block sections of the open track segment (increased by the time needed to clear the second section and reaction time of the signaling system). This way, the minimum headway time is independent from the running time of the second train, which is a limitation since the minimum headway time is overestimated if the second train is slower than the first one and can be underestimated if the second train is faster than the first. Deriving minimum line headways as in blocking time theory would require sequence dependent minimum headway times, computed with respect to blocking times of both trains.

6.1 Comprehensive evaluation

The five models were applied to the corridor test case. Solution procedure described in Section 4.2 was used to minimize secondary delay in all models. The complete equivalence of all models is achieved in terms of departure and arrival times of trains, when they were applied without delays.

In the following subsections the quality of solutions obtained by the macroscopic models will be evaluated by comparisons with the microscopic model (reference model). The smaller the differences, in terms of relative orders of trains, between the solutions obtained using the microscopic model and those obtained using macroscopic model, the better is the performance of the macroscopic model under evaluation. Comparisons between the objective values will be performed only among the macroscopic models due to the different way of computing the minimum headway times in the microscopic model.

A comprehensive evaluation of the models was performed over 200 delay instances. All trains from the timetable shown in Figure 13 are delayed in each instance according to the Weibull distributions as in Corman et al.[4]. The maximum primary delay is 326.80 s and the average primary delay is 30.15 s (both values are average over all instances).

The evaluation consists of two parts: models are compared in terms of (i) delay propagation and (ii) train orders, both on the solutions computed by the exact algorithm for each model.

6.1.1 Quantitative analysis

In the quantitative part of evaluation, presented in Table 3, the size of the resulting AG for each model is given in number of nodes, number of fixed arcs and number of alternative pairs (Columns 2-4). We also present the average computation time (CTF) to obtain the first solution using initial heuristics and average computation time (CTB) to compute the best solution or prove optimality for the initial solution over all instances (Columns 5-6). Moreover, average (ASD) and maximum (MSD) values of secondary delay over all instances are presented for each model (Columns 7-8).

Table 3: Quantitative assessment of the 5 models.

Mode	Nodes	Fixed arcs	Alt. pairs	CTF (s)	CTB (s)	ASD (s)	MSD (s)
Model1	394	505	558	< 1	< 1	4.38	89.16
Model2	394	505	1116	< 1	< 1	6.75	120.00
Model3	410	521	1164	< 1	< 1	7.43	124.00
Model4	410	521	1636	< 1	< 1	10.57	173.00
Micro	1018	1155	2312	< 1	1.20	5.88	119.56

As expected, the size of the graph increases together with the number of operational constraints considered in each model. There is a large difference in terms of ratio number of nodes/number of alternative pairs, between Model 1 and the microscopic model on the one side, and Models 2, 3 and 4 on the other. That can be explained by the fact that Models 2, 3 and 4 employ IC+FIFO resource type for modeling open track segments. Therefore, those models need twice as many pairs of alternative arcs to model train runs along open tracks compared to Model 1 (see Section 4.4.3 for a description of resource type IC+FIFO).

Savings in computation time between the microscopic model and macroscopic models are captured but for applications on this relatively small test case all five models show excellent performance in terms of computation time to obtain the first as well as the best solution. For this set of instances, the optimal solution was always found for all models.

The last two columns of Table 3 show that the average and maximum secondary delay increase along with the number of operational constraints taken into account in each macroscopic model, meaning that the more realistic models are able to capture more interactions between trains and therefore compute more realistic delay propagation (the microscopic model is not considered in this analysis due to different computation of minimum headway times).

6.1.2 Comparison of train reordering actions

Reordering trains (changing the order of departures) is a common dispatching action for reducing delay propagation. In this section, we will investigate how close are the solutions of macroscopic models in every instance to the solution of the reference microscopic model in terms of orders of departures. The analysis has been carried out on trains running from Den Bosch towards Utrecht. There are three checkpoints where the relative order of trains in the direction toward Utrecht is determined: through runs in Htda, departure from Gdm and arrival in Ut. By checking the orders of through runs in Htda we are able to estimate the effect of considering inter-track conflicts that are possible to occur in the junction Htda. According to the published timetable, intercity trains are scheduled to overtake slower regional trains in Gdm. Therefore, checkpoints in Gdm and Ut are used to verify if some macroscopic models provide solutions with a different point of overtaking (which in reality is unfeasible). The first three rows of Table 4 give the percentage of train sequences (for each macroscopic model) that are different from the corresponding sequences produced by the microscopic model, in each check point on the 200 instances. The last row of the table shows the percentage of different sequences aggregated over all three check points.

Table 4: Difference in orders between the microscopic and each macroscopic model. Direction Ht $\rightarrow$ Ut

	Model 1	Model 2	Model 3	Model 4
Through run Htda (%)	1.0	1.0	0.5	0.0
Departure from Gdm (%)	33.5	20.0	20.0	2.0
Arrival to Ut (%)	4.5	4.0	4.0	2.0
Average (%)	13.0	8.3	8.2	1.3

In almost all instances, the solutions of four macroscopic models suggest identical sequences of departures from Htda as the microscopic model. Therefore, differences in operational constraints, included in the models (Table 2), are only to small extent manifested in different relative order of trains running through Htda.

By comparing the percentage of different sequences of departures from Gdm and arrivals to Ut for each model, it is visible that in the large number of instances, Models 1, 2 and 3 allow overtaking between Gdm and Ut (the number of differences at arrival to Ut is much smaller than the number of differences at departure from Gdm). In Model 4, the percentage of different sequences is the same in both check points which implies that the relative order of trains that depart from Gdm is maintained until Ut.

This comparison of aggregated differences shows that Model 4 gives solutions closest to the accurate microscopic model compared to other macroscopic models. Only 1.3% departure sequences are different on the three checkpoints. Other macroscopic models show greater deviation from solutions provided by the microscopic model. This deviation percentage is again correlated to the number of operational constraints included in the models.

7 Test case B: Dutch national railway network

The primary purpose of this section is to test the applicability of the macroscopic models presented in Section 5 for the management of large and busy networks. Figure 14 shows the test case of the Dutch national network that comprises approximately 6800 km of tracks, and represents one of the busiest railway networks in the world with more than 700 passenger trains operating during peak hours.

7.1 Description of the tested instances

Input data for the macroscopic models of traffic on the Dutch national network is obtained from the macroscopic timetabling tool DONS (Designer Of Network Schedules) [12], that is able to generate a periodic hourly timetable on the national level with all scheduled event times in all timetable points (departures and arrivals) and scheduled process times (running and dwell times, connection times and headways) rounded to full minutes. Slack times and time reserves are not included in the DONS constraints database, which is used to build the timed event graph.

In order to reduce the size of the problem without loosing validity we have computed all strongly connected components in the graph as explained in Goverde [10]. If a primary delay occurs within a strongly connected component, it cannot propagate to other strongly connected components. Therefore, each strongly connected component of a TEG corresponds to an autonomous model. The strongly connected component considered in this example comprises the largest part of the Dutch national hourly timetable and takes into account all trains operating on the lines depicted by black solid lines in Figure 14. Thick solid lines represent double and multiple-track segments, whereas the thin solid lines stand for single-track segments.

Table 5 reports specific information on the instances used to test the macroscopic models. Size and properties of the network in this test case are presented. We take into account all intercity, regional and freight trains (reserved slots).

7.2 Comprehensive evaluation

The four macroscopic models have been tested on 200 delay instances in which all trains were delayed according to Weibull distribution, similar as in Section 6.1. The maximum primary delay is 18.22 min and the average primary delay is 1.41 min (both values are average over all instances).

Table 6 reports average results for the network-wide instances on each macroscopic model (Column 1): the number of nodes, fixed arcs and alternative pairs (Columns 2-4), the average computation time (CTF) to obtain the first solution using initial heuristics

Figure 14: Dutch railway network considered (in black), with main stations.

Table 5: Characteristics of the network-wide test case

Instance property	Number
Stations	298
Other timetable points	294
Unidirectional open track segments	1119
Bidirectional open track segments	324
Trains	679
Connections	84

(Column 5), average computation time (CTB) to compute the best solution or prove optimality for the initial solution over all instances (Column 6) and the average (ASD) and maximum (MSD) secondary delays (Columns 7–8). All values in Columns 5–8 are average over 200 instances.

Table 6: Quantitative assessment of the macroscopic models on test case B.

Mode	Nodes	Fixed arcs	Alt. pairs	CTF (s)	CTB (s)	ASD (min)	MSD (min)
Model1	17490	20591	16494	5.03	5.06	0.21	4.95
Model2	17490	20591	32380	45.71	45.78	0.25	5.26
Model3	18968	22069	33956	50.93	51.00	0.29	5.42
Model4	18968	22069	42750	83.91	84.01	0.36	7.62

From Table 6, pairwise comparison between the macroscopic models in terms of the graph size, computation time and average secondary delays lead to the following conclusions. Model 4, the most realistic macroscopic model, captures the largest amount of secondary delays compared to the other macroscopic models.

The more precise information comes with a cost in the alternative graph size and in the computation time of solution algorithms. Initial heuristics are used to compute the first solution. The branch and bound algorithm proves optimality for all 200 instances of Models 1, 199 instances of Model 2 and 3. For Model 4, the optimal solution is proved for 163 instances while for the remaining 37 instances the branch and bound algorithm is not able to compute the optimal solution within the given time limit of computation (5 min).

Avoiding redundancies in modeling inter-track conflicts (as described in Section 5.3) results in large savings in the size of AG and consequently in computation time. This becomes visible when both variants of Model 4 are applied to the network-wide instance. If we model all intra-track conflicts with redundant alternative arcs, the size of the graph increases to 73 655 alternative pairs. The average computation time to obtain the first solution is 222.11 s and the best solution is produced within the time limit in 5 instances.

7.3 Network-wide effects of reducing delay propagation

In order to demonstrate the effect of minimization of secondary delays on the national network, we compare the delay propagation that arise if the relative order of events (departures and arrivals of all trains) remains as scheduled in the timetable, with the secondary delays that occur as a result of the solution procedure on Model 4. The maximum primary delay in the generated instance is 16 min and the average primary delay is 1.24 min. The total secondary delay accumulated in all stations is 3093 min when the order of events is fixed and 1611 min if secondary delays are minimized by applying the solution procedure (Section 4.2) on Model 4.

Figure 15 presents the maximum secondary delays in major stations in the Netherlands with fixed order of events (left) and after modifying the order of events as proposed by the optimal solution (right). Without rescheduling actions, the maximum secondary delays are the largest in the busiest part of the network around Amsterdam (Asd) and Utrecht (Ut), as well as in Leiden (Ledn), Apeldoorn (Apd) and Tilburg (Tb). Secondary delays still occur after optimization in the busiest part of the network but the network-wide effect of rescheduling actions is clearly visible compared to the left part of Figure 15.

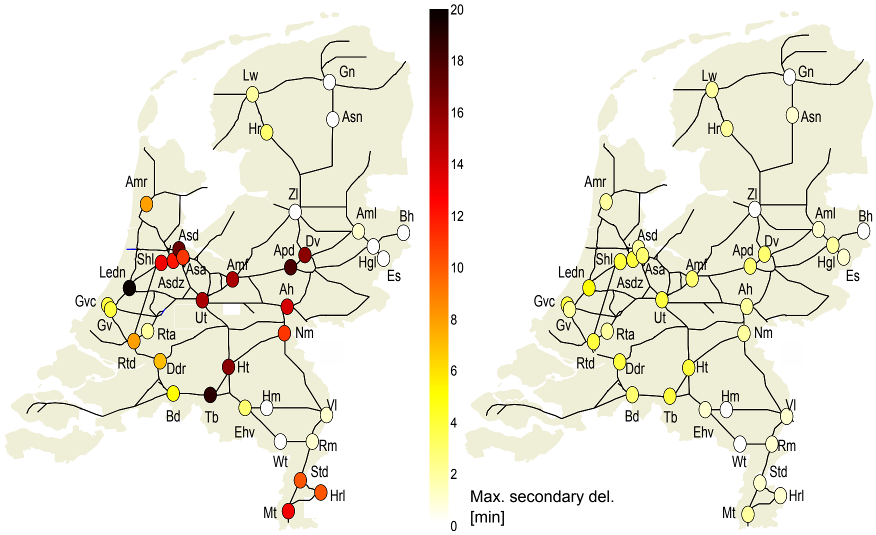

Figure 15: Delay propagation without (left) and with (right) rescheduling

8 Conclusions and outlook

The potential further growth of both passenger and freight flows in already busy railway networks in western Europe will mostly have to be accommodated over the existing railway infrastructure. This will lead to an increase of capacity utilization thus reducing reliability and punctuality of railway services. Improvements in traffic management and control have to be made in order to prevent a decrease of traffic reliability. In that context, this contribution leads to an improvement of global delay propagation indicators.

This paper presented four models of railway traffic flows at a macroscopic level. Trade-off between the level of detail included in each model and the number of considered operational constraints was examined in terms of minimization of secondary delay and computation time. A comprehensive evaluation was performed on two real-world case studies. We were able to handle very large instances such as the Dutch national network within reasonable time even with the most complex macroscopic model.

Further work will be dedicated to study other traffic disturbances and dispatching measures, such as global rerouting. Additional levels of detail of the macroscopic models will be further investigated, e.g. by introducing sequence dependent headway times rather than using conservative headways. Regarding the solution algorithms, on the one hand we are investigating possible speed-ups of computation related to the reduction of the alternative graph size in a pre-processing step. On the other hand, further research should be dedicated to the development, implementation and testing of sophisticated meta-heuristics in order to compute good quality solutions in a short computation time.

Acknowledgements

This work is partially funded by the Dutch Technology Foundation STW, research project: Model-Predictive Railway Traffic Management (project no. 11025).

References

[1] Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M.: Stochastic Delay Prediction in Large Train Networks. In: A. Caprara, J. Kontogiannis (eds.) 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, pp. 100–111. Dagstuhl, Germany (2011)

[2] Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M.: Bi-objective conflict detection and resolution in railway traffic management. Transportation Research Part C: Emerging Technologies **20**(1),79-94, (2012)

[3] Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M.: Dispatching and coordination in multi-area railway traffic management. Tech. Rep. RT-DIA-190-11, pp. 1–30, Dipartimento di Informatica e Automazione, RomaTre (2011)

[4] Corman, F., D'Ariano, A., Pranzo, M., Hansen, I.: Effectiveness of dynamic reordering and rerouting of trains in a complicated and densely occupied station area. Transportation Planning and Technology **34**(4), 341–362 (2011)

[5] D'Ariano, A.: Improving Real-Time Train Dispatching: Models, Algorithms and Applications. Phd thesis, TU Delft (2008)

[6] D'Ariano, A., Pacciarelli, D., Pranzo, M.: A branch and bound algorithm for scheduling trains in a railway network. European Journal of Operational Research **183**(2), 643–657 (2007)

[7] D'Ariano, A., Pranzo, M.: An Advanced Real-Time Train Dispatching System for Minimizing the Propagation of Delays in a Dispatching Area Under Severe Disturbances. Networks and Spatial Economics **9**(1), 63–84 (2008)

[8] D'Ariano, A., Pranzo, M., 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)

[9] Goverde, R.M.P.: A delay propagation algorithm for large-scale railway traffic networks. Transportation Research Part C: Emerging Technologies **18**(3), 269–287 (2010)

[10] Goverde, R.M.P.: Railway timetable stability analysis using max-plus system theory. Transportation Research Part B: Methodological **41**(2), 179–201 (2007)

[11] Hansen, I.A., J, P. (eds.): Railway timetable & traffic - analysis, modelling, simulation. EUrail press, Hamburg (2008)

[12] Hooghiemstra, J.S.: Design of regular interval timetables for strategic and tactical railway planning. In: J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto, S. S (eds.) Computers in Railways V, pp. 393–402. Computational Mechanics Publications, Southampton (1996)

[13] Jacobs, J.: Reducing delays by means of 'on the spot' reschdeduling. In: J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto, S. Sone (eds.) Computers in Railways IX, pp. 603–612. WIT Press, Southampton (2004)

[14] Mascis, A., Pacciarelli, D.: Job-shop scheduling with blocking and no-wait constraints. European Journal Of Operational Research **143**, 498–517 (2002)

[15] Min, Y.H., Park, M.J., Hong, S.P., Hong, S.H.: An appraisal of a column-generation-based algorithm for centralized train-conflict resolution on a metropolitan railway network. Transportation Research Part B: Methodological **45**(2), 409–429 (2011)

[16] Nash, A., Huerlimann, D.: Railroad simulation using OpenTrack. In: J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto, S. Sone (eds.) Computers in Railways IX, pp. 45–54. WIT Press, Southampton (2004)

[17] Schachtebeck, M., Schöbel, A.: To Wait or Not to Wait–And Who Goes First? Delay Management with Priority Decisions. Transportation Science **44**(3), 307–321 (2010)

[18] Schöbel, A.: Integer programming approaches for solving the delay management problem. In: F. Geraets, L. Kroon, A. Schöbel, D. Wagner, C. Zaroliagis (eds.) Algorithmic Methods for Railway Optimization, *Lecture Notes in Computer Science*, vol. 4359, pp. 145–170. Springer Berlin / Heidelberg (2007)

[19] Siefer, T., Radtke, A.: Evaluation of Delay Propagation. In: Proceedings of 7th World Congress on Railway Research. Montreal (2006)

[20] Tomii, N., Yoshiaki, T., Noriyuki, T., Chikara, H., Kunimitsu, M.: Train rescheduling algorithm which minimizes passengers dissatisfaction. In: M. Ali, F. Esposito (eds.) Innovations in Applied Artificial Intelligence, *Lecture Notes in Computer Science*, vol. 3533, pp. 37–43. Springer Berlin/Heidelberg (2005)

[21] Törnquist, J., Persson, J.: N-tracked railway traffic re-scheduling during disturbances. Transportation Research Part B: Methodological **41**(3), 342–362 (2007)

[22] Törnquist Krasemann, J.: Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances. Transportation Research Part C: Emerging Technologies **20**, 62–78 (2011)

[23] Van den Boom, T.J.J., De Schutter, B.: On a model predictive control algorithm for dynamic railway network. In: I.A. Hansen, A. Radke, J. Pachl, E. Wendler (eds.) Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis (RailHannover2007), Hannover, Germany (2007)

[24] www.sporenplan.nl