Journal of Rail Transport Planning & Management 28 (2023) 100417

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

Optimizing rollout strategies for migration to moving block signaling – A MINLP-based approach for on-board train integrity monitoring technology

Jakob Geischberger [a,*], Alessa Isberner [b], Norman Weik [a,c]

[a] German Aerospace Center, Institute of Transportation Systems, Lilienthalplatz 7, 38108, Braunschweig, Germany
[b] German Aerospace Center, Institute of Transportation Systems, Rutherfordstr. 2, 12489, Berlin, Germany
[c] Technical University of Munich, TUM School of Engineering and Design, Department of Mobility Systems Engineering, Professorship Design and Operation of Public Rail Transport Systems (RTS), Parkring 35, 85748, Garching, Germany

ARTICLE INFO

Keywords:
Railways
On-board train integrity
Network migration
Optimization

ABSTRACT

Increasing demand on heavily-used rail corridors in line with the modernization of the signaling architecture are key drivers for migrating to modern, moving-block based train control in the European railway network. In order to maximally profit from the increase of reliability and reduction of costs associated with shifting towards full ETCS Level 3 from a network management perspective, additional requirements on the fleet management level arise. Amongst other things, if track vacancy detection equipment is to be eliminated, all trains operating on these lines need to be equipped with on-board train integrity (OTI) monitoring solutions. In order to facilitate the planning of the OTI network migration processes, a MINLP-model is proposed which allows economic optimization of OTI migration in view of fleet allocation and the removal of trackside equipment for train integrity verification within the network. The model is tested in a case-study based on a generic network abstracted from the Austrian mainline network and found to significantly enhance planning compared to heuristic migration strategies.

1. Motivation

Increasing demand for rail on corridors with high capacity utilization as well as the obsolescence of currently operated class-B signaling systems have increased pressure on European railway network operators to accelerate the migration process to modern, digital ETCS signaling technology. With respect to the projected growth of the overall rail traffic volume associated with political actions to reduce the carbon foot print in transportation (see, e.g. European Rail Master Plan (Europe's Rail Joint Undertaking, 2021)), the increase of transport efficiency which can be achieved with moving-block train operation is of particular interest.

The shift of rail traffic control from fixed-block operation to moving-block signaling marks a key transition from trackside-systems for train occupation detection to communication-based and on-board systems for train localization and train integrity verification. As a result, failure risks resulting from wear of weather-exposed trackside occupation detection systems can be reduced and the reliability of operations is increased.

To fully exploit the potential of moving-block operations and to dismantle track-side occupancy detection systems on lines

* Corresponding author.
E-mail address: jakob.geischberger@dlr.de (J. Geischberger).

https://doi.org/10.1016/j.jrtpm.2023.100417
Received 11 May 2023; Received in revised form 26 September 2023; Accepted 6 October 2023
Available online 26 October 2023

equipped with ETCS Level 3, all trains servicing these lines need to be equipped with on-board technology for train integrity verification. This poses considerable effort associated with the equipment of vehicles, especially for freight transport, as freight trains have varying compositions and therefore migration needs to be planned and executed on a wagon-based rather than a train-based level. Consequently, a gradual upgrade or replacement of rolling stock without OTI capabilities spanning several years or decades will be required. As a result of technological challenges in the migration process (EEIG ERTMS Users Group, 2022), transitory solutions, such as e.g. ETCS Level 3 Hybrid system design (Vergroesen, 2020; Ranjbar et al., 2022) have been proposed.

While the overlay of signaling architectures in transitory ETCS Level 3 Hybrid is providing additional track capacity by making use of the capabilities provided by trains equipped with OTI technology (especially passenger trains), track-side infrastructure (e.g. axle counters and track circuits) is further required for operating non-integral trains. As these infrastructure elements require high maintenance and investment costs, an economic focus on reducing or fully removing them in the course of the shift from track-to train-side technology emerges. Otherwise, the increase of capacity does not go along with the full potential of reduced network operating costs and does not allow to fully profit from increased traffic reliability by reducing failure risks from weather-exposed trackside systems. Moreover, the ETCS Level 3 Hybrid system configuration could yield an incentive to increase block section lengths over Level 2 in order to decrease operating and maintenance costs. Due to longer fixed-block sections w.r.t ETCS Level 2, Hybrid Level 3 could pose negative effects on capacity whenever a train (even if OTI-equipped) follows another one not equipped with OTI in this case.

In the present paper, we focus on the joint optimization of fleet allocation and network migration processes to full moving-block operation without track vacancy detection equipment. Given restrictions on the number of non-integral train sets which can be replaced by or upgraded to trains with OTI technology, a MINLP (mixed-integer nonlinear program) is developed which allows to economically optimize transition processes based on a stable, optimal allocation of trains with OTI technology - favoring cost savings due to the removal of trackside equipment as early as possible without restricting the operating program. The proposed methodology contributes to the following subjects:

- Support of decision making within practical migration planning of OTI (on-board train integrity) fleet equipment
- Quantifying the respective decisions, more precisely which trains/train categories to equip with the proposed technology at what time, given restricted workshop capacities
- Comparison of different heuristic migration strategies to help pondering between different migration strategies (e.g. focus on corridors or lines with high amounts of trackside elements)
- Deepening the scientific understanding of how different optimization and heuristic approaches work by the example of migration optimization

The paper outline is as follows: section 2 presents the current state of OTI technology as well as related work to put the proposed approach into context. Next, the methodological approach is explained and proposed in section 3. In section 4, the method is tested in a case study based on a generic network resembling the Austrian network and results are assessed by comparison to heuristic migration strategies in network planning to pinpoint the optimization potential. Section 5 summarizes the analysis and gives an overview of future research potential.

2. State of technology and related work

The prevalent track vacancy detection system serves the purpose of locating and determining the infrastructure possession of trains. In current systems, this is achieved relying on trackside elements, such as axle counters or track circuits. Wear resulting from weather or traffic results in high maintenance costs, requiring both capital and operational expenditure at trackside level. In the transition towards ETCS Level 3, train localization and train integrity verification will be performed by on-board technology.

Depending on the available communication technology as well as the availability of energy supply on the train, the OTI technology solutions differ. Therefore, in the course of two Shift2Rail projects (X2Rail-2 D.4.2, 2020), three different OTI product classes have been defined to develop tailored solutions for the requirements per train type (X2Rail-2 D.4.1, 2020). The OTI product classes describe the different design of OTI functional architecture in terms of three different approaches. In particular, a distinction is made between the type of communication technology, the energy supply, the type of odometry, as well as the OTI architecture itself.

Product class 1 is being developed for train compositions with a wired communication network where the integrity is monitored based on the communication liveliness between the head (leading cabin) and the tail (last wagon) of the train. For trains without wired communication technology, Product class 2 provides a solution where the integrity is determined based on comparing kinematic data of the leading cabin and the train tail (e.g., the coherent movement of the train tail in respect to the front cabin). Product class 2 and Product class 3 are applicable for variable train compositions. For Product Class 3, separation detection sensors are installed in each wagon and the train integrity is verified by communication liveliness as well as the separation monitoring of adjacent wagons (X2Rail-2 D.4.2, 2020).

Passenger trains with fixed train compositions and a wired communication network are applicable for Product class 1 solution. Fixed composition freight block trains can be applicable for Product class 1 as well, if a wired communication is available through innovations such as, e.g., equipment with DAC (Digital Automatic Coupling) technology. Single wagons are equipped with Product class 3 solution due to their nature of varying train composition and a lack of wired communication system. As train integrity verification is one important aspect of the migration of the command control and signaling it needs to be seen in the context of the rollout of ETCS Level 3 in Europe. OTI is an enabler and mandatory functionality for ETCS Level 3 and thus relevant for the objective of an increase in track capacity.

In the context of this study, we focus on the economic aspects of OTI migration, where the main motivation is the decrease of life-cycle costs (LCC) for track vacancy detection systems. As stated above, the shift of equipment from track to rolling stock train integrity components aims at causing lower maintenance costs compared to current systems. To gain the full benefit of moving-block operations and to reduce track-side occupancy detection systems on the lines, all trains using a route need to be equipped with on-board technology for train integrity verification. As different train categories require different OTI product classes (see above), a distinction between train types in the later performed optimization approach is implemented (cf. section 3 onwards). The focus of the study is on railway control and operations. It is noted that some track vacancy detection solutions (e.g. track circuits) also provide additional functionalities such as, e.g. broken rail detection, which are not discussed in the context of this paper. Still, the preservation of all necessary safety functions must, of course, be ensured. However, the presented approach could be adapted to other functionalities by adapting the problem formulation (see below).

Currently two different approaches can be identified for the European ETCS migration: First, corridor approaches, where the most important corridors are equipped with ETCS Level 3 technologies initially, and the rest of the network will be retrofitted at a later time. The second approach, by contrast, consists of completely converting the rail network to ETCS Level 3 in a relatively short time in the form of an overall program. Another important migration aspect is the definition of the time period during which dual equipment of vehicles and infrastructure is foreseen (this refers to the simultaneous presence of ETCS and Class B systems) (Obrenovic et al., 2006).

Due to the amount and variability of trains and wagons within the European train network, equipping trains with the described technology poses a very complex and extensive undertaking. Further, technology migration is defined as a migration which takes place when the renewal of a system has to be carried out without interrupting the operating process (Lackhove, 2014). Such systems are characterized by network-wide implications. This can also be due to a pronounced interface character to surrounding systems.

Analogously, the phase from the introduction of a technology to a full equipment of a fleet is referred to as migration phase. During this phase, both legacy and new technology are operated at the same time, often referred to as coexistence (Kawano et al., 2016). For many technologies this state poses technological, organizational or communicational challenges. In context of the OTI technology, a major restriction is that only once a majority of trains is equipped, positive effects on capacity could possibly be observed, while cost reduction due to removal of track-side elements may even require full equipment (compare section 3). As a result, migration planning and execution need suitable decision support methods in order to minimize or alienate the beforementioned negative effects.

Migration processes for railway control and signaling systems are typically the subject of strategic and political reflections involving many factors. To the best of our knowledge, mathematical optimization techniques (that are widespread for other operation research domains in railway engineering) have not been applied as a means of providing decision support for strategic ETCS network migration processes discussed in this paper. Comparable technology migration problems have, however, been discussed for sub-aspects involving other domains, such as, e.g., train communications.

In the context of communication systems, the related problem of migrating from GSM-R to higher communication standards (e.g. 5G FRMCS) is of particular interest: Wen et al. (2021) propose a methodical approach with an optimization model based on MOEA/D (Multiobjective Evolutionary Algorithm Based on Decomposition) in order to minimize migration costs for a network-wide data communication by example of the potential migration from GSM-R to LTE. While the mentioned methodology is applied on a smaller, synthetic network, further work focusses on the same technological migration for the Polish railway network utilizing simulation tools (Kochan and Gruba, 2018). Moreover, a parametrized file-based model with a stronger economic focus is presented (European Railway Agency, 2016).

In the context of European Train Control System (ETCS) migration, Lackhove et al. (2010) give an overview of possible migration paths, goals and strategies, quantifying different solutions and visualizing them as a Petri-Net. An optimization approach based on cost-benefit-relations is further proposed. All these solutions share quantitative approaches to model migration processes.

Similar to the aforementioned approaches, in this paper, we propose a methodology to optimize technology migration in the railway sector on the network level by quantifying the underlying processes. Extending previous approaches with a strong strategic

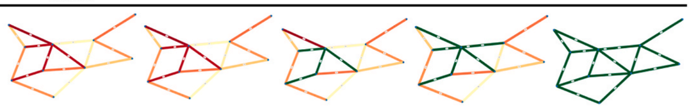
	Initial state	Year 1	Year 2	Year 3	Year 4
Network state					
Equipment rate (trains/ year)		100	100	100	100
Equipped trains	0	100	200	300	400
Unequipped trains	400	300	200	100	0

Fig. 1. Illustration of the train OTI migration and its impact on network migration. The illustration is based on a generic example based on 400 train sets and a total of 4 years.

focus by a tactical dimension by taking into account train rides on the level of single train and wagon units as well as distinct routes, an in-depth comparison is achieved. This unfolds economic implications of different migration strategies, enabling optimization on a tactical level by means of a MINLP.

3. Methodological approach

The methodological approach is described in the following. The goal of our MINLP approach is to support technology migration by optimizing the network-rollout of OTI-technology for ETCS Level 3 and abandoning of system trackside vacancy detection technology. In the course of a network-wide OTI-migration of trains, trackside elements can be removed step by step as track vacancy detection is shifted from trackside-to onboard-solutions. Being aware that some trackside elements – especially in station areas with shunting movements – cannot be replaced, this paper focusses on trackside elements on line segments. Taking an economic perspective as mentioned above, removed trackside elements mean a reduction of costs and should therefore be abandoned as early as possible in order to save costs. The overall idea of the migration process and methodology discussed in this paper is illustrated in Fig. 1. For modelling purposes, the underlying railway network is transformed into a mathematical graph. Edges are attributed with the number of trackside elements required for train occupation detection on this edge, which is referred to by different colors (red = high number of trackside elements, green = low number of trackside elements).

From left to right, successive steps in the migration process are depicted. Each year, a specific number of trains is equipped with OTI-technology. This leads to a constant rise in the share of equipped trains over the years. Depending on the routes of these trains within the network and when these are equipped, trackside elements can be removed on a certain line. The transformation of the underlying network infrastructure is not necessarily correlated to the steps in the train migration, as all trains on a given line need to be equipped with OTI, in order to be able to abandon track vacancy detection equipment. The goal is to migrate and allocate train sets in such a way that large subnetworks can be migrated early in order to save corresponding infrastructure and maintenance costs. This leads to an optimization problem, seeking to equip trains in such an order that the amount of trackside elements per edge can be removed as early as possible. The next subsection explains how this idea is translated into a MINLP.

3.1. MINLP model

The MINLP-model for optimization of the migration process is described in the following:

The objective function consists of two nested sums, where the inner sum adds up the amount $c_{e,t}$ of trackside elements over all edges $e \in E$ at a certain timestamp t, while the outer accumulates these inner network-wide trackside elements sums over all timestamps $t \in T := \{t_0, \ldots, t_N\}$, where $t_i := i^* \backslash \Delta T$ (defined as one year). By minimizing this nested sum, a reduction of trackside elements as early as possible is accomplished.

The trains within the network are assigned an id x and a category $cat \in CAT$ (the set of all existing train categories). By adding all trains of all categories, the total amount of trains X can be identified. This can be analyzed on the edge-level accordingly, where X_e is the set of trains running over edge e, equalling the sum of trains per category and egde over all categories as shown in (1).

Equation (2) is of core importance, as it models the correlation between equipment of trains with OTI and the reduction of trackside elements on edges. As described above, it is assumed that only once all trains that run over a certain edge e are equipped with OTI technology, all trackside elements on this specific edge e can be removed (however, other defined amounts would be theoretically implementable as well). This is mathematically implemented by means of an if-statement as follows: the binary OTI equipment status of train x, running over edge e at timestamp t, is defined as $o_{x_e,t} \in \{0;1\}$, where 0 means unequipped and 1 means equipped. Hence, if all trains running over edge e are fully equipped, the sum of their status variables $o_{x_e,t}$ will equal the total amount of trains assigned to edge e, $|X_e|$. Only if this is true, the amount of trackside elements $c_{e,t}$ on edge e can be removed and their amount set to 0. In any other case, the sum of equipment variables $o_{x_e,t}$ will be below the total amount of trains $|X_e|$, leaving $c_{e,t}$ at its initial value $c_{e,t} = c_{e,t=0}$ and thus prohibiting trackside element removal.

Constraint (3) assures that equipped trains stay equipped. Equation (4) restricts how many trains of a certain train category $cat \in CAT$ can be equipped in a defined time, by defining the sum of equipped trains at a certain timestamp t (cf. section 3.4).

$$\min \sum_{t=t_0} \sum_{e \in E} c_{e,t}$$

subject to

$$\sum_{cat \in CAT} X_{e,cat} = X_e \qquad e \in E, cat \in CAT \tag{1}$$

$$\left(\sum_{x_e \in X_e} o_{x_e,t} = |X_e| \rightarrow c_{e,t} = c_{e,t=0}\right) \bigwedge \left(\sum_{x_e \in X_e} o_{x_e,t} < |X_e| \rightarrow c_{e,t} = 0\right) \qquad o_{x_e,t} \in \{0;1\}, t \in T, e \in E \tag{2}$$

$$o_{x_e,t} \geq o_{x_e,t-1} \qquad o_{x_e,t} \in \{0;1\} \tag{3}$$

$$\sum_{t=t_0}^{t}\sum_{x_e\in X_e} o_{x_e,t} = r_{t,cat} \qquad o_{x_e,t}\in\{0;1\}, t\in T, e\in E, cat\in CAT \tag{4}$$

$c_{e,t}$ = amount of trackside elements on edge e at timestamp t

$c_{e,t=0}$ = initial amount of trackside elements on edge e at timestamp $t=0$

CAT = set of train categories

$X_{e,cat}$ = set of trains of category cat running over edge e

X_e = set of trains running over edge e

$|X_e|$ = amount of trains running over edge e

$o_{x_e,t}$ = OTI equipment status of train x running over edge e at timestamp t

$r_{t,cat}$ = accumulated amount of trains of category cat that can be equipped with OTI at timestamp t

3.2. Heuristic approaches used for assessment of the MINLP approach

In order to compare MINLP optimization results, additional equipment strategies are defined. In this context it is worth noticing that in general migration strategies might not be purely defined by quantitative or technological approaches as dealt with in this paper, but can also be driven by political or business-related strategic aspects. These can be included into the mathematical model (as e.g. has been done in the high-speed strategy below). The five chosen strategies include four related heuristic approaches, as well as a random equipment to perform benchmarking. Restrictions are identical to the MINLP described in the subsection above. Table 1 gives an overview over all approaches.

The heuristic approaches are defined by a loop, where trains are equipped subject to applying restrictions and the assumed time horizon for the migration. Therefore, a list of trains is generated where they are ordered according to their equipment priority. In the next step, these trains are equipped year by year, beginning from the start of the list until all trains are equipped. The equipment rates per train category and year (as explained in section 3.4. below) constrain this process and thus implement the dimension of time into the heuristic approach. The mentioned list can be defined with different strategies:

For the *least trains-strategy*, a loop-wise identification of the total amount of unequipped trains per edge is performed. The respective train priority list is set up in accordance to the route of trains over these edges, starting with trains running over edges with the least amount of unequipped trains. The explanatory equivalent of this strategy would be equipment according to least equipment costs (as the least amount of trains has to be equipped in order to effect removal of trackside elements).

Second, a strategy with special focus on *high-speed operation* is defined, where trains of this category are automatically set to the top of the equipment list. This approach is characterized by prioritizing high-speed traffic in comparison to other modes, as often seen in operational reality. As high-speed-operations are often concentrated to a few corridors, the strategy shows resemblance to the corridor strategies mentioned in section 2 in the context of ETCS migration. Apart from the mentioned aspects, the approach is identical to the least trains-setup.

A third heuristic approach is defined to depict the strategy of equipping trains running over edges with a *high number of track-side elements* first. The respective train equipment list is defined depending on how many track-side elements are on the according edges of their routes. This approach seeks high gains at first and is inspired by Greedy - Algorithms.

The *combined heuristic strategy* aims to merge the least trains and the trackside element strategy by simultaneously prioritizing their objectives. Therefore, the quotient of trackside elements per unequipped train is calculated and used as a target value for ordering the above-mentioned train priority list.

At last, a *random train equipment approach* is implemented to create a lower performance level for comparison purposes. This approach is expected to have only very few removed trackside elements in the beginning – as by chance all trains of a certain edge would have to be equipped – and therefore a rather steep fall to the end of the migration period. Results are averaged over 1000

Table 1
Different approaches.

Approach	Category
Network-wide optimization	MINLP
Least trains strategy	Heuristic
High-speed strategy	Heuristic
Trackside elements strategy	Heuristic
Combined strategy	Heuristic
Random equipment	Randomized

different randomized approaches.

3.3. Data acquisition and processing

The strategies shown above are operationalized in a Python environment. Application on network level requires methodological preparatory steps and assumptions, as well as data acquisition and processing as shown in Table 2:

Publicly available train count data on the level of granularity of line segments and train categories is used as main input for two purposes.

First, network information is derived in order to accomplish a mapping consistency between train operation and infrastructure. As this analysis is conducted on a network-wide level and hence incorporates a macroscopic perspective on railway operations, we refrain from processing more detailed input data on the level of single infrastructure elements typically used in simulation-based analyses with stronger operative focus. Instead, the network is defined as a mathematical graph on a node-edge-basis. As train count data can be very heterogenous in terms of observation areas (segments) and thus granularity, an aggregation of major nodes has been performed. In the course of this operational points within large cities have been merged. Further, the processed topological information is mapped with geo information for visualization purposes.

Second, specific train runs are derived from train count data, as the formulated optimization model is highly dependent on representative train routes and numbers and purely random or synthetical train operation would reduce expressiveness of results. In order to extract information of such quality from segment-wise train count data as best as possible, the same is further processed by means of the tool dfrouter (part of SUMO simulation tool (Alvarez Lopez, 2018)): initially developed for generating validated motorized traffic flow data for urban areas on the basis of car counting data, it can be easily applied to railway operations as well. If counting data for as many edges as possible within a network is provided, it algorithmically generates single trips (in the sense of single means of transports running a specific route from origin to destination), matching the input counting data by approximation. An important restriction in the context of train counting data is that only physical trains can be equipped with a technology, while counting data usually lists counted train rides, not physical trains. This means that, e.g., physical regional trains might have been counted multiple times within the observation time window – in contrast to freight trains or wagons. As a result, timetables and resulting cycle times have to be considered and/or assumed in order to extract the respective number of physical (and hence "equippable") trains/wagons.

After the definition of the network and both train numbers and routes, the number of track side elements has to be identified in a next step. It is assumed based on German regulations for infrastructure (DB Netze, 2013), calculated edge length and operational density. In context of the latter representative numbers for axle counters and signals are defined for different line categories. Trackside elements numbers were assumed according to Mixed-Traffic-Line categories, as this applies to large parts of the Austrian Network.

3.4. Train categories and equipment intervals

Finally, the observation interval is defined to eight years, marking the migration period. This has been set to match the maintenance interval of each the train categories (EBO, 2022). This interval can vary between the train categories. The intervals are defined as two years for high-speed trains, four years for regional trains and eight years for freight trains and wagons respectively. It is worth noticing that the length of the migration period highly correlates with the complexity of the optimization model and therefore computational calculation time. Definition of train categories and respective equipment intervals are shown in Table 3:

It is assumed that within the respective interval all trains in the system can be equipped and workshop equipment rates are constant over the migration time. This subsequently results in the definition of amounts of trains per category that can be equipped within a certain timespan. For high-speed trains, for example, this will mean that half of the amount of HS trains can be equipped in Year 1 (as a total equipment interval of two years for HS trains is assumed). This leads to a very dynamic modelling of migration capacities, suiting the respective migration problem and targeted time span.

4. Use case

To assess the methodology a case study based on a generic, mid-size railway network derived from the Austrian railway network is presented. It allows to assess a European railway network with mixed-traffic operation and sufficient complexity while remaining computationally tractable.

Publicly available infrastructure and network operation data on the basis of node-edge-relations is implemented as a first step

Table 2
Data requirements and processing.

Data category	Source/Processing
Network topology	Derived from train count data on a segment level – mapped with geo information
Operational program and train routes	Train count data, algorithmically approximated with dfrouter (SUMO, see below)
Time horizon	Defined on the level of maintenance intervals (cf. subsection train categories and equipment intervals below)
OTI equipment rate per category	Calculated based on the assumption that within the defined time horizon all trains can be equipped
Amount of trackside elements	Calculated based on literature

Table 3
Equipment interval per train category.

Train category	Abbreviation	Equipment interval (years)
High-speed	HS	2
Short-distance regional	R1	4
Short-distance commuter	R2	4
Freight (block)	F1	8
Freight (single wagons)	F2	8

(European Commission, 2022). Pre-Processing includes smaller node aggregation operations and geo mapping. Next, the edge-wise count data is transformed to train trips (methodology cf. section 3). Hereby turnaround times according to Table 4 are assumed. Moreover, trackside elements per single edge are assumed as described above. Table 4 shows details of both the network and the created operational program.

Table 4 also shows the approach followed in order to model freight train operation realistically: assuming freight trains as single "train units" similar to passenger trains would be an oversimplification, narrowing expressiveness of the analysis performed. For this reason, for freight trains an equal share of freight block trains – always operating in the same constellation of wagons – and single wagon load trains is assumed. As the latter can operate in a steadily changing constellation of wagons, perspective is down-shifted by declaring single wagons as modelling "units", comparable to passenger trains. This has turned out to be a highly ambivalent assumption, as it results in a drastic increase in complexity. It leads, however, to a much more realistic model of freight train operation, considering wagons as single units that can be equipped with OTI technology. In the context of equipment, it is further assumed that single wagons have a specific route and wagons with identical routes are preferably equipped at the same time. This "bundling" of wagons is implemented to reduce complexity and hence calculation time of the model, which has proved necessary in pre-testing. Unfortunately, this assumption can also lead to slightly weaker results than optimizing with purely independent train/wagon units. Pre-tests have also shown that, for the heuristic approaches, this assumption can be lifted. In general, the resulting numbers in Table 4 illustrate the high dependency of OTI migration on the freight sector.

4.1. Model setting

The Network is modelled in Python using freely available libraries. Optimization is conducted with the Gekko solver library (Beal et al., 2018), which allows to model the if-constraint in the MINLP problem formulated above, and applying the open source APOPT-solver (Hedengren et al., 2012). The observed calculation time on standard hardware is ~18 h. This high computation time is notable due to the non-linearity of the problem. The nonlinearity originates from the fact that trackside elements of a certain edge can only be removed once all trains running over it are equipped, which gives rise to a cascading correlation in form of an if-statement (cf. section 3, equation (3)). Results from the combined heuristic strategy (cf. Table 1) were used as starting vectors when solving the MINLP. Table 5 shows details on the size and setting of the program:

4.2. Network migration results based on optimization approach

The results of the optimization of network migration are depicted in Fig. 2 and Table 6, respectively. The figure shows the development of the network-wide sum of trackside elements over the years as well as the number of trains of specific categories which have been replaced by trains with OTI. The blue curve depicts the number of trackside elements in the network. At the beginning, no trains have yet been equipped and the number of elements corresponds to the pre-migration state with 7096 trackside elements. As OTI equipment of trains and wagons proceeds, lines can be upgraded and the number of trackside elements decreases. At the end of the migration process after 8 years, the entire network has been migrated and the number of trackside elements is 0. The total amount of equipped trains per category is marked by bars growing from left to right, with no equipped trains in year 0 until all of them are equipped in year 8. The moment of full equipment (when bars stay at the same level) is different between the train categories (cf.

Table 4
Network and operational program.

Attribute	Amount	Assumed turnaround times (days)
network		
Nodes	87	
Edges	94	
trackside elements (whole network)	7096	
Operational program		
High-speed (HS)	46	½
Short-distance regional (R1)	15	¼
Short-distance commuter (R2)	16	⅙
Freight (block, F1)	261	5
Freight (single wagons, F2)	13488	10
Considered distinct train routes	292	

Table 5
Size and settings of the MINLP model.

Attribute	value
Equations	11455
Connections	50866
Variables	11464
Solver-settings according to APM.SOLVER (2022)	
minlp_branch_method	3 = lowest objective leaf
nlp_maximum_iterations (per subproblem)	500
minlp_gap_tol	0.01

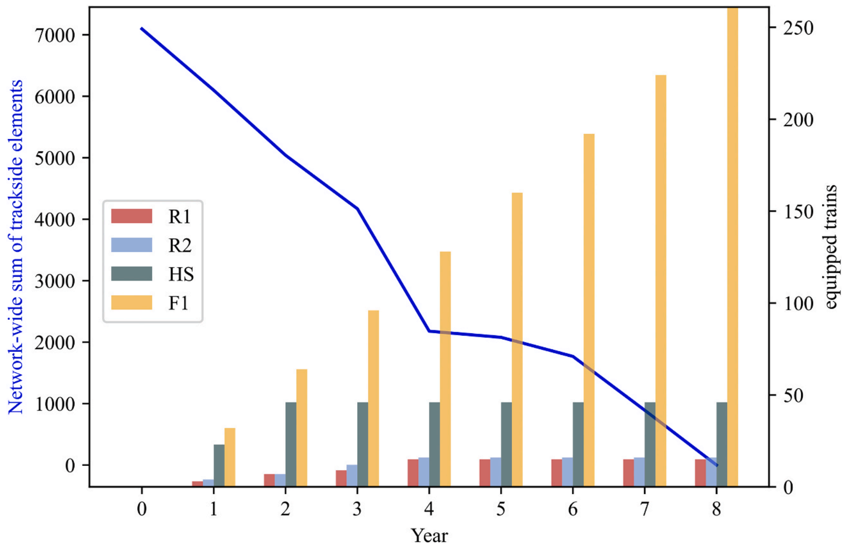


Fig. 2. Development of trackside elements.

Table 6
Number of equipped trains (HS, R1, R2, F1) and wagons (F2).

Cat.	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
HS	0	23	46	46	46	46	46	46	46
R1	0	3	7	9	15	15	15	15	15
R2	0	4	7	12	16	16	16	16	16
F1	0	32	64	96	128	160	192	224	261
F2	0	1604	3406	5108	6757	8443	10155	11767	13488

section 3). The absolute number of equipped trains and wagons over the years are shown in Table 6:

It is worth noticing that, for visualization purposes, Fig. 2 only depicts numbers of trains running in “blocks”, namely R1, R2, HS and F1. As category F2 incorporates wagons and thus has equipment rates in a different scale, it was not included in the figure. However, absolute amounts for category F2 were included in Table 6. The table also shows that all categories are equipped with very constant rates. However, slight deviations from average equipment rates can be observed, e.g. for category F2. For this category, this can be explained by the “bundling” of wagons (cf. section 4). For combinatory purposes, the optimization model is allowed to deviate the number of equipped wagons by 5% from average. Moreover, this also intends to model natural variations in wagon equipment capacities in maintenance yards. For categories R1 and R2 equipment slightly differs due to rounding issues or equipment in the last year.

In Fig. 2, gradients are comparably steep in the first four years, followed by a flattening section in the middle of the observation time to steep gradients again to the very end of the time span, slightly resembling the form of an “S-curve”. This can be explained by several influencing factors, one of them being the amount of trains equipped within the respective time. For example, the green bar marking the amount of equipped high-speed trains has reached full height in year 2, as all of these trains are equipped with OTI technology by then. This leads to a higher total amount of equipped trains in the first two years (and four years respectively), allowing higher gains to the minimization approach. This effect is enhanced by the sheer number of train categories that can be equipped in the first years, effecting in high amounts of combinatory options for the solver in this timespan. The curve flattens after most of the passenger trains are equipped (in year 4). As the share of equipped trains rises from year to year, the probability of all trains of a certain edge being equipped rises from year to year, marking higher trackside element removal rates again to the end of the observation period.

The network migration process is illustrated in Fig. 3, depicting the shape of the network and its migration state for selected years. The color-code describes the number of trackside elements for train occupation detection which are operated on a specific line (red = high amount, green = low amount). For year 0, the initial state of the network is shown where all edges still have trackside elements (thus, no dark green parts can be seen there). Over the years, trains are equipped with OTI technology, allowing trackside elements to be removed and thus gradually changing the color-code to green. Consequently, the color code of the edges of the network is all-green in year 8 (the end of the OTI-equipment period), where all trains are equipped and thus no more trackside elements remain in the network.

It can be seen that after two years a considerable amount of edges has already been removed from trackside elements. Thereby the green elements partly represent one major route of high-speed trains, as these are fully equipped after year 2 and thus optimization focusses on these lines within the first years. The figure to the bottom left shows that after year 5 (where all passenger trains have been fully equipped for a year already), trackside elements have already been removed in a geographically large part of the network. However, it takes another three years to equip the remaining freight trains, hence allowing full network migration shown to the bottom right.

4.3. Comparison to heuristic approaches

In the following, the MINLP results are compared to heuristic approaches described in Table 1. Fig. 4 illustrates the different approaches. Again, the blue curve marks the network-wide amount of trackside elements over the years for the MINLP optimization approach. The other curves represent the equipment behavior of the other approaches, where the purple, red, orange and grey lines stand for the combined, least trains, high-speed and trackside elements strategies and the green line marks random equipment. All curves have a common start, beginning at the initial state of the network in year 0, where all 7096 trackside elements are still in the network. Likewise, all five curves share the same value in year 8 - the end of the migration period - where no more trackside elements are left in the network. The labels are supplemented by the value of the objective function as defined in (1), marking the network-sum of all trackside elements over all years.

The Figure shows that the MILNP approach produces best results, having the lowest value of the objective function. Consequently, the course of the blue curve is below the other curves for most of the time. Amongst the heuristic strategies, the combined strategy shows best results. Interestingly, the combined strategy shows even better results than the MINLP approach within the first three years, meaning that by year three more trackside elements are still remaining in the network for the MINLP approach compared to the combined heuristic strategy. However, gains within the last five years are significantly higher with the MILNP method, leading to a considerably better result in total. This provides a deeper understanding of how optimization and heuristic approaches differ: While heuristic strategies calculate and equip trains on a year-by-year-basis, the optimization approach interprets the observation period as a whole, refraining from too high gains in the beginning in the favor of even higher gains towards the end.

The three other heuristic methods tend to show increasing negative gradients from left to right (with some exceptions), rather

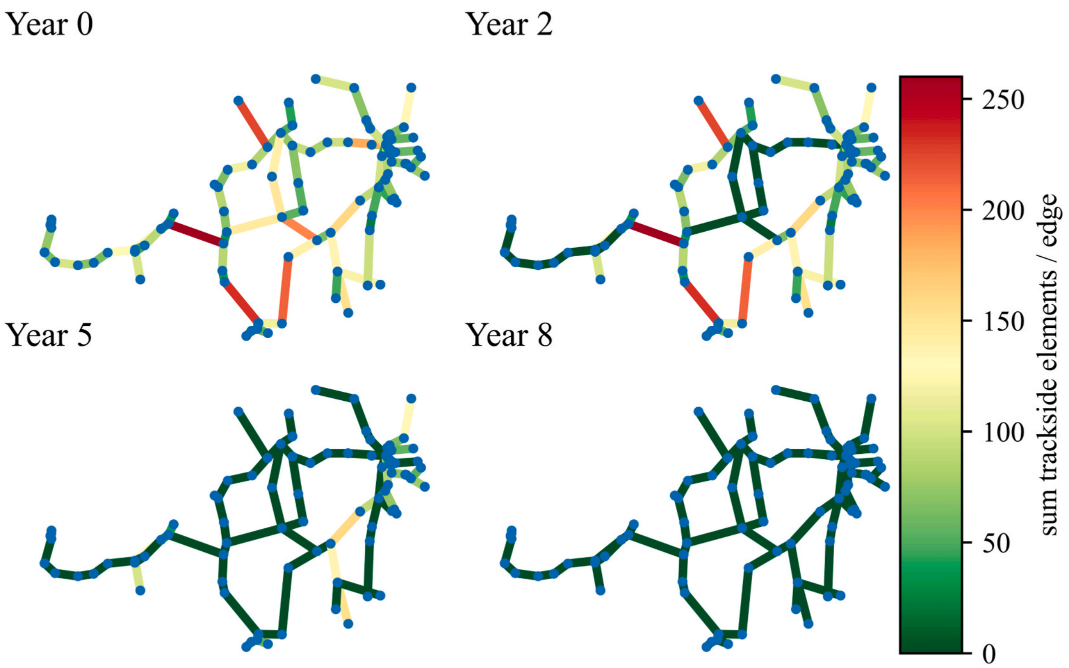


Fig. 3. Migration state on a network level for selected years.

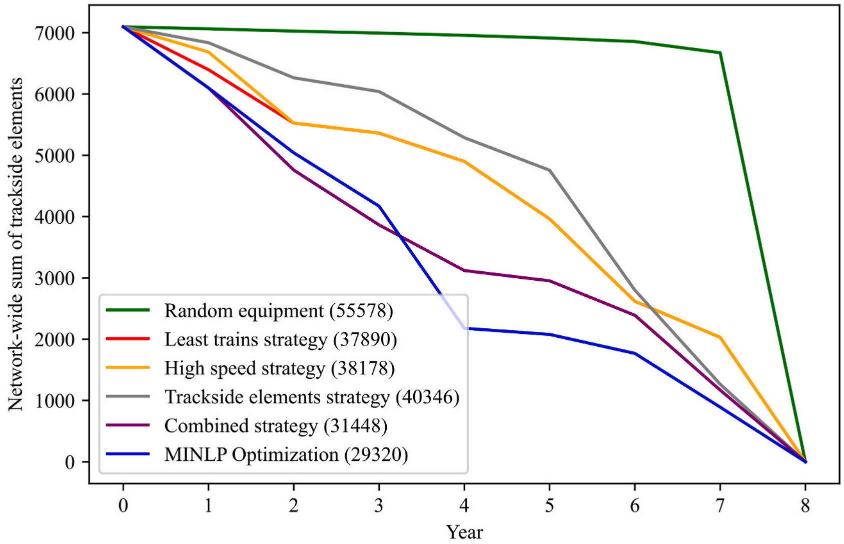


Fig. 4. Comparison of methods.

resembling a linear or flat negative exponential behavior than an "S". Moreover, it can be observed that in general the two train-focused approaches (red and orange) bear advantages compared to the trackside element strategy (grey). The latter approach has slow gains in the beginning, despite equipping trains on the edges of the network with the most track side elements. In contrast, the least train and highspeed strategies first equip trains running over edges with the least amount of total trains. Still, from a heuristic perspective, it seems advisable to simultaneously focus on both strategies, as the considerably better results for the combined approach shows. Within train-focused strategies, the focus on high-speed trains seems to bring slight disadvantages compared to the basic approach (see red and orange curve). Apparently, due to high-speed trains being only equipped in the first two years, the difference does only appear in the very beginning of the equipment period and afterwards merges with the course of the other heuristic, having identical values until the end. The high-speed approach was included to present a special focus on high-speed passenger corridors, as high-speed trains often have operational priority. The results show, however, that a pure focus on these trains and their respective corridors leads to a slightly higher objective value – an effect that has to be pondered against other decision variables not included in this study. Finally, the random approach has – as expected – only very few gains in the beginning and does not become effective until the very last year.

To evaluate and understand the working of the different approaches, the number of trackside elements per network edge is analyzed in Fig. 5. Here, the initial amount of trackside elements of each edge is scatter-plotted over the year where all trackside elements were removed from the respective edge (as all trains running over it are equipped). This correlation is shown for four selected strategies, where blue scatter points stand for the MINLP optimization strategy, orange points indicate the combined strategy, green marks the heuristic least trains strategy and red the trackside elements strategy.

It can be seen that the trackside elements strategy (red scatter points) equips trains as expected, starting with the edge with the most track side elements to the top left corner and step by step equipping the edge with next higher amounts, having no or only little "side effects" in the first three years, meaning that only few other edges are fully removed from track side elements. In contrast, all three other approaches equip trains on edges with medium or low numbers of trackside elements in the beginning. For the three edges with the highest amount of trackside elements, it can be seen that the MINLP approach removes their trackside elements chronologically between the other three heuristic approaches. In year 1 and 2, the MINLP approach and the combined strategy show quite similar behavior (blue and orange points), marking high gains as visualized in Fig. 4 above. In general, the figure shows that the optimal approach can ponder the influence of heuristic strategies against each other, finding many optimal combinations.

Summarized, the methodology introduced in section 3 was applied to a mid-size network derived from the Austrian main-line network as a use case. Results show that the MINLP optimization approach enables the removal of a high amount of trackside elements in the first four years of migration, similar to the combined heuristic strategy. Even though it comes with lower gains in the middle of the migration phase (years four to six), it shows considerably higher gains towards the end of the observation period, effecting in advantages in total compared to heuristic migration strategies. Amongst those, simultaneously focusing on the amount of unequipped trains and the amount of trackside elements per edge goes along with best results (combined strategy).

5. Conclusions and further research

In the present work, a MINLP model for describing the network migration process for OTI technology migration is proposed. Optimizing on the level of single trains, wagons and their respective distinct routes, it was shown that the model is capable of achieving the goal of early migration of significant areas within the network. The model is applied and evaluated by comparison to heuristic

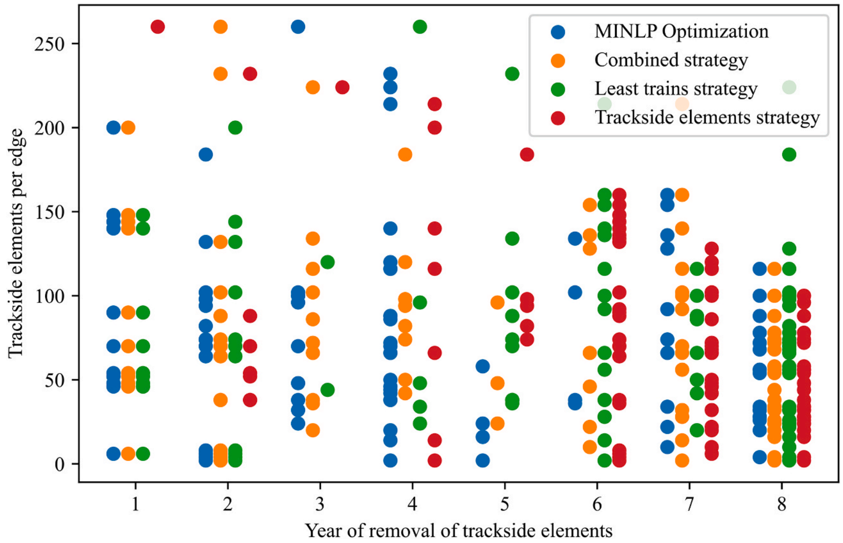


Fig. 5. Trackside elements removal moment (year).

migration strategies, showing significant advantages and a considerably different train equipment strategy. In the context of the paper the objective was to accelerate network migration to save operating costs by the removal of trackside elements and increase reliability. Besides application to wider scopes in ETCS technology migration, more complex objectives and considerations of migration scenarios can be considered. For instance, different migration periods can be compared to initialize in-depth economic analyses based on the methodology presented in this study.

On a network level, nodes and junctions can be added to the proposed approach, where the removal of trackside elements might be subject to very different constraints: while these critical parts of the network are of special importance in terms of both capacity and safety, upkeeping trackside vacancy detection might at least partially be inevitable to ensure safe switch movements in interlocking areas.

Also, the present work has not yet implemented stakeholder-specific perspectives. Moreover, capacity evaluation can be implemented into the approach, considering capacity rises due to moving block operation as an alternative objective. Finally, operational resilience in the event of system failures has not been discussed. Here, existing track-bound systems can play an important role in upholding and strengthening the operational resilience of railway networks, an important aspect, which should be addressed in future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has received funding from Shift2Rail Joint Undertaking (JU) under grant agreement No 881806. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Shift2Rail JU members other than the Union.

References

Alvarez Lopez, P., et al., 2018. Microscopic traffic simulation using SUMO. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. The 21st IEEE International Conference on Intelligent Transportation Systems 2575–2582. Maui, USA.

APM.SOLVER, 2022. APMonitor Option: APOPT Options. [Online] https://apmonitor.com/wiki/index.php/Main/OptionApmSolver. (Accessed 22 November 2022).

Beal, L., Hill, D., Martin, R., Hedengren, J.D., 2018. GEKKO optimization suite. Processes 6 (8). https://doi.org/10.3390/pr6080106.

DB Netze, 2013. Richtlinie 413.0301 "Infrastruktur Gestalten.

EBO, 2022. Eisenbahn-Bau- und Betriebsordnung: § 32 Abnahme und Untersuchung der Fahrzeuge [Online] Available at: https://www.gesetze-im-internet.de/ebo/__32.html. (Accessed 29 November 2022).

EEIG ERTMS Users Group, 2022. Principles: ERTMS/ETCS Hybrid Train Detection. [Online] Available at: https://ertms.be/wp-content/uploads/2023/06/16E0421F_HTD.pdf (Accessed 28 July 2023).

European Commission, 2022. eurostat. [Online] Available at: https://ec.europa.eu/eurostat/databrowser/view/rail_tf_ns10_at/default/table?. (Accessed 22 November 2022).

European Railway Agency, 2016. Study on Migration of Railway Radio Communication System from GSM-R to Other Solutions. [Online] Available at: https://www.era.europa.eu/sites/default/files/library/docs/studies/systra_study_on_migration_of_railway_radio_communication_from_gsm-_to_other_solutions_en.pdf. (Accessed 18 November 2022).

Europe's Rail Joint Undertaking, 2021. Master Plan (DRAFT). [Online] Available at: https://rail-research.europa.eu/wp-content/uploads/2022/03/EURAIL_Master-Plan.pdf (Accessed 22 November 2022).
Hedengren, J., Mojica, J., Cole, W., Edgar, T., 2012. APOPT: MINLP Solver for Differential Algebraic Systems with Benchmark Testing. INFORMS Annual Meeting, Phoenix, AZ.
Kawano, T., Matsumoto, M., Hirao, Y., 2016. Application and Effectiveness of Assurance Technology for Autonomous Decentralized Railway Signalling Systems. WIT Transactions on The Built Environment, pp. 167–178. https://doi.org/10.2495/CR160161.
Kochan, A., Gruba, Ł., 2018. Analysis of the migration process in the ERTMS system from GSM technology to LTE on the polish railway, 897. In: Mikulski, J. (Ed.), Management Perspective for Transport Telematics. TST 2018, Communications in Computer and Information Science. Springer, Cham. https://doi.org/10.1007/978-3-319-97955-7_17.
Lackhove, C., 2014. *Optimierung der Migration von ETCS. Dissertation,* s.l. Institut für Verkehrssystemtechnik.
Lackhove, C., Jaeger, B., Lemmer, K., 2010. Generating and optimizing strategies for the migration of the European train control system. WIT Trans. Built Environ. 114, 383–394. https://doi.org/10.2495/CR100361.
Obrenovic, M., Jäger, B., Lemmer, K., 2006. Migration of the European Train Control System (ETCS) and the Impacts on the International Transport Market. ETC Proceedings. European Transport Conference 2006.
Ranjbar, V., Olsson, N.O., Sipilä, H., 2022. Impact of signalling system on capacity – comparing legacy ATC, ETCS level 2 and ETCS hybrid level 3 systems. Journal of Rail Transport Planning & Management 23 (9), 100322.
Vergroesen, R., 2020. *ERTMS/ETCS Hybrid Level 3 and ATO: A Simulation Based Capacity Impact Study for the Dutch Railway Network.* s.l.:PhD Thesis. TU Delft.
Wen, T., et al., 2021. A cost-effective wireless network migration planning method supporting high-security enabled railway data communication systems. J. Franklin Inst. 358 (1), 131–150. https://doi.org/10.1016/j.jfranklin.2019.01.037.
X2Rail-2 D.4.2, 2020. D.4.2: Functional Architecture & Interfaces Specifications & Candidate Technologies Selection.
X2Rail-2 D.4.1, 2020. D4.1: Train Integrity Concept and Functional Requirements Specifications.