

Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions

Rob M.P. Goverde^{1,§}, Francesco Corman^{2,3}, Andrea D’Ariano⁴

¹Dep. of Transport & Planning, Delft University of Technology, The Netherlands

²Centre for Industrial Management, Katholieke Universiteit Leuven, Belgium

³Dep. of Maritime and Transport Technology, Delft University of Technology

⁴Dipartimento di Ingegneria, Università degli Studi Roma Tre, Italy

[§]Corresponding author: *r.m.p.goverde@tudelft.nl*

Abstract

This paper evaluates the capacity consumption on a Dutch railway line both under scheduled and disturbed traffic conditions. For the scheduled condition the standard UIC compression method is used, while the computation of capacity consumption under disturbed conditions requires multiple simulation runs via a Monte Carlo simulation set up. For the analysis we use the optimization-based train dispatching system ROMA that combines the alternative graph formulation of train rescheduling with blocking time modelling of the signalling constraints. For this study ROMA is extended to various signalling/ATP systems so that the braking behaviour of hindered trains is accurately simulated in the different configurations. In addition, ROMA computes the compressed timetable with conflict-free train paths without rescheduling. For the disturbed scenarios, we consider a case study of the Utrecht-Den Bosch line equipped with different signalling/ATP systems. Two traffic control scenarios are considered for the disturbed scenarios: a first come first served rule and an advanced branch-and-bound algorithm. The results show that the scheduled capacity consumption improves from the Dutch NS’54/ATB to ETCS Level 2 with short blocks. In presence of delayed operations, the capacity consumption declines for NS’54/ATB, since trains have to brake and run at lower speeds. With ETCS, there is a considerable gain in terms of capacity consumption and punctuality compared to NS’54/ATB, since the braking distances decrease when delayed trains run at lower speeds, having a stabilizing effect on headway times, delay propagation and throughput.

Keywords

ERTMS, ETCS, capacity consumption, schedule optimization, disturbance handling

1 Introduction

The characteristics of a signalling and automatic train protection (ATP) system have a significant impact on the capacity and stability of a railway line. One way to improve railway infrastructure capacity is to update the signalling or ATP system to one that allows a closer headway between successive trains. A capacity assessment with respect to the different signalling and ATP variants then reveals how much capacity gains can be achieved. In Europe,

the new state-of-the-art European Train Control System (ETCS) has been implemented successfully on various new (high-speed) lines and moreover it is being implemented on six international freight corridors as decreed by the European Commission (EC) to improve interoperability of the European railways [28]. In 2007, all European Union member states also had to submit a national implementation plan to the EC. The European countries are facing the strategic dilemma to how, where, and when to install ETCS on conventional railway lines. Different countries might have different reasons for replacing their safety and signalling systems, including interoperability, improved safety, increased capacity, replacing legacy systems at the end of their life cycle, and improved or extra functionality like higher supervised speeds. Since the investments of signalling equipment replacements are considerable, both in the infrastructure and rolling stock, a Social Cost-Benefit Analysis (SCBA) is typically applied to balance the costs to benefits. A crucial element in such analysis is the quantification of benefits. This paper focuses on the evaluation of capacity benefits when implementing ETCS by comparing it with the legacy system for a case study in the Netherlands. Part of this study was performed in commission of the Dutch Parliament [10].

The capacity consumption of a railway line can be calculated using the timetable compression method for given infrastructure characteristics, rolling stock characteristics and timetable pattern. This compression method is based on a deterministic microscopic calculation of conflict-free train paths with minimum headway times using blocking time theory [11]. This approach is also adopted as the standard method for assessing capacity consumption by the International Union of Railways [25], which also gives empirically derived guidelines on the total required buffer time in a timetable pattern to be stable for delays. The UIC timetable compression method has been applied to evaluate the capacity consumption of various ETCS variants [26, 27, 28]. In the literature, this capacity consumption assessment method has been critically reviewed; see e.g. Landex [14], Lindner [15], and Lindner and Pachl [16]. For an overview of other capacity assessment methods, see Abril *et al.* [1].

Capacity planning is normally applied to planned train paths and scheduled conditions (see e.g. [3, 28]) and does not consider the actual capacity consumption under disturbed conditions and the unavoidable stochastic phenomena. As a result, although theoretically conflict-free, a timetable could be very sensitive to train path deviations, even with a low capacity consumption [19]. In the presence of delays, train path conflicts might occur depending on the scheduled buffer time between train paths. In practice this means that a train has to brake in response to the signalling system and possibly wait in rear of a stop signal. This leads to changed train trajectories with increased blocking times so that following trains may also be affected. The actual response to the signals depends on the specific signalling system and constraints of the ATP system, and can be quite different. In a standard three-aspect signalling system, trains have to brake from two signals in rear of an occupied block section to a restricted speed and continue with this restricted speed until a final brake before the red signal. In contrast, a cab signalling system as ETCS allows a train to proceed until an approach indication point that is determined by a dynamically computed braking curve on-board. This braking curve depends on the actual train speed and braking characteristics as well as on the infrastructure description until the end of the movement authority (such as slopes). Also the time of re-acceleration after an improved signal aspect depends on the (intermittent or continuous) ATP system.

The contribution of this paper is the quantitative evaluation of the capacity consumption on a Dutch mixed-traffic railway corridor both under scheduled and disturbed conditions. For the scheduled condition the standard UIC compression method is used; while for the

disturbed conditions, we present a new concept of *dynamic capacity consumption*. Random distributions of the initial delay are evaluated for each train when entering the station at the beginning of the line, which model the unavoidable variability of real-life operations. The resulting conflicts between trains are then detected and solved for multiple simulation runs. In each run, a *dynamic minimum cycle time* is computed, while the average capacity consumption of all replications is used as a measure of the capacity consumption for disturbed situations. Note that dynamic evaluations depend on the (time allowances in a) given timetable, the assumed delay distributions, and the train and traffic control applied.

For the capacity analysis we adopted the train dispatching system ROMA [9]. ROMA is based on the combination of alternative graph model [18] and blocking time theory [11], and is thus applicable to any signalling/ATP system. For this study, ROMA is extended with various signalling/ATP systems so that the braking behaviour of hindered trains is precisely simulated in the different configurations. Also the route-locking sectional-release principle is taken into account for accurate blocking time calculations in the station areas [6]. ROMA has been adapted in order to compute the compressed timetable with conflict-free train paths without and with rescheduling. For the calculation of the capacity consumption for disturbed scenarios with rescheduling, ROMA is applied in a Monte Carlo simulation set up. We use two rescheduling approaches: a first come first served (FCFS) rule and an advanced branch-and-bound (B&B) algorithm [8]. The optimization approach was first developed in a software, named AGLIBRARY, for solving general scheduling and rerouting problems and then incorporated in ROMA. We evaluate the two approaches to demonstrate the potential advantage of real-time intelligent traffic control in combination with the ERTMS Level 2 two-way communication architecture.

The capacity consumption assessment is applied to a case study of the Dutch Utrecht-Den Bosch railway line with different signalling/ATP systems. Four signalling/ATP system variants are considered: the current Dutch NS'54 speed signalling system with ATB train protection, the NS'54/ATB system but with optimized blocks near the stops, ETCS Level 2 with current block lengths, and ETCS Level 2 with shorter blocks. The disturbances are modelled as stochastic initial delays for all trains with train-type dependent distributions given by three-parameter Weibull distributions fitted by empirical data.

The next section explains the considered signalling and ATP systems, and the timetable compression method applied to these different systems. Section 3 explains the ROMA system and the extensions that we developed for modelling the signalling and ATP systems, and for the computations of capacity consumption in this work. Section 4 illustrates the results to a practical case study regarding the Utrecht-Den Bosch corridor. Section 5 gives the conclusions of this paper and recommendation for further investigations.

2 Railway signalling and capacity consumption

2.1 Signalling and ATP systems

Railway safety systems can be partitioned in four components [23]:

1. Track-free detection: detecting the occupation and release of track sections,
2. Interlocking: setting technically protected routes for safe train movements,
3. Signalling: indicating a movement authority to train drivers, and
4. Automatic train protection (ATP): guard against driver errors.

Track-free detection devices such as track circuits or axle counters monitor the occupation of track sections. Interlocking systems guarantee that a signal is only released if the route to a next signal is safe, i.e., when the following route requirements are met: all points properly set and locked, conflicting routes locked, flank protection, and tracks clear. Open tracks between station layouts are usually operated by automatic block systems which can be viewed as a simple and automatic form of interlocking. By definition an open track does not contain any points so that signals only have to guard against following and opposing movements. The open track is then partitioned in block sections each protected by a block signal in rear. The block signals react automatically on messages from the track-free detection and show signal aspects according to a signalling logic that varies from country to country. Finally, ATP systems have supervision and intervention functions as a fallback to driver errors and again the functionalities differ largely amongst the different systems. This study analyzes two signalling and ATP systems that are applied in the Netherlands: the legacy NS'54 speed signalling system with ATB automatic train protection and ETCS Level 2, see Figure 1.

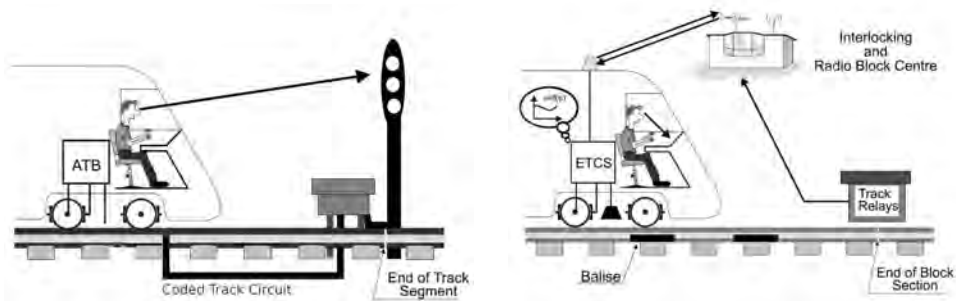


Figure 1: Basic setup of signalling/ATP systems NS'54/ATB and ETCS L2

NS'54/ATB

The signalling system NS'54 (originating from 1954) operates by light signals that give speed commands to drivers such that trains can always brake before a signal at danger. NS'54 uses fixed block sections with lengths that are at least equal to the maximum braking distance for a worst braking train plus a margin for the reaction time of the driver (and ATP system). The maximum braking distance depends on the entry speed into the block and is laid down in the Dutch railway law for a downhill gradient of 5 %. A *normal block* has a length corresponding to the line speed in which case NS'54 is just a three-aspect two-block system with a clear (green, G), approach (yellow, Y) and stop (red, R) aspect. A green signal indicates that the train may pass unhindered with the (local) maximum speed, a yellow signal orders to reduce speed to a restricted speed of 40 km/h and prepares to stop before a red signal, and a red signal orders to stop before the signal. As an example, for a line speed of 130 km/h the maximum braking distance is 1000 m and the corresponding block length with margin is 1181 m.

Near stations also *short blocks* are applied which have a length shorter than the maximum braking distance from the line speed by which trains can follow at a shorter headway. In this case, NS'54 indicates already one (or more) signals before that the train has to slow down to an indicated speed that must be reached before the next signal, so that the train will

enter the short block with a lower speed associated to the short block length. This speed signalling is given by a yellow signal plus a white numeral indicating the permitted speed at the next signal. For instance, 'yellow 8' (Y8) indicates that the train has to reduce speed to 80 km/h before the next signal. This speed corresponds to a maximum braking distance of 800 m and block length of 918 m. For 40 km/h a maximum braking distance and block length of 400 m is defined. Consider for instance a block sequence from an open track to a station platform track with successive block lengths 1350 m, 1000 m and 400 m. Then a train is guided from the open track to standstill at the platform track by progressive speed signalling over a signal aspect sequence G-Y8-Y-R, assuming a line speed of 130 km/h.

The NS'54 wayside signalling system is complemented with the on-board automatic train protection system ATB to guard against errors of train drivers. ATB is a continuous ATP system that supervises overspeed and braking orders by the signalling system in rough steps (40, 60, 80, 130, 140 km/h). These speed steps are transmitted to the train via coded track circuits. After a speed reduction order (a yellow signal with possible speed indication) the driver has to apply the brakes until the permitted speed is reached. In the case of a yellow signal the driver has to reduce speed to a restricted speed of 40 km/h and then may drive on-sight for the remainder of the block. If a driver does not brake sufficiently after a speed reduction order then ATB warns the driver and if the driver still does not react ATB will intervene with an emergency brake to standstill. Speeds below 40 km/h are not supervised assuming that the driver runs on-sight and stops in time before a red signal. The supervised speeds correspond to the nearest ATB speed step above the permitted speed. For instance, if the permitted speed is 100 km/h then the ATB-code corresponds to 130 km/h which is shown to the driver as the supervised speed in the cabin. It is the responsibility of the driver to stay below 100 km/h. So the supervised speeds in the cabin are not equal to the actual permitted speeds. The driver can therefore not rely on the ATB cabin information but needs to pay attention to the wayside speed signs and signals. If the signal ahead improves to an aspect with an improved ATB-code (e.g. from red to yellow) then this is directly communicated to the train via the coded track circuits and the driver may increase speed again promptly. Since the 1990s ATB is also called ATB-EG (for 'first generation') to differentiate it from the then newly introduced system ATB-NG (for 'new generation') which is a completely different intermittent ATP system installed mainly on regional lines that did not have ATB at the time. In the computational experiments of this paper, only ATB-EG is considered.

ETCS Level 2

ETCS Level 2 is an integrated cab signalling and train protection system which will probably follow up the NS'54/ATB system in the near future. Also ETCS L2 is a fixed block system with track-free detection that the interlocking system uses to set routes. However, now the interlocking communicates the set route(s) to a Radio Block Centre (RBC). The RBC translates the set route(s) into a Movement Authority (MA) and sends this together with a track description to the ETCS L2 on-board computer. The ETCS on-board computer computes a dynamic speed profile based on this incoming message, the current speed, and the train characteristics (maximum speed, braking curve), taking into account all track speed restrictions until the End of Authority (EoA). The MA and dynamic speed profile are displayed to the driver in the cabin, and are supervised continuously, i.e., the on-board computer supervises both the permitted speed at each location as well as the braking curves at each speed reduction. The driver must start braking only at the braking distance before an intermediate speed restriction or the EoA. This indication point depends on the track and

train characteristics and the current speed. Note that an ETCS L2 train does not have to brake at the entry of a block but just before the computed braking curve. Trackside signals are no longer used. If a driver does not brake sufficiently and exceeds the (warning) braking curve a warning is given; if he then still does not brake enough an emergency brake intervention to standstill is carried out.

Moreover, ETCS L2 has a two-way safe communication with the RBC via the GSM-R radio connection. The train sends (semi-)continuously its position and speed to the RBC by which the RBC can also request new routes to the interlocking which are then again translated in new MA's that are send back to the train. As soon as a new MA has been received the current MA is extended with the new MA. So if for instance a train was braking before an EoA and the MA is extended then the driver can immediately accelerate again.

2.2 Capacity consumption: the timetable compression method

The timetable compression method is based on blocking time theory [11]. The time duration over which a block section is allocated exclusively to a specific train is called blocking time. During this blocking time the block is thus blocked for other trains. In station areas an interlocked route between two successive signals can also be seen as a (route) block section. The blocking time of a block for a running train consists of the following parts, see Figure 2:

1. Setup time to set and lock the route in the block,
2. Sight and reaction time before the approach signal,
3. Approach time from the approach signal to the signal in rear of the block,
4. Running time in the block,
5. Clearing time (running time over the train length to clear the block), and a
6. Release time to release the route in the block.

For three-aspect fixed-block signalling systems the blocking times thus exceed the running time over two blocks required for unhindered running. The blocking time of an outbound route block after a scheduled stop in a station has zero approach time, although the setup time associated to setting the outbound route is here larger than for an automatic block signal. In an interlocking area with sectional-release route-locking the route block is released in steps after clearing the points. For more details on blocking times, see Pacht [20].

For the progressive speed signalling of short blocks the approach indication to the end of the short block already starts two (or more) signals in advance. The approach time is then given by the (free) running time from the first signal that would order a speed reduction to be able to stop at the end of the short block. For ETCS L2 the approach indication is provided by the braking indication point according to the train-, track- and speed-dependent braking curve which is independent from block boundaries. The approach time is in this case the free running time over the braking distance from the indication point to the EoA (red signal). Hence, in ETCS L2 a fast braking train has a smaller approach time (and distance) than a slower braking one. And likewise, if a train runs at a lower speed then it has a smaller approach time than when it is running faster. For more details on blocking time theory specifically for ETCS, see Winter [28].

Drawing the successive blocking times for each train over a railway line in a time-distance graph leads to a *blocking time diagram*, cf. Figure 8 in Section 4. The subsequent blocking times of a train form a so-called *blocking time stairway* in the blocking time diagram. The blocking time stairways represent the operational use of a railway line by a train

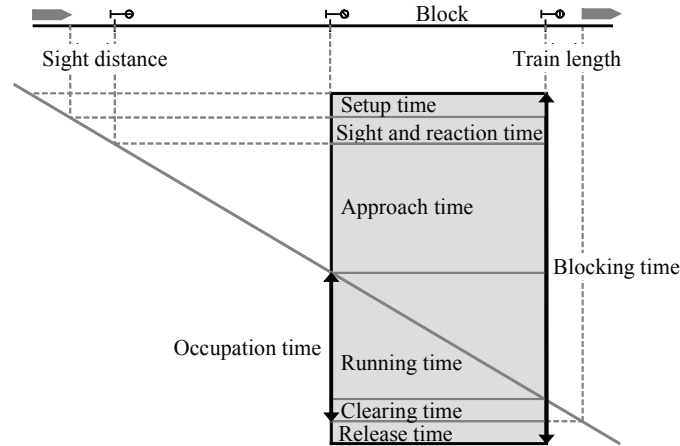


Figure 2: Blocking time for a running train in three-aspect signalling

and can be compared to tetris pieces. The capacity consumption of a given train sequence on a railway line is easily derived by pushing the tetris pieces against each other, which is basically the timetable compression method. The minimal time interval between successive departure times then denotes the minimum (line) headway time between the associated trains: the second train cannot start earlier than this minimum headway time after the preceding train. The block(s) where the blocking times of the two successive stairways touch each other are the *critical blocks* for this train sequence. Figure 8 shows the compressed timetables for two signalling systems. As can be seen in this figure (without going into details which is postponed to Section 4) the critical section of a slow after fast train is one of the first blocks of the line (depending on the block lengths), and a fast after slow train is one of the last blocks of the line (or before an overtaking facility).

For a cyclic timetable where the same timetable pattern repeats at a fixed cycle time – such as the Dutch basic hour pattern – the *minimum cycle time* associated to the compressed timetable is obtained by the time interval between the departure times of the first train in two successive cycles of a compressed timetable. The ratio of the minimum cycle time and the planned cycle time is called *capacity consumption* and represented as a percentage. The minimum cycle time is also called ‘infrastructure occupation’ [25] but to avoid confusion with the physical occupation time of the infrastructure we prefer the former term. As prescribed by the UIC leaflet we include running time supplements in the running times. However, the present study does not include buffer time or supplements for maintenance in the computation of capacity consumption. Note that the minimum cycle time of a basic hour timetable pattern must be smaller than 60 minutes to be realizable. For mixed-traffic lines the UIC advises a capacity consumption of at most 75% in the peak hours and 60% during the daily period [25].

For this paper the following parameters have been used for the computation of the blocking times. For NS’54/ATB the sight and reaction time was assumed 9 s, the combined setup and release time in station interlocking areas $12+2=14$ s, and the setup and release time on open tracks 2 s. For ETCS, the reaction time was assumed 10 s (there is no sight time in cab signalling), the sum of setup and release time in station interlocking areas 15 seconds, and

the latter on open tracks 3 seconds. The setup time for ETCS includes a delay for communicating the MA to the train via the RBC and GSM-R. The reaction time for ETCS includes displaying a new MA on the Driver Machine Interface. For trains departing from a station after a scheduled stop 24 s setup time is calculated in the blocking time from the signal clearing to the departure corresponding to the departure process. This relates to the reaction time to the signal clearing, the procedure of closing the doors with all safety checks, informing the driver, and finally achieving the actual departure. So, this study assumed the fixed system times for ETCS 2 is larger than for NS'54/ATB. The ETCS values are rather conservative and add up in all the minimum headway time. When actual data comes available about real communication delays for ETCS L2 this can easily be updated. In the mean time our results cannot be accused for being too optimistic.

3 The train dispatching system ROMA

In the recent literature, train dispatching approaches are based on optimization models and algorithms for intelligent rescheduling, see e.g. [4, 9, 12, 17, 22, 24]. This work uses the train dispatching system ROMA [9], which is based on a microscopic model of traffic predictions and computes optimal rescheduling decisions based on an exact global conflict resolution algorithm. Traffic management is considered in a large area, detecting future conflicts affecting train movements, computing a conflict-free train schedule in the whole area by using global information that might result in changes of train routes, orders and times. In this work, train routes are fixed and the optimization is limited to train orders and speeds, since our aim is to investigate the combined effect of optimal train rescheduling for various signalling/ATP systems.

Figure 3 presents the modular structure of the dispatching system. In the next subsections, each module is described and we provide a description of the extensions developed in this study to deal with the different signalling systems.

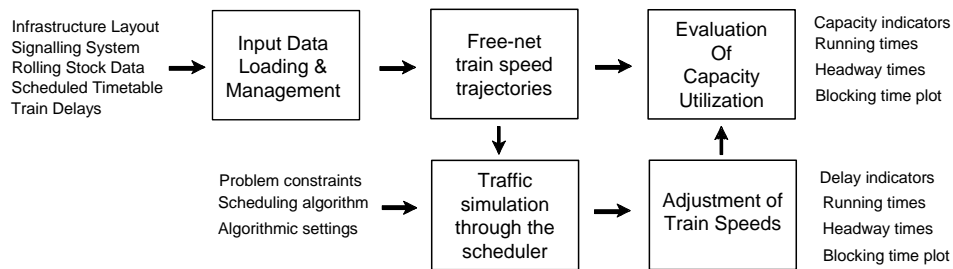


Figure 3: The modules of the dispatching system

3.1 Input data loading and management

The module for input data loading and management is in charge of gathering all the information that is required by the other modules. This module loads static data (data such as infrastructure and timetable information) and dynamic data (train detection and other

real-time information that varies in time) from the field.

The timetable contains a list of arrival/departure times for a set of relevant points in the network, including all the station platforms visited by each train. The infrastructure consists of the network layout with block sections delimited by signals. Infrastructure data includes the status and length of each block section and other characteristics, such as speed restrictions and travel direction. The data associated with each train includes speed and position at the entry of the network, maximum speed and acceleration and braking curves (calculated accurately based on rolling stock characteristics). At any time a train route is available if none of its block sections is blocked. Finally, the blocking time for each pair (train, block section) is computed by this module on the basis of current rolling stock characteristics and infrastructure data.

The static data is stored in a format compatible with the RailML specifications which are becoming the standard interface for railway data in Europe. For the dynamic data, we assume that the actual position and speed of all trains are available to the conflict detection and resolution algorithm. With ETCS L2 the train sends this dynamic data semi-continuously via GSM-R to the RBC which is connected to the traffic control centre. For traditional systems like NS'54/ATB train positions are available online from the train describers using track-free detection [13], but speed must be estimated based on the train number steps of the train describers or the underlying track section occupation messages [2]. Also on-board GPS systems may be used, connected to traffic control centres via e.g. GSM-R.

3.2 Traffic simulation through the scheduler

The scheduling module (AGLIBRARY) of the ROMA system [9] is used to determine train schedules in each dispatching area. These schedules can be the timetable if no delay is considered, or an updated plan of railway operations if delays are considered. The problem is viewed microscopically as a job shop scheduling problem with additional constraints. The mathematical formulation of [8] is based on the alternative graph of [18] (a generalization of a disjunctive graph), and on the blocking time theory [11] (used to compute minimum headways at the level of signals and track sections). This model is also rich enough to represent the detailed train movements within complex station areas [5].

The alternative graph is a triple $G = (N, F, A)$, where N is the set of nodes, F is a set of fixed arcs, and A is a set of pairs of alternative arcs. For fixed block signaling, each block signal corresponds to a node in G , and we name *operation* the passage of train over a block section or the dwelling of a train at a scheduled station stop. Each arc (i, j) , either fixed or alternative, is directed and weighted by a minimal time $weight_{ij}$ required between the start t_i of operation i and the start t_j of operation j . Arcs represent the interval times between operations, as described by the blocking time theory [9]. Hence, in general $t_j \geq t_i + weight_{ij}$.

Graph G models the train routes in a dispatching area along with their precedence constraints. Since a train must traverse the block sections on its route sequentially, a train route is modelled in the alternative graph with a job that is a chain of operations (modelled by nodes from the set N) and the associated precedence constraints corresponding to running and dwell times modelled by fixed arcs from the set F . In this work, a train route is given for each train and a running time run_i for each operation i is known in advance (similarly, a minimum dwell time $dwell_i$ for scheduled stops), except for possible additional waiting times between operations added to solve train conflicts. Additional fixed arcs from a start

node $0 \in N$ model the entrance of a train in the network or the departure from a scheduled stop, according to a minimum start time $ready_i$ associated. Finally, fixed arcs to an end node $n \in N$ model the exit of a train from the network or the arrival to a scheduled stop.

The variables of the train rescheduling problem are the start time t_i of each operation. Since a block section cannot host two trains at the same time, a conflict might occur and a sequence order between trains must be defined. This is modelled by a pair of alternative arcs $[(i, j), (h, k)] \in A$ from which exactly one must be chosen. Each alternative arc corresponds to an order in which the two trains enter the conflicting block section. The weight of the alternative arc (i, j) corresponds to (part of) the minimum time headway time h_{ij} between the corresponding pair of trains. If two consecutive trains run at different speeds (i.e. one fast train and one slow train), we use a sequence-dependent time headway [6].

A selection S of alternative arcs from A is obtained by choosing at most one arc from each alternative pair. The selection S is *complete* if exactly one arc is chosen from each pair. A *train schedule* is represented by an alternative graph solution $(N, F \cup S)$ in which S is a complete selection. The schedule is *feasible* if the graph contains no positive length cycles. This latter situation corresponds to an operation preceding itself, i.e., a system deadlock. The problem of deciding whether a feasible schedule exists or not is NP-hard [18].

This paper uses the following two rescheduling algorithms. A Branch and Bound algorithm (B&B), which is an exhaustive algorithm that explores all the reordering alternatives and chooses the one minimizing the maximum consecutive delay. Here we consider a truncated branch and bound [8] that returns near-optimal schedules for practical size problems within a short computation time. Furthermore, a simple First-Come-First-Served (FCFS) algorithm is used, which gives precedence to the train arriving first at a conflict point. The FCFS algorithm was used in the report for the Dutch Parliament [10].

3.3 Free-net speed trajectories and speed adjustments

The free-net speed trajectory module computes the free speed profile for each train assuming no train conflicts. The free speed profile then depends on the signalling system, block configuration, and scheduled stops. This free-speed profile is used to compute the minimum cycle time in the timetable compression method. Figure 4 gives an overview of the speed profile computed in the case study for NS'54/ATB (top) and ETCS L2 (bottom) with the same block configuration. For every signalling system the speed profile (in red) is plotted against the infrastructure with the associated block boundaries. For a description of the typical speed dynamics of the two signalling/ATP systems, see Section 2.1.

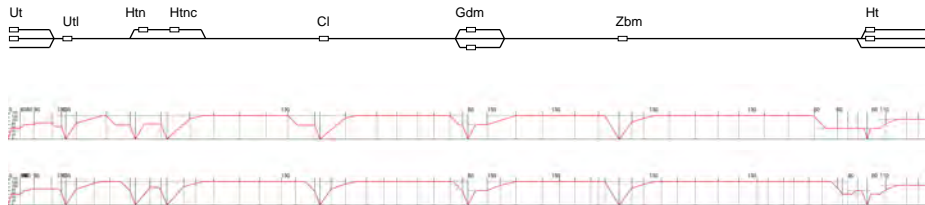


Figure 4: Schematic Sprinter speed profiles for NS'54/ATB and ETCS L2

The speed profile of trains is computed based on the maximum speed allowed by the operational signalling and ATP rules, except for a common running time supplement of

5%. For NS'54/ATB, the braking curves are calculated according to the signal aspects with a typical driving strategy. This driving style relates to accelerating with maximum force from standstill, cruising at maximum speed, and decreasing speed as soon as a signal or speed sign prescribes a reduced speed. For ETCS, the braking curve corresponds to the service brake intervention curve. The track description associated to the movement authority specifies the static speed profile defining the maximum train speeds. It is assumed that the driver starts braking at the indication points, which means that the train slows down at the last moment, thus saving time compared to a traditional signalling system. Note that in ETCS a local restricted speed on a switch or bridge within a block section is also taken into account in the dynamic speed profile. After a switch or bridge the train accelerates again as soon as the train has passed the object with its full length or when the train enters a new block section.

During operations, trains do not always follow unhindered movements. Due to delays, varying driving behaviour, longer dwell times, and overall failures, the signalling system might actually show restrictive aspects (yellow and red signals) to the trains. While computing the scheduling solution, only a technical part of the minimum headway is considered, excluding the approach time. Apart from the block running time this thus covers the setup, sight and reaction, clearing and release time. An optimal scheduling solution might thus contain trains that come within the approach distance of the preceding train. Therefore the module that adjusts train speeds is called to recompute feasible speed profiles respecting the actual speed restrictions, corresponding to e.g. yellow signals for the NS'54 signalling system and a limited movement authority for ETCS.

3.4 Evaluation of capacity consumption

We now describe the implementation of the UIC 406 compression method proposed in this paper, which is valid both for scheduled and disturbed situations. We consider in all cases the minimum headway, thus including the approaching, running, sight, reaction, clearing and switching times. We also recall that the order between trains (i.e., choice of which train goes first at shared elements) cannot be changed in this phase, but their timing only. So for the timetable compression we consider a complete selection of the alternative arcs.

A straightforward approach to perform timetable compression consists in the relaxation of all constraints and in the computation of a ready time for each operation i , apart from the entrance of the first train in the network which will be the reference for the capacity consumption. In fact, this approach needs to guarantee that operations of a single train are executed one after each other without any further interruption or delay between them, i.e. $t_i = \max(t_j + weight_j, t_k + h_{ki})$, with j the operation immediately preceding i on the train route (either a running arc run_j or a dwell arc $dwell_j$), and k an operation having a potential conflict with i over a shared block section. To achieve this, a complex no-wait extension of the alternative graph formulation can be adopted, as in Corman *et al.* [7].

This paper follows, instead, a simpler and more compact approach made possible by the fact that ready times $ready_i$ and minimum start times t_i have been computed already by the scheduler when capacity is evaluated, differently from [7], where they are decision variables. Define the aggregate process time $P(i, j)$ from an operation i to j on the same train route, and hence corresponding to the same train, as:

$$P(i, j) = \sum_k run_k + \sum_m dwell_m ,$$

where k (or m) runs over all the running (or dwell) operations from operation i to j over the train route associated to i . Then for every operation i with a ready time constraint, associated to e.g. an entrance in the network or departures from scheduled stops, the ready time is adjusted to the new value:

$$\overline{ready}_i = \max_{k,j,m} \{t_k + h_{ki}, t_j + P(j, i), t_m - P(i, m)\}$$

where k ranges over all operations with a headway constraint to operation i , j ranges over all the ancestors of i over the train route (with running and dwell operations) until the last overtaking location, and m ranges over the descendants of i over the train route until the next overtaking location. In the computation of ready times, both ancestors and descendants are thus relevant, on top of the headway constraints. For the process times from i to j we use the aggregate function $P(i, j)$ as it allows for an easy setup and a low computational complexity, since the dynamic capacity computation would result in a large number of iterative adjustments of the ready times if just the immediate predecessor (successor) is considered, depending on the amount of nodes.

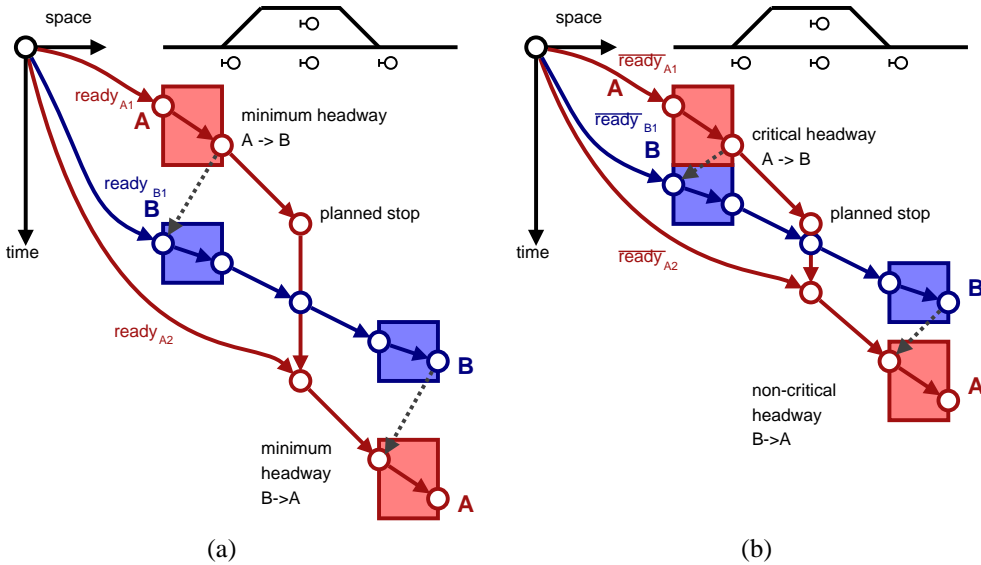


Figure 5: Uncompressed and compressed time-distance graph

Figure 5 illustrates the compression routine for an example based on the alternative graph formulation. In Figure 5 (a), the alternative graph is plotted, with its nodes arranged on a time-space axis. The node on the origin of the axes defines the reference time for all other nodes. Two trains (A and B) are running on a simple network, defining two chains of nodes. Solid arcs represent running constraints (oblique arcs), minimal dwell times (vertical arcs), or ready times when indicated so (namely, $ready_{A1}$ and $ready_{B1}$ refer to the entrance in the network of trains A and B; $ready_{A2}$ defines a minimum departure time for train A at a scheduled stop). Alternative arcs defining orders and headways between the trains are displayed as dotted arcs. Train B overtakes train A, while the latter has a scheduled stop. In

the plot, the blocks corresponding to the minimum headway before and after the overtaking are highlighted.

Figure 5 (b) describes the result of our compression procedure. Orders between trains are considered fixed, as well as running times. One node acts as a reference and remains fixed in time, in this case the earliest node of train A associated with $\overline{ready}_{A1} = ready_{A1}$. Train paths are shifted upwards until the critical headway with their predecessor is reached. Updated ready times \overline{ready}_{B1} and \overline{ready}_{A2} are computed, based on the given formula. Minimum headways, such as $A \rightarrow B$, can result in being the critical headway, i.e., those that are actually limiting a train path from being shifted further earlier in time.

Figure 5 also describes a cyclic relation between the two trains, due to the overtaking. Before the overtaking, train A constrains train B to keep distance. After the overtaking, train B constrains the start time of train A from the scheduled stop. While compressing the train paths, and investigating the headway constraint $B \rightarrow A$, train A is pushed upward, but this action is prevented by the other headway constraint $A \rightarrow B$ on the previous section together with the preceding fixed running times and minimum dwell time. Depending on the running times and the two headway times, it might be that only a single headway constraint ($A \rightarrow B$) becomes critical, while the other ($B \rightarrow A$) is not.

The *minimum cycle time* can be computed as the time difference between two operations describing the same service in successive time periods. When a stochastic variable is sampled into a set of realizations, the minimum cycle time is computed as the average of the minimum cycle times of the different realizations.

4 Computational experiments

This section presents the computational experiments. The simulation and optimization routines have been written in C++ and are able to run on Windows and Linux. The tests reported in this section were run on a 64-bit, Intel i5 processor at 2.67 GHz, under Windows 7. For the evaluation of the various signalling/ATP variants, we use the extended ROMA system. For the disturbed condition the train traffic is simulated over a period of two hours, which includes one warm-up period and one hour of full detail on train traffic flows.

4.1 Case study Utrecht-Den Bosch

We studied the corridor Utrecht - Den Bosch (see Figure 6) for the basic (morning peak) hour timetable 2011. The corridor has a length of about 40 km and contains six intermediate stops including the larger station Geldermalsen (Gdm), about halfway in the corridor depicted in Figure 6, where overtaking is possible. The railway corridor is operated by mixed traffic with ten trains (per direction) in a basic hour pattern including four Intercity (IC) trains, four regional trains called Sprinter (Spr), and two freight train paths (F). The Intercities stop in Utrecht (Ut) and Den Bosch (Ht) only, while the Sprinters serve all stops until Geldermalsen after which two Sprinters continue to Den Bosch with another intermediate stop and two Sprinters divert to another corridor. The freight trains pass Utrecht and Den Bosch without stopping; one freight train path has a stop in Geldermalsen to allow overtakings while the other runs non-stop. In the study only the direction from Utrecht towards Den Bosch is considered. A minimum dwell time of 24 s is considered for short stops on open tracks, 1 minute in Geldermalsen, and 2 minutes in Utrecht and Den Bosch.

The infrastructure characteristics and (platform) track allocation of 2011 were provided



Figure 6: Schematic track layout of railway line Utrecht-Den Bosch

by ProRail. The infrastructure characteristics come from the InfraAtlas and consist of an accurate description of all track sections, points, speed signs, and signals over the entire track layout from Utrecht until Vught Aansluiting (after Den Bosch). For the blocks of the ETCS L2 Short Block variant we matched as close as possible the block boundaries and projection rules of the ProRail study [21], which was however based on the 2009 timetable and infrastructure. Since then the station layout of Houten was adjusted and the Sprinters have an additional stop in Houten Castellum (Htnc). For the static speed profiles of the railway line, see Figure 4. The infrastructure data did not contain gradients so we assumed flat infrastructure. The rolling stock characteristics for the passenger trains were provided by NS. The IC's are assumed to use VIRM-IV rolling stock and the Sprinters use SLT's. For the freight trains, typical values were used: a length of 550 m, a constant braking rate of 0.3 m/s^2 , and an approximate linear acceleration of 0.2 m/s^2 for speeds up to 60 km/h, and 0.18 m/s^2 for speeds from 60 km/h to 80 km/h (the maximum speed for the freight trains).

Figure 7 shows a sketch of the traffic pattern in the basic hour pattern 2011 with Intercities (blue), Sprinters (green) and freight trains (red). The Intercities overtake the Sprinters in Geldermalsen, and they also overtake one freight train.

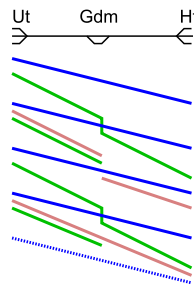


Figure 7: Scheme of the basic hour pattern for the 2011 timetable

Four signalling variants are considered in the experiments:

- (a) NS'54/ATB: the existing NS'54 signalling system with ATB train protection;
- (b) NS'54/ATB Opt: NS'54/ATB system with optimized blocks (near stations);
- (c) ETCS L2: ETCS Level 2 with current block section lengths;

(d) ETCS L2 SB: ETCS Level 2 with short block sections.

The configurations (a) and (c) are the reference cases for NS'54 and ETCS L2. In case (b), the optimized infrastructure is based on the optimized block projection of the ProRail report [21], see also Bartholomeus *et al.* [3]. This variant is included for a fair comparison of the benefits of ETCS over NS'54/ATB, and in particular to check whether the existing NS'54 block lengths configuration is not particularly bad.

In case (d), the short block sections refer to an infrastructure with block lengths of about 700 m with the block boundaries projected on the nearest existing track circuit border. This ETCS infrastructure is not optimized but rather it is easily obtained from the existing block projection through halving the blocks on the open track and using the existing track circuits. This implies that the ETCS L2 SB results are by no means the best that can be obtained using ETCS L2 but rather corresponds to a pragmatic implementation of ETCS L2 on the existing infrastructure. In practice, blocks of 400 m with either track circuits or axle counters could also be considered. The optimal choice of block lengths is out of the scope of this paper.

4.2 Scheduled capacity consumption

ROMA computes the capacity consumption for the railway line Utrecht-Den Bosch using the timetable compression method for the four signalling and ATP variants. Figure 8 shows the blocking time diagrams with the compressed timetables of the NS'54/ATB and ETCS L2 variants. Capacity consumption depends on blocking times and running time differences of the different train types. Changing the (signalling) infrastructure influences the technical running times, which on its turn affects capacity consumption by both modifying the running time differences and the blocking times (via the approach, running, and clearing times). To understand the differences between signalling variants we therefore not just report the capacity consumptions, but also the running times and minimum headway times. Specifically, we use the following indicators for the evaluation of the train schedules:

- *Running time*: the required time from the departure to the arrival of a train between a pair of stations, consisting of the technical free running time calculated for the given infrastructure and rolling stock plus for passenger trains 5% running time supplement.
- *Minimum headway time*: the minimum time separation between the departure or arrival times of two successive trains at a station associated to conflict-free train paths.
- *Minimum cycle time (MCT)*: the minimum time required to run all trains conflict-free according to the scheduled train sequence of the basic hour pattern.
- *Capacity consumption (Cap.)*: the ratio between the minimum cycle time and the planned cycle time (60 minutes) given in percentage.

Note that the minimum headway time depends on the critical block section, and it is influenced by the order of the trains, running time differences including intermediate stops, and the speed profiles. The minimum cycle time equals the sum of all minimum departure headway times from Utrecht (the minimum line headway times).

Running times

Table 1 shows the running times per train category (Intercity, Sprinter, and Freight trains) from Utrecht to Geldermalsen and from Geldermalsen to Den Bosch, including the dwell

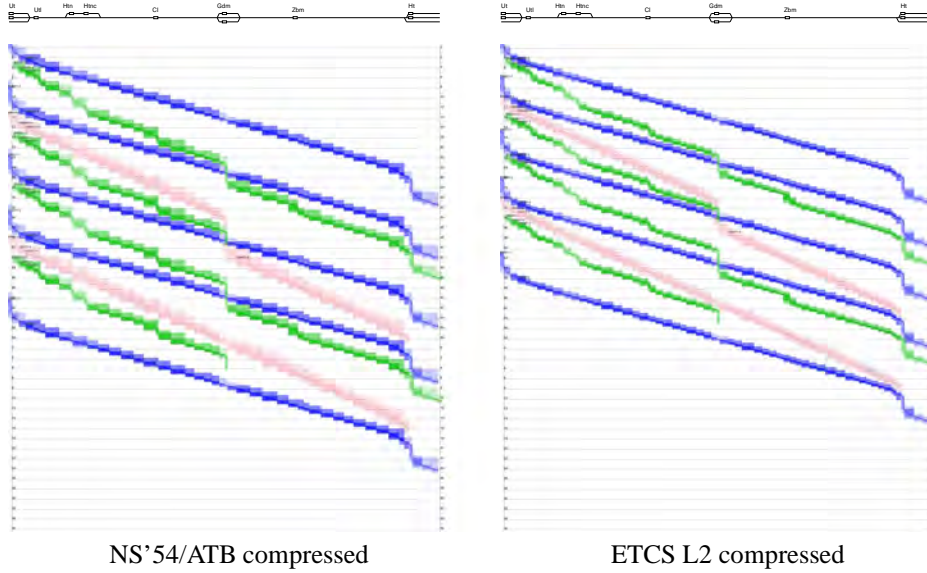


Figure 8: Scheduled capacity consumption according to timetable compression method

Table 1: Running times (scheduled conditions)

Scenario	Ut → Gdm			Gdm → Ht		
	IC	Spr	F	IC	Spr	F
NS'54/ATB	14'19"	20'25"	21'13"	12'27"	14'50"	18'18"
NS'54/ATB Opt	14'20"	20'30"	21'13"	12'25"	14'43"	18'15"
ETCS L2	14'19"	19'42"	21'12"	12'13"	14'30"	18'05"
ETCS L2 SB	14'20"	19'36"	21'12"	11'55"	14'11"	17'55"

times at intermediate stops for the Sprinter. The running times for the freight train are the ones with a stop in Geldermalsen. The numbers are given in minutes (') and seconds (").

The running times for NS'54/ATB and NS'54/ATB optimized differ only little by at most 7 s, so we compare the ETCS variants with the existing NS'54/ATB. ETCS L2 on the existing block sections leads to significant running time savings with respect to NS'54/ATB. The running time of the Sprinter is 43 s shorter on Ut-Gdm and another 20 s on Gdm-Ht, so that the Sprinter can run one minute faster from Utrecht to Den Bosch (besides the dwell time in Geldermalsen). The running time for the Intercity is only 14 s shorter in the last part before Den Bosch, and likewise the Freight train gains only 13 s on this last part before Den Bosch. The many stops of the Sprinter explain its higher time savings: before stops the trains with ETCS L2 can maintain a higher speed for a longer time before they have to brake due to the braking curve, that is train-dependent and independent from the block boundaries. When short blocks are taken into account the time savings of ETCS L2 are further improved to 88 s from Utrecht to Den Bosch for the Sprinter, 31 s for the Intercity, and 23 s for the Freight train.

Minimum headway times

Tables 2 and 3 give the minimum headway times at four locations: Departure Utrecht, Arrival Geldermalsen, Departure Geldermalsen, and Arrival Den Bosch. The minimum headway times for the Freight train are measured for the one stopping in Geldermalsen. The larger minimum headway times set in *italic* include a considerable running time difference on top of the critical blocking times. For instance, with NS'54/ATB the Intercity can only depart 7'30" after the Sprinter from Utrecht. This minimum headway consists of a minimum headway time of 1'24" due to the critical block just before Geldermalsen (see arrival headway at Geldermalsen) and the speed difference of 6'06" between the Intercity and Sprinter from Utrecht to Geldermalsen. In the reverse order, the Sprinter cannot depart earlier than 3'07" after the Intercity from Utrecht otherwise it will have to brake due to a yellow signal. The critical block is just after Utrecht. The minimum headway times for NS'54/ATB Optimized are similar to NS'54/ATB with at most 10 s improvement. When a running time difference is included, the minimum headway times may also be slightly worse, up to 5 s. Hence, we compare the ETCS variants with the existing NS'54/ATB.

Table 2: Minimum headway times on Utrecht-Geldermalsen (scheduled conditions)

Scenario	Departure Ut				Arrival Gdm			
	IC→Spr	Spr→IC	IC→F	F→Spr	IC→Spr	Spr→IC	IC→F	F→Spr
NS'54/ATB	3'07"	<i>7'30"</i>	2'49"	<i>3'41"</i>	<i>9'13"</i>	1'24"	<i>9'43"</i>	2'29"
NS'54/ATB Opt	2'58"	<i>7'34"</i>	2'49"	<i>3'32"</i>	<i>9'08"</i>	1'24"	<i>9'42"</i>	2'25"
ETCS L2	2'12"	<i>6'32"</i>	1'39"	<i>3'04"</i>	<i>7'35"</i>	1'09"	<i>8'32"</i>	1'10"
ETCS L2 SB	2'13"	<i>6'20"</i>	1'37"	<i>3'09"</i>	<i>7'29"</i>	1'04"	<i>8'29"</i>	1'09"

Table 3: Minimum headway times on Geldermalsen-Den Bosch (scheduled conditions)

Scenario	Departure Gdm				Arrival Ht			
	IC→Spr	Spr→IC	IC→F	F→IC	IC→Spr	Spr→IC	IC→F	F→IC
NS'54/ATB	1'42"	<i>11'55"</i>	1'42"	<i>9'25"</i>	<i>4'05"</i>	9'32"	<i>7'33"</i>	3'34"
NS'54/ATB Opt	1'37"	<i>11'55"</i>	1'26"	<i>9'21"</i>	<i>3'55"</i>	9'37"	<i>7'16"</i>	3'31"
ETCS L2	1'34"	<i>9'21"</i>	1'34"	<i>8'14"</i>	<i>3'51"</i>	<i>7'04"</i>	<i>7'26"</i>	2'22"
ETCS L2 SB	1'34"	<i>9'05"</i>	1'34"	<i>7'57"</i>	<i>3'50"</i>	<i>6'49"</i>	<i>7'34"</i>	1'57"

There are seven critical minimum headway times per variant that do not include a running time difference (set in normal font in Tables 2 and 3) which will be discussed in more depth in the following. The critical minimum headway times decrease significantly with ETCS L2 relative to NS'54/ATB. In particular, the headway times decrease at departure from Utrecht of the IC after the Sprinter (55 s) and freight train (70 s), and at the arrivals of the passenger trains after the freight train with the Sprinter in Geldermalsen (79 s) and the IC in Den Bosch (72 s). With short blocks the latter minimum arrival headway time at Den Bosch decreases by more than 1.5 minute (97 s). The departure headway reductions can be explained largely by the smaller approach times of ETCS L2 trains, due the smaller train-dependent braking distance at the lower speeds of departure, independent of the block lengths. The arrival headway reductions can be explained also by the later braking according to the ETCS dynamic speed profile just before the points of restricted speeds and the

End of Authority. This way the trains run at a higher speed for a longer time, decreasing the approach, running and clearing times. Note that the running time savings correspond to the locations with reduced arrival headway times. The shorter blocks decrease the block running times further. The overall reduced minimum headway times would lead to less capacity consumption for the ETCS trains.

Computation of capacity consumption

Table 4 gives the results obtained in terms of minimum cycle time and capacity consumption (in percentage), considering three timetable variants with zero, one (with a scheduled stop in Geldermalsen), or two (with and without a scheduled stop) freight train paths. The values for NS'54/ATB Optimized are slightly better than the existing block configuration, with minimum cycle times about 30 s lower, resulting in an improved capacity consumption.

Table 4: Minimum cycle time and capacity consumption (scheduled conditions)

Scenario	2 freight paths		1 freight path		0 freight path	
	MCT (s)	Cap. (%)	MCT (s)	Cap. (%)	MCT (s)	Cap. (%)
NS'54/ATB	52'58"	88.3	45'36"	76.0	41'22"	68.9
NS'54/ATB Opt	52'25"	87.4	45'01"	75.0	41'02"	68.4
ETCS L2	44'43"	74.5	37'53"	63.1	34'02"	56.7
ETCS L2 SB	43'40"	72.8	36'51"	61.4	33'01"	55.0

As expected from the evaluation of the minimum headway times, for the case with two freight paths the minimum cycle time decreases considerably with ETCS L2 by around eight minutes (a decrease of 16%-18%) when compared to NS'54/ATB. With short blocks the minimum cycle time decreases one minute more. The associated capacity consumption reduces with 13.8% and 15.5%, respectively.

Table 4 also shows that the capacity consumption drops quickly if one (or zero) freight path is scheduled. Going from two freight paths to one freight path reduces capacity consumption by around 12%, while going further from one to zero freight path leads to a lesser decrease by around 7%, which can be explained by decreasing heterogeneity (speed differences). Short blocks result in minor savings for ETCS compared to the existing blocks, which can be explained by the fact that the block lengths in station areas are already quite optimized as the small improvements for the optimized NS'54/ATB show.

For two freight trains, the capacity consumption of the existing NS'54/ATB is very high (88.3%) and exceeds the UIC-norm for peak hours considerably (of 75%). In practice, this means that two freight paths per hour would lead to unstable railway operations, where delays propagate easy. Under ETCS L2 the capacity consumption is 74.5% which does satisfy the UIC-norm, and with short blocks it drops further to 72.8%. Hence, with ETCS L2 the existing timetable with two freight paths is very well possible, but under the existing signalling system there would be delays rapidly. After inquiry with ProRail we obtained the freight path realizations from which became clear that in practice only one freight train should run per two hours. For one freight train the capacity consumption with NS'54/ATB is with 76% just above the UIC-norm, without freight path the capacity consumption is with 68.9% no more problematic.

4.3 Dynamic capacity consumption and delay minimization

For disturbed conditions trains are delayed and they may run at reduced speeds due to signal restrictions. Consequently, traffic controllers may reschedule the timetable to reduce the delays. The performance of the railway traffic system under disturbances depends on the different signalling and ATP systems, as well as the traffic control decisions. Therefore, in this section we compute the capacity consumptions and delays under disturbed conditions for the original timetable with two freight paths.

A *train delay* is the (positive) difference between the actual (or simulated) arrival time and the scheduled one. In this work, delays are divided into two parts: the *initial delays* of trains, that are due to the entrance disturbances, and the *consecutive delays*, that are caused by the interaction of trains running in the network. The latter delays can be linked to longer dwell times, reduced speeds and unscheduled stops in response to restricted signals. The goal of train rescheduling is to minimize the consecutive delays.

We consider the signalling variants under various delay scenarios. In the disturbed traffic conditions, trains are given an initial delay when entering the network and rescheduling is simulated by two types of traffic management algorithms: simple traffic management according to First-Come-First-Served (FCFS) rule, and advanced re-ordering and re-timing decisions computed by the ROMA Branch-and-Bound (B&B) algorithm. For both the train scheduling algorithms, the train speeds are adjusted to obtain feasible movements with respect to the signalling constraints.

To evaluate the behaviour of the various signalling variants and scheduling algorithms, we compute both delay statistics and capacity consumption for each delay scenario using the following indicators:

- *Average delay (AD)*: the average of all arrival delays at all scheduled stops in stations.
- *Maximum consecutive delay (MCD)*: the maximum of all consecutive delays.
- *Average consecutive delay (ACD)*: the average of all consecutive delays.
- *Punctuality (Punc.)*: percentage of punctual arrivals (with three minutes tolerance).
- *Dynamic minimum cycle time (MCT)*: minimum cycle time for one hour of traffic, for the delayed and (possibly) rescheduled trains.
- *Dynamic capacity consumption (Cap.)*: the ratio of dynamic minimum cycle time and planned cycle time (one hour).

The dynamic minimum cycle time is computed by compressing the rescheduled train paths for the trains scheduled in the second hour of the simulations (after one hour warm-up). This corresponds to the minimum cycle time of one hour of traffic with disturbances and rescheduling (by FCFS or B&B). This indicator represents the capacity that is actually used for running trains during operations. The dynamic capacity consumption is the capacity used by trains when dealing with disturbances and rescheduling. These can be compared with the scheduled capacity consumption given in the previous subsection. The reported performance indicators are the averages over a given set of traffic simulations.

For the disturbed conditions we modelled the train entries into the network (i.e., approaching Utrecht) as random initial delays. Specifically, we modelled the entries at Utrecht by three-parameter Weibull distributions. The Weibull distributions were fitted using realizations data of Utrecht, with separate estimated parameters for each train type (Intercity, Sprinter, and Freight train), see Figure 9. The results are averages over 50 delay instances according to a Monte Carlo simulation setup. The delay instances have a maximum delay

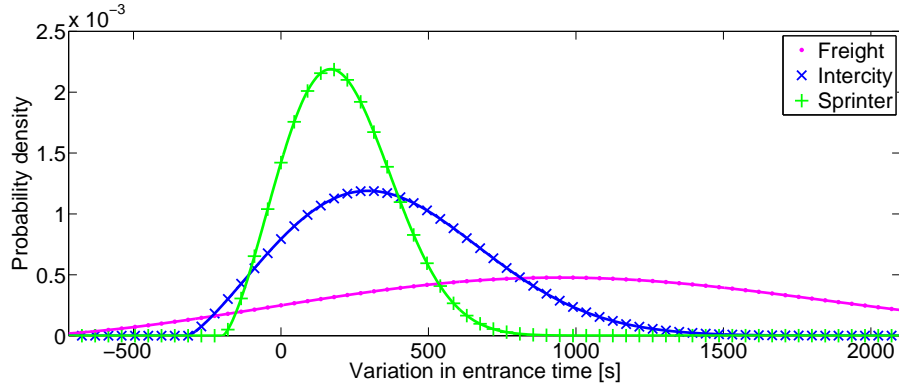


Figure 9: Distributions of initial delays

of 832 s (13'52''), a mean delay of 30.9 s, and 12% of the delays exceed 3 minutes. The passenger trains do not depart early from Utrecht, but the freight trains (running non-stop through Utrecht) may leave early the railway line Utrecht-Den Bosch. For more details on these distributions, we refer to Corman *et al.* [5].

Table 5 reports the results of the performance indicators averaged over the 50 traffic simulations for the four signalling/ATP variants. Note that each simulation corresponds to a specific initial delay instance and rescheduling by either the FCFS or B&B algorithm.

Table 5: Performance indicators under disturbed conditions (case with two freight trains)

Scenario	MCD (s)	ACD (s)	AD (s)	Punct. (%)	MCT	Cap. (%)
FCFS						
NS'54/ATB	617	21.6	59.2	89.7	52'47''	88.0
NS'54/ATB Opt	607	20.5	57.8	89.7	50'49''	84.7
ETCS L2	566	13.7	48.1	91.6	44'06''	73.5
ETCS L2 SB	561	13.3	47.5	91.6	43'24''	72.4
B&B						
NS'54/ATB	365	16.3	53.8	91.7	53'46''	89.6
NS'54/ATB Opt	356	15.0	52.4	91.9	51'34''	86.0
ETCS L2	253	7.6	42.0	93.7	44'07''	73.5
ETCS L2 SB	260	7.4	41.6	93.8	43'27''	72.4

When comparing the dynamic capacity consumption of Table 5 with the scheduled capacity consumption of Table 4 (for two freight trains), we see that the capacity consumption is generally lower in the disturbed cases. This can be explained as a consequence of the ROMA conflict resolution which adjusts train speeds so that train paths become more homogeneous, i.e., the speed difference of a train is adjusted to its preceding train (after possible re-timing and re-ordering) if a conflict occurred. It is well recognized that homogeneous traffic has less capacity consumption.

In the case of ETCS L2 the length of the approach distance depends on speed, so that a reduced train speed leads to smaller blocking times and to smaller minimum headway

times, if the reduced speeds occur at critical blocks. This also leads to a smaller dynamic minimum cycle time and thus to reduced capacity consumption. There is one exception: the dynamic capacity consumption of NS'54/ATB with B&B rescheduling is 1.3% percentage points higher than the scheduled one. This can be explained by the increasing blocking times when trains have to brake and run at lower speeds. However, the B&B solution is also 1.6% higher with respect to the simple FCFS rescheduling for this case. This can be explained by the objective function of the ROMA scheduling algorithms, that minimize the maximum consecutive delay (MCD). Note that the MCD of the B&B solution is much lower than the FCFS solution. So although the dynamic minimum cycle time gets increased by 48 s, the maximum consecutive delay gets reduced by 4'12" over the FCFS solution. Obviously, minimizing delays does not necessarily lead to minimized capacity consumption.

The delay results for the existing NS'54/ATB are similar to the optimized configuration, but the dynamic capacity consumption (and minimum cycle time) differ a lot with NS'54/ATB optimized having a considerable decrease in capacity consumption, again more for FCFS (-2.7%, corresponding to 96 s less MCT) than B&B (-1.4%, corresponding to 51 s MCT), although again B&B reduces the MCD by more than four minutes less than the FCFS solution. Apparently the optimized blocks near stations lead to different solutions and reduced capacity consumption. Overall, the capacity consumption of the NS'54/ATB variants seem much more sensitive to disturbances, while the performance of the ETCS L2 variants stay close to the scheduled ones.

The performance of the rescheduling algorithms FCFS and B&B are best measured in the delay and punctuality indicators. With respect to these indicators, both NS'54/ATB variants are comparable, and so are both ETCS variants. However, both algorithms perform better with the ETCS variants than the NS'54/ATB variants, with B&B clearly the winner. With B&B, the average consecutive delay (ACD) of the ETCS variants is over 50% less compared to the NS'54/ATB variants, and the maximum consecutive delay (MCD) is 96 to 112 s smaller. The average total delay improves by 9% to 13% when B&B is applied instead of FCFS. Also, punctuality is better for ETCS than NS'54/ATB by 2% points for both rescheduling algorithms. The advanced B&B resolution algorithm succeeds in improving punctuality by 2% more than FCFS.

Overall, the punctuality of NS'54/ATB with FCFS rescheduling can be improved with 4% from 89.7% to 93.8% with ETCS and B&B, demonstrating the usefulness of both ETCS and advanced traffic management systems, and their integration. A final remark is on the assumption that accurate train speed information is available for real-time traffic management. For ETCS L2 this assumption is justified since ETCS L2 trains communicate their speeds at regular intervals to the traffic control centre via the RBC. On the other hand, for NS'54/ATB the train speeds are not directly available so that real-time traffic management is much more difficult to realize.

5 Conclusions

This paper presents a thoughtful assessment of the potential gain achievable by the combination of next generation signalling and train dispatching systems. The main contribution is the quantification of the scheduled and dynamic capacity consumption, and the possibility to reduce delay propagation. The resulting methodology extends the usual capacity consumption assessment method in order to evaluate disturbed traffic conditions. To this aim, the ROMA system has been adopted and a dynamic timetable compression method has been

proposed for calculating capacity consumption. The proposed concept of dynamic capacity consumption is to capture the effect of the variability of train dynamics and the effect of rescheduling in presence of traffic disturbances. The method has been applied to a practical case study of the Dutch Utrecht-Den Bosch railway line.

The results of our capacity assessment study show that the scheduled capacity consumption decreases considerably from the Dutch speed signalling system NS'54/ATB to the state-of-the-art ETCS Level 2. The capacity consumption gap between NS'54/ATB and ETCS Level 2 increases when delays are considered. This can be explained by the increasing blocking times for NS'54/ATB when trains have to brake and run at lower speeds. For ETCS Level 2 the capacity consumption reduces when delays are considered: the braking distances shorten when trains run at lower speeds, which has a stabilizing effect on headway times and the throughput. This is a strong advantage of cab signaling systems with braking curve supervision such as ETCS L2, which it has not yet been fully recognized in the existing literature.

An integrated approach of ETCS L2 with a traffic management system like ROMA shows real potential to improve operational performance in disturbed conditions. We found that ETCS L2 combined with optimization-based rescheduling (ROMA's B&B algorithm) significantly increases performance compared to simple (FCFS) traffic management. This is, in fact, the novel contribution of this paper compared to the report [10] in which the improvement achievable by ROMA was not fully investigated.

Further research will be devoted to include rerouting and optimal speed control in the ETCS/ROMA integration, and evaluate the performance of ETCS in the presence of large disruptions like track obstructions and adverse weather conditions. A careful cost/benefit analysis has to be addressed before deciding on the investment of new expensive infrastructure and train equipment. The evaluation of advanced railway signalling systems via simulation and optimization can shed light on the benefits of these systems in practical conditions.

Acknowledgment

This work was funded by the Dutch Parliament for the investigation of railway maintenance and innovation in the Netherlands, and partially supported by the Italian Ministry of Research, Grant number RBIP06BZW8, project FIRB "Advanced tracking system in intermodal freight transportation". We thank NS and ProRail for providing the railway data.

References

- [1] Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P., Lova, A. (2008). An assessment of railway capacity. *Transportation Research Part E*, 44 (5), 774–806.
- [2] Albrecht, T., Goverde, R.M.P., Weeda, V.A., Van Luipen, J. (2006). Reconstruction of train trajectories from track occupation data to determine the effects of a driver information system. In: J. Allan, C.A. Brebbia, A.F. Rumsey, G. Sciutto, S. Sone, C.J. Goodman (Eds.), *Computers in Railways X*, WIT Transactions on The Built Environment, 88, WIT Press, Southampton, 207–216.
- [3] Bartholomeus, M.G.P., van Touw, B.L.J.C., Weits, E.A.G. (2011). Capacity effects of ERTMS Level 2 from a Dutch perspective. *Signal+Traht*, 103 (10), 34–40.

- [4] Caimi, G., Fuchsberger, M., Laumanns, M., Lüthi, M. (2012) A model predictive control approach for discrete-time rescheduling in complex central railway station areas. *Computers & Operations Research*, 39 (11), 2578–2593.
- [5] Corman, F., D’Ariano, A., Pranzo, M., Hansen, I.A. (2011). Effectiveness of dynamic reordering and rerouting of trains in a complicated and densely occupied station area. *Transportation Planning and Technology*, 34 (4), 341–362.
- [6] Corman, F., Goverde, R.M.P., D’Ariano, A. (2009). Rescheduling dense train traffic over complex station interlocking areas. In: R.K. Ahuja, R.H. Moehring, C.D. Zorliagis (Eds.), *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*, LNCS 5868, Springer, Berlin, 369–386.
- [7] Corman, F., D’Ariano, A., Pacciarelli, D., Pranzo, M. (2009). Evaluation of green wave policy in real-time railway traffic management. *Transportation Research, Part C*, 17 (6), 607–616.
- [8] D’Ariano, A., Pacciarelli, D., Pranzo, M. (2007). A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183 (2), 643–657.
- [9] D’Ariano, A., Pranzo, M. (2009). 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.
- [10] Goverde, R.M.P., Hansen, I.A., Corman, F., D’Ariano, A., Trinckauf, J. (2012). Innovatie op het spoor en mogelijkheden van ERTMS in Nederland (Innovation on the railways and opportunities of ERTMS in the Netherlands, in Dutch). *Kamerstuk kst-32707-11*, Dutch Parliament, The Hague.
- [11] Hansen, I.A., Pachl, J. (Eds.) (2008). *Railway Timetable & Traffic: Analysis, Modelling, Simulation*. Eurailpress, Hamburg.
- [12] Hirai, C., Tomii, N., Tashiro, Y., Kondou, S., Fujimori, A. (2006). An algorithm for train rescheduling using rescheduling pattern description language R. In: *Computers in Railways X*, WIT Transactions on The Built Environment, 88, WIT Press, Southampton, 551–561.
- [13] Kecman, P., Goverde, R.M.P. (2012). Process mining of train describer event data and automatic conflict identification. In: C.A. Brebbia, N. Tomii, P. Tzieropoulos, J.M. Mera (Eds.), *Computers in Railways XIII*, WIT Transactions on The Built Environment, 127, WIT Press, Southampton, 227–238.
- [14] Landex, A. (2009). Evaluation of railway networks with single track operation using the UIC 406 capacity method. *Networks and Spatial Economics*, 9 (1), 7–23.
- [15] Lindner, T. (2011). Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. *Journal of Rail Transport Planning & Management*, 1 (1), 49–57.
- [16] Lindner, T., Pachl, J. (2010). Recommendations for Enhancing UIC Code 406 Method to Evaluate Railroad Infrastructure Capacity. In: *TRB 89th Annual Meeting Compendium of Papers*, Transportation Research Board, Washington.
- [17] Lusby, R., Larsen, J., Ehrgott, M., Ryan, D.M. (2013) A set packing inspired method for real-time junction train routing. *Computers & Operations Research*, 40 (3), 713–724.

- [18] Mascis, A., Pacciarelli, D. (2002) Job shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research*, 143 (3), 498–517.
- [19] Medeossi, G., Longo, G., de Fabris, S. (2011) A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management*, 1 (1), 1–13.
- [20] Pachel, J., 2002. *Railway Operation and Control*. VTD Rail Publishing, Mountlake Terrace.
- [21] ProRail (2010). *Technische vergelijking tussen NS54 ATB-EG en ERTMS Level 2 (Technical comparison between NS'54 ATB-EG and ERTMS Level 2, in Dutch)*. Technical report EDMS 2472676, ProRail, Utrecht, 26 januari 2010.
- [22] Rodriguez, J. (2007). A constraint programming model for real-time train scheduling at junctions. *Transportation Research Part B*, 41 (2), 231–245.
- [23] Theeg, G., Vlasenko, S. (Eds.) (2009). *Railway Signalling & Interlocking: International Compendium*. Eurailpress, Hamburg.
- [24] Törnquist, J. (2007). Railway traffic disturbance management: An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon. *Transportation Research Part A*, 41 (3), 249–266.
- [25] UIC (2004). *UIC Code 406: Capacity*. International Union of Railways, Paris.
- [26] UIC (2008). *Influence of ETCS on the line capacity: Generic study*. International Union of Railways, Paris.
- [27] UIC (2010). *Influence of the European Train Control System (ETCS) on the capacity of nodes*. International Union of Railways, Paris.
- [28] Winter, P. (Ed.) (2009). *Compendium on ERTMS*. Eurailpress, Hamburg.