DITTO Project Deliverable 4.2 Milestone 10

Draft Good Practice Guide

Edited by John Preston (University of Southampton) with contributions from:

Phillip James, Faron Moller, Markus Roggenbach (Swansea University);

Hoang Nga Nguyen, Coventry University;

John Armstrong, Tolga Bektas, Attila A. Kovacs, Chris N. Potts (University of Southampton);

Ronghui Liu and Hongbo Ye (University of Leeds).

Version 2, June 2017

Abstract

This report summarises the findings of the DITTO project to date with particular respect to optimising the rail life cycle. This consists of three main stages in terms of optimising the overall system, optimising the plan (or timetable) and optimising (real-time) operations. This is underpinned by continuous performance monitoring with a particular emphasis on the relationship between capacity utilisation and service reliability. A series of good practices are identified with respect to using safety and capacity analysis to determine theoretical capacity limits, using optimisation techniques to identify practical capacity limits and using simulation techniques that in the future will allow optimised timetables to be put into practice.

Table of Contents

List of Figures

List of Tables

1. Introduction

In the last 20 years, rail traffic on the national network in Britain has grown by around 100% in terms of passengers and freight and by 50% in terms of train movements, whilst the overall quantum of infrastructure has barely changed (ORR, 2016). To meet the challenges that such growth presents, the UK rail sector has established the Future Traffic Regulation Optimisation (FuTRO) research programme which is examining the ways that advances in technology, including those associated with the digital railway, can improve rail operations. FuTRO is thus developing the control, command and communications theme of the Rail Technical Strategy (RSSB, 2012). One of the projects that has been commissioned by FuTRO is Developing Integrated Tools to Optimise (DITTO) Railway Systems, funded by RSSB (formerly the Rail Safety and Standards Board) for three years from September 2014 – see www.dittorailway.uk. DITTO is a consortium of researchers based at universities in Leeds, Southampton and Swansea. Industrial support has been provided by Arup, Siemens Rail Automation and Tracsis. It builds upon separate projects undertaken by the three Universities for the RSSB/EPSRC Capacity at Nodes programme that ran from 2010 to 2012. The three projects were Challenging Established Rules for Train Control (Leeds), Overcoming Capacity Constraints: A Simulation Integrated with Optimisation of Nodes (OCCASION – Southampton) and SafeCap (Swansea) (see Goodall et al., 2013).

DITTO contributes to FuTRO by establishing basic principles and proofs of concept and by developing optimisation formulations, algorithms and processes that will help deliver a step change in rail system performance and help to meet future customer needs. This will be done by taking into account developments in human and automatic control on trains and in control centres (particularly related to ERTMS) and by making better use of data, particularly with respect to time and position of trains.

DITTO's objectives are thus to:

1. Develop optimisation activities that maintain safe operating conditions and do not exceed theoretical capacity limits.
2. Develop timetables that optimise capacity utilisation without compromising service reliability.
3. Combine dynamic data on the status of individual trains to produce an optimal system-wide outcome in terms of traffic management.
4. Use Artificial Intelligence to produce tractable solutions to real-time traffic control.

Objective 1 relates to network optimisation. It determines the theoretical capacity of a given infrastructure scheme plan that is operated in a safe manner. By inference, it can be used to optimise infrastructure provision. Our findings with respect to objective 1 are discussed in section 2. Objective 2 relates to plan optimisation. It involves matching trains to the infrastructure so as to maximise the throughput of trains subject to acceptable levels of performance, primarily in terms of punctuality. Our findings with respect to objective 2 are outlined in sections 3 (where we deploy analytical methods) and 4 (where we deploy optimisation techniques). Objective 3 relates to traffic management optimisation. It involves dynamically controlling trains to minimise the impact of service disruptions. Our findings with respect to objective 3 are outlined in section 5, based on the simulation tools we have developed. Objective 4 attempts to integrate the three optimisation processes described above by using machine learning tools based on performance monitoring. Our initial findings are discussed in section 6. The overall approach adopted in optimising the railway life-cycle is illustrated by Figure 1.

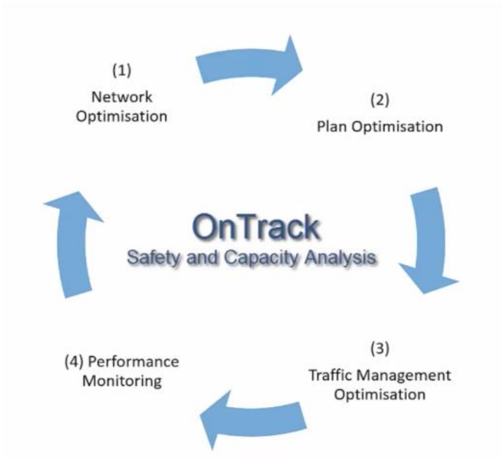


Figure 1: Optimising the Rail Life-Cycle.

The DITTO project thus consists of four inter-related and complementary technical strands that are innovative both on their own and in combination with each other.

Safety – this strand allows optimisation activities to proceed in the knowledge that safe operating conditions are being maintained and that theoretical capacity limits are not being exceeded.

Reliability – this strand quantifies the trade-offs between the provision of additional train services and the maintenance of service quality so as to develop timetables that optimise capacity utilisation without compromising service reliability.

Dynamic simulation – micro-level data on the status of individual trains will be combined to produce an optimal, macro-level outcome, transmitting the system-wide needs back to the micro-level, so that individual train movements can be optimised within overall system requirements.

Network integration – using artificial intelligence, optimised timetables are produced that can be adjusted in real time through dynamic simulation. Our work in this area has not yet completed but we discuss some of our intentions in section 6.

1.1 System Optimisation

Figure 2 shows that our starting point for system optimisation is to put a Scheme Plan (SP) through a safety and capacity verification process (see also James et al., 2015a). This might use either RailML or output from Computer Assisted Design (CAD) and the OnTrack editor (see James et al., 2015b). These approaches are brought together by the OnTrack Domain Specific Language (DSL) developed by Swansea University. Safety verification can then be performed using a variety of languages such as: CSP (Communicating Sequential Processes), a specification language for concurrent systems defined by Sir Tony Hoare in the early 1980s; CSP Parallel B, a combination of CSP and the specification language B, defined by Swansea's research partners at Surrey University around 2000; and CASL (Common Algebraic Specification Language). As section 2, illustrates the key output is the maximum number of trains (and their sequence) that constitutes the safety limit for a given infrastructure. This can than set the theoretical capacity limit for the plan optimisation.

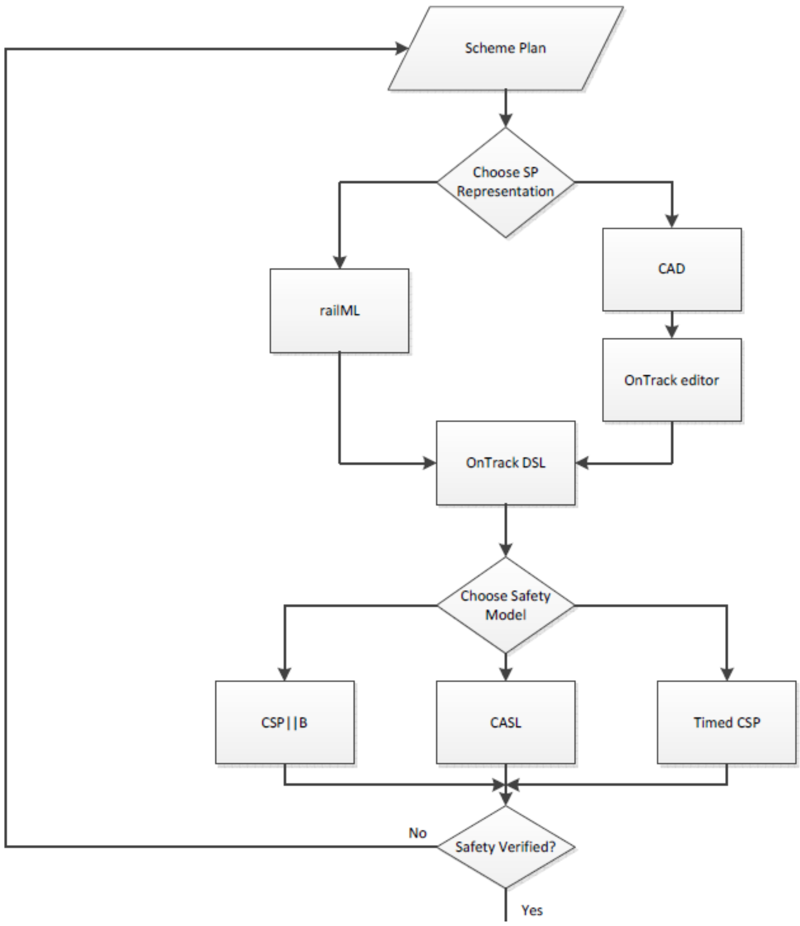


Figure 2: System Optimisation - Safety and Capacity Validation.

1.2 Plan Optimisation

Figure 3 shows that the next stage is to undertake the plan optimisation. This takes the existing Timetable (TT) in CIF (Common Interface Format) and the safety limits established by the verification and, using Capacity Utilisation Indices or other related approaches, assesses the likely performance in terms of Congestion Related Reactionary Delay (CRRD). Performance scenarios are then developed to feed into a stochastic optimisation based on a variant of job shop scheduling, in which the railway infrastructure (track and signalling) is treated as a machine shop and train movements are treated as jobs to be processed. This involves a two stage stochastic program. In the first stage, new trains are inserted into the timetable. The second stage involves optimising for reliability for various random scenarios. This is undertaken at the meso-level, for example for a node such as Peterborough on the East Coast Main Line (ECML). The implications are assessed at a macro-level, for example for the ECML between Doncaster and Alexandra Palace. Initially, we had intended to use a variation of the Multi-Commodity Network Design Problem (MCNDP) but this did not prove to be practical. Instead, a deterministic job shop optimisation is applied. Constraints ensure that the revised timetable is within safety limits. Once the optimisation is confirmed at the meso- and macro-levels, it is fed into the final stage.

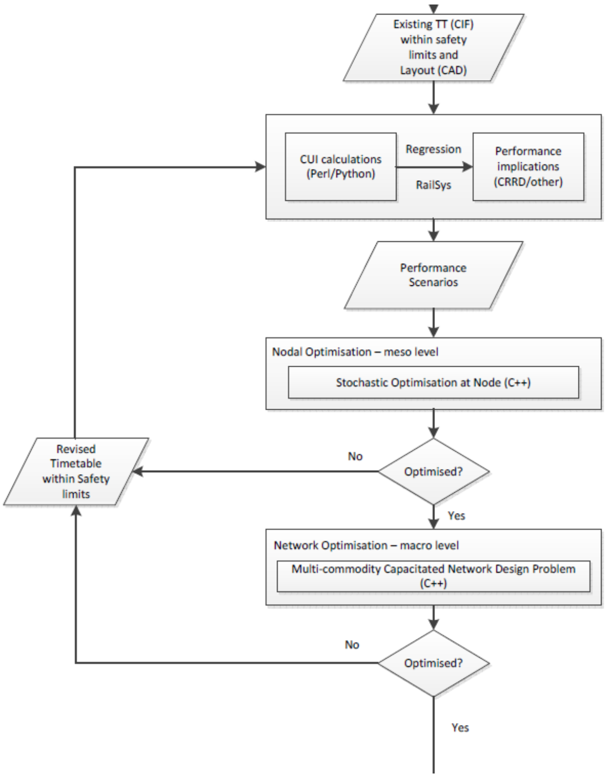


Figure 3: Plan Optimisation

1.3 Operations Optimisation

The third stage involves operations optimisation by examining the scope for dynamic rescheduling and this is done by using the TrackULA train simulator, together with consideration of traditional algorithms, and alongside human control and artificial intelligence based on machine learning. This is informed by historic data on performance (in terms of delays) that has also informed the static optimisation and may be used to consider a wider range of scenarios. The final output, as illustrated by Figure 4 is an optimised timetable, along with a series of rescheduling plans, if needed.

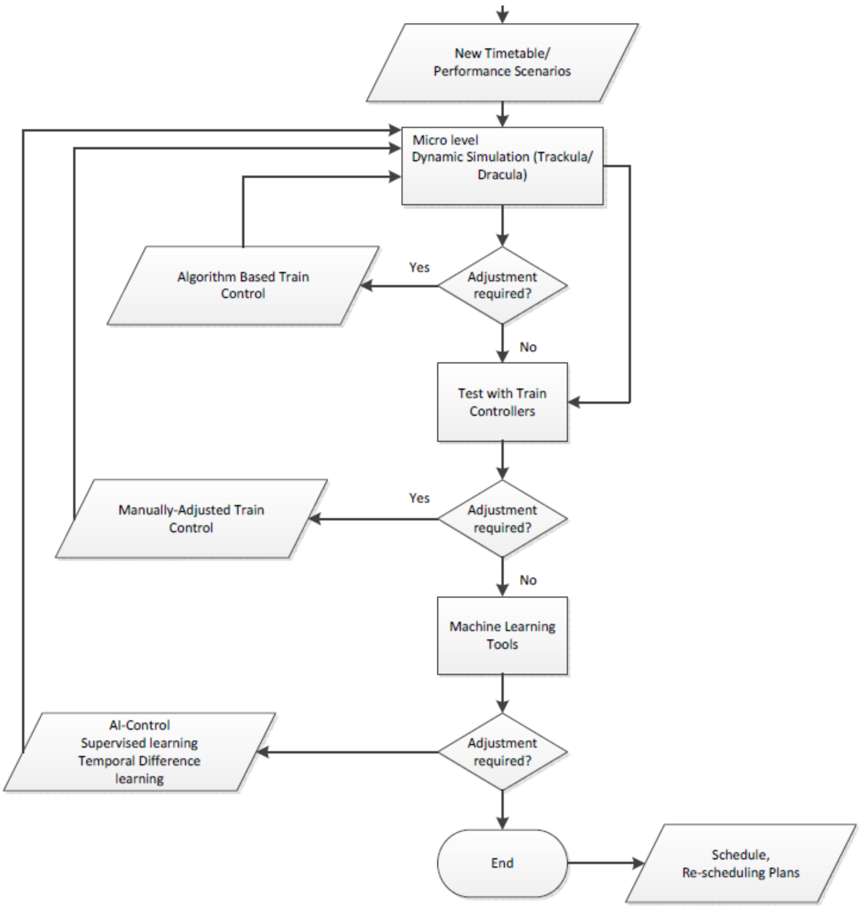


Figure 4: Operations Optimisation by Dynamic Rescheduling

1.4 Monitoring and Evaluation

As can be seen from Figure 1, the optimisation life-cycle will be informed by monitoring and evaluation. This will be used to continuously improve the system, as indicated by the feedback loops in Figures 2 to 4. Key performance indicators will include capacity utilisation indices (see section 3) and measures of punctuality and reliability, such as

CASL (cancellations and serious lateness), CRRD (congestion-related reactionary delay) and PPM (public performance measure).

1.5 Integrated Assessments

Our work draws on the rich literature in this application domain, with a particular emphasis on rail capacity (for reviews, see Abril et al., 2008 and Kontaxi and Ricci, 2012). These reviews have highlighted a number of approaches to rail capacity management, including analytical methods (non-parametric and parametric), simulation, optimisation and integrated assessment. DITTO is attempting to provide an integrated assessment by combining analytical, optimisation and simulation approaches with formal methods for safety and capacity verification.

2. Safety Verification and Capacity Assessment

FuTRO has the objective of improved management and control via system optimisation, where measures are constrained by theoretical capacity, and safety is an underlying indispensable precondition. Through our previous SafeCap project, the Swansea team has developed expertise in this topic as well as scientific results on safety and capacity, which were turned into applicable tool sets that can be scaled up to complex rail nodes. Within DITTO, we have continued and expanded on this work:

- We further developed the OnTrack Tool and integrated it with the Birmingham Railway Simulation Suite (BRaSS) – former called BRaVE (Birmingham Railway Virtual Environment) -- developed by the Birmingham Centre for Railway Research and Education, therefore reaching out to the DEDOTS project, which is also funded within the FuTRO programme – see Section 2.1 and Good Practice X.

- We devised a new method for formal safety analysis of computer based interlocking at the design level based on the process algebra CSP. This method has been implemented in the OnTrack tool and tested on real world examples with verification times now in seconds or minutes. We believe this approach to be mature enough to be used in industrial practice -- see Section 2.1.

- We developed a new method for formal safety analysis of ERTMS Level 2 systems at the design level based on the algebraic specification language Real-Time Maude. From an industrial perspective, Siemens Rail Automation, Chippenham, considers our work to have high potential to improve quality assurance within their software development process of ERTMS level 2 inter-lockings and RBCs – see Section 2.3.

- There is ongoing, promising work to analyse track plans for capacity using the process algebra Timed CSP. Here, the models are timed extensions of the CSP models that we use for safety analysis. This is because we see safety and capacity as two sides of the same coin: in the interest of safety, trains must be separated by headways; in the interest of capacity, trains should run closely together. The current status is that our models allow one to demonstrate predictable effects on models for capacity, addressing calibration and scalability is ongoing work – see Section 2.3.

2.1 The OnTrack Tool

OnTrack (James et al., 2013, 2016)[1] is an open toolset for railway verification developed between Swansea University, Coventry University and Surrey University. Within the DITTO project, OnTrack has been developed further for railway optimisation and serves as a common platform for tool integration.

Within the railway industry, defining graphical descriptions is the de facto method of designing railway networks. These graphical descriptions enable an engineer to visually represent the tracks and signals etc., within a railway network. The OnTrack toolset (see Figure 5) achieves the goal of encapsulating formal methods for the railway domain. Overall, the OnTrack toolset provides a modelling and verification environment that allows graphical scheme plan descriptions to be captured and supported by formal verification. Thus, it provides a bridge between railway domain notations and formal specification. This in turn makes formal methods accessible to domain engineers.

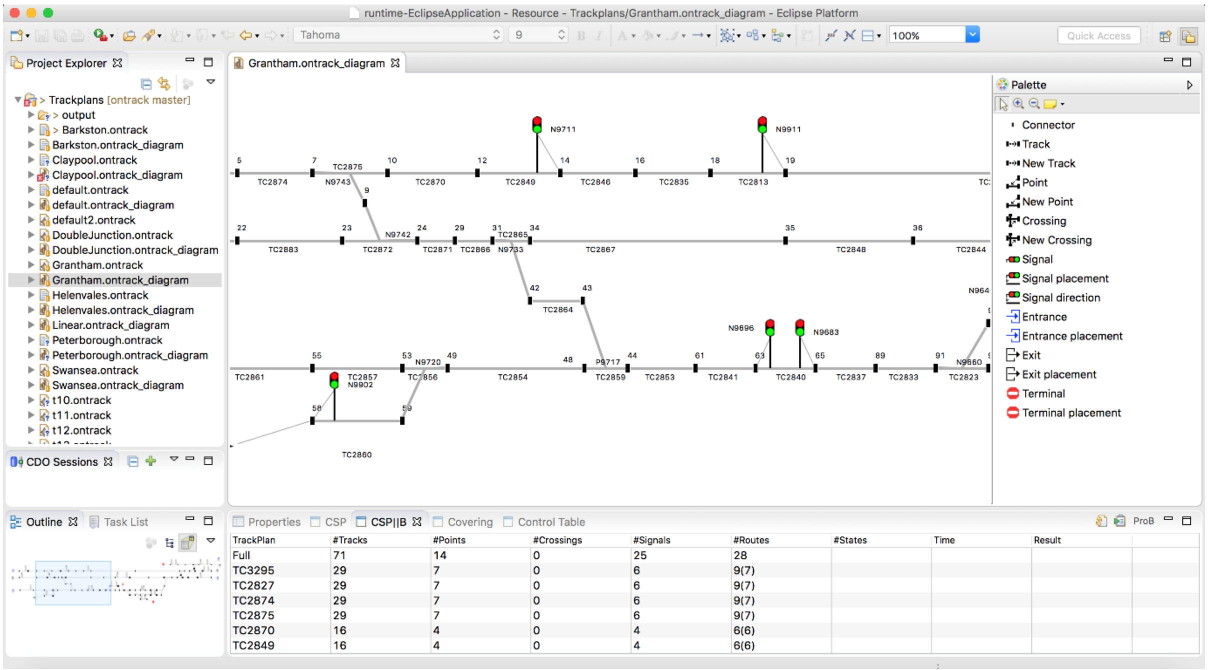


Figure 5: The OnTrack Editor

In OnTrack, we emphasise the use of a Domain Specific Language (DSL) and the decoupling of this DSL from the verification method. One of the novelties of this is that we can define abstractions on the DSL in order to yield an optimised description prior to formal analysis. Importantly, these abstractions allow benefits for verification in different formal languages. Also, due to the way OnTrack has been designed, it is easily extendable to allow the generation of formal models in any given modelling language.

[1] See also the OnTrack toolset – Webpage: http://www.cs.swansea.ac.uk/~csmarkus/ProcessesAndData/ontrack

This means that the graphical editor of OnTrack can be used as a basis for generating different formal specifications in different languages. Finally, OnTrack is designed for the railway domain, but the clear separation of an editor with support for abstractions from the chosen formal language is a principle that is more widely applicable. For full details on OnTrack, see James et al., 2015b.

2.2 Formal Safety Verification

Formal verification of railway control software has been identified as one of the "Grand Challenges" of Computer Science. But in respect of this challenge, a question has been asked by the community: "Where do the axioms come from?" Bluntly expressing a view common to the Formal Methods community, Paulson states, "I have seen many pieces of work spoilt by unrealistic models, incorrect axioms or proofs of irrelevant properties" (Paulson, 2012). The modelling of systems, as well as of proof obligations, needs to be *faithful*.

In this section we report on two faithful and formal modelling and verification approaches on the design level: safety analysis of interlocking designs in CSP and ERTMS Safety Verification in Real Time Maude. As interlockings are also part of the ERTMS Level 2 systems, our CSP analysis applies to both traditional railway systems and ERTMS level 2 systems.

Good Practice I: Safety Analysis in CSP

In order to develop a faithful model, we first developed an abstract view of a classical railway system. To this end, we produced a hierarchy of components and their communications in the form of an information flow diagram to visualise the communication between railway elements, as shown by Figure 6.

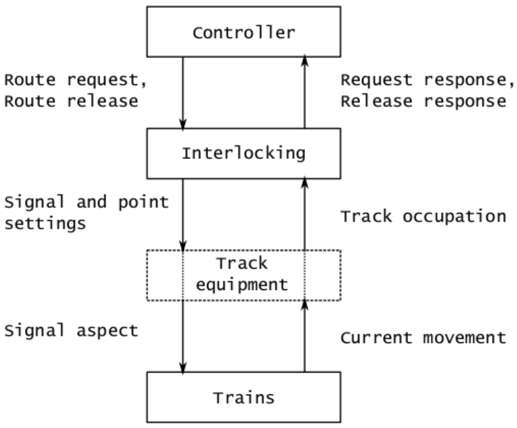


Figure 6: Information Flow Diagram

This abstract view was then modelled in a so-called *process algebra*, a framework for describing processes (agents, systems) and their interactions with each other and their environment. As the name implies, a process algebra provides algebraic laws which allow for formal analyses of the behaviour of the processes being modelled. We provide here only the briefest glimpse into process algebra and how it is used.

Process algebra research began with Robin Milner's seminal work starting in 1973 on the *Calculus of Communicating Systems* (CCS) (Milner, 1980) – though this was itself influenced by *Petri nets* (Petri, 1962) and the *actor model* (Hewitt et al., 1973). Tony Hoare's *Communicating Sequential Processes* (CSP) first appeared in 1978 and was subsequently developed into a fully-fledged process algebra with the publication of his CSP textbook in 1985 (Hoare, 1985). There are various other modelling languages in the category of process algebra, but we have adopted CSP for this work.

A process algebra has two main constituents: *processes* – these are the entities with a "behaviour", in our case, for example: the Controller, the Interlocking, the signals, and the trains; and *events* – these are the things that we can "observe" from the processes, in our case, for example: that the Controller makes a Route request, or a train moves from one track to the other. It then provides a number of *algebraic operations* for defining and combining processes; typical amongst these are: *sequential composition* (running two processes one after the other); *concurrent composition* (running two or more processes together in parallel with their events happening in an interleaving fashion, with the synchronous execution of events modelling a communication between processes); and *choice* (running just one of a given collection of processes, with the choice determined by the system or being made by the environment non-deterministically). Processes are then defined by algebraic equations, and the execution of a system (the parallel composition of the processes) is represented by a *labelled transition system* (LTS), which is a set of states with transitions (arrows) between them labelled by events. If in our LTS we have S — a —> T where S and T are states with a transition labelled a going from S to T, this indicates that if the system is in state S, it can do an a event, and by doing so it will evolve into state T.

In order to model railway systems in CSP, we have first systematically described their dynamics in a number of tables. As an example, Tables 1 and 2 describe how trains 'interact' with a green, respectively red, signal.

Table 1: Move and Cancel Route Behaviours (for a Green signal)

Event	Explanation	Condition
move.x.y	A train moves from track *x* to track *y* past the signal, which is changed to red.	The train is on track *x* which contains the signal, and track *y* is a next track.
cancelRoute.r	The route *r* is cancelled and the signal is changed to red.	None.

Table 2: Hang Move and Set Route Behaviours (for a Red signal)

Event	Explanation	Condition
hangMove	A train passes the signal whilst it is red.	The train is on the track which contains the signal.
setRoute.r	A route *r* is set and the signal is changed to green.	The route must begin at this signal.

We can get an intuitive impression of the high-level behaviour we are trying to capture as an LTS with the above tables from Figure 7:

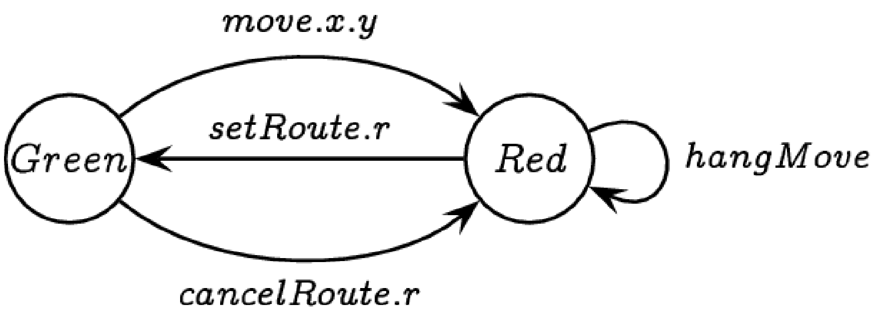


Figure 7: Rail Application of a Labelled Transition System

To give a flavour of CSP, the representation of this behaviour is shown in Figure 8.

Signals process in CSP

$SignalBehaveGreen(Id) =$
$(\Box\, np : next(signalAt(Id)) \bullet move.signalAt(Id).np \rightarrow SignalBehaveRed(Id))$
$\Box$
$(\Box\, r : routes(Id) \bullet cancelRoute.r \rightarrow SignalBehaveRed(Id))$

$SignalBehaveRed(Id) =$
$(\Box\, r : routes(Id) \bullet setRoute.r \rightarrow SignalBehaveGreen(Id))$
$\Box$
$(\Box\, np : next(signalAt(Id)) \bullet hangMove \rightarrow SignalBehaveRed(Id))$

Figure 8: Behaviour Representation in CSP.

The 2017 MRes dissertation by Michael Smith details this approach (Smith, 2017). Thanks to such systematic modelling utilising the strength of CSP and the fast model checker FDR3 we have built up a fast verification method, which is now automatically implemented in the OnTrack tool. Table 3 summarises some verification times from real world rail nodes on the East Coast Main Line. Note that, thanks to further tool development of OnTrack, the times have dramatically improved during the course of the DITTO project, and also that now all the verifications are fully automatic.

Table 3: Verification Times on the East Coast Main Line

Rail Node	Verification Time (for the whole plan)
Allington	0m23.199s
Barkston	0m18.371s
Werrington	0m14.546s
Grantham	45m27.161s

Good Practice II: ERTMS Safety Verification in Real Time Maude

ERTMS extends classical signalling systems by adding a radio block centre and adding control computers to trains. This allows, in ERTMS/ETCS Level 2, speed and braking curves of each individual train to be taken into account. These determine, for each train individually, the train's braking point well in advance of the end of the movement authority that the ERTMS signalling system had granted to the train. This will separate trains by shorter margins (compared to classical signalling systems) and thus increase capacity. Concerning formal safety analyses, for ERTMS it is necessary to develop and analyse timed or hybrid models. Note that – as ERTMS level 2 still includes interlockings – the challenges for formal safety analysis for classical interlocking designs remain, and are extended by new dimensions.

More specifically, an ERTMS/ETCS system consists of a controller, an interlocking (a specialised computer that determines if a request from the controller is "safe"), a radio block centre, track equipment, and a number of trains. Whilst the ERTMS/ETCS standard details the interactions between the trains and track equipment (e.g., in order to obtain concise train position information) and the radio block centre and trains (e.g., to hand out movement authorities), the details of how the controller, interlocking and radio block centre interact with each other are left to the suppliers of signalling solutions, such as our industrial partner Siemens Rail Automation UK. In this example, we work with the implementation as realised by Siemens and in the following we refer to this system simply as ERTMS.

One development step when building an ERTMS system consists of developing a so-called detailed design. Given geographical data such as a specific track layout and what routes through this track layout shall be used, the detailed design adds a number of tables that determine the location-specific behaviour of the interlocking and radio block centre. To the best of our knowledge, our modelling of ERTMS is the first one comprising all ERTMS subsystems required for the control cycle in ERTMS Level 2.

The objective of our modelling is to provide a formal argument that a given detailed design is safe. Here we focus on collision freedom, though our model is extensible for dealing with further safety properties such as derailment and run-throughs, and potentially with performance analysis.

We base our modelling approach on Real-Time Maude, a language and tool supporting the formal object-oriented specification and analysis of real-time and hybrid systems. In order to obtain a faithful model of ERTMS/ETCS level 2 on the design level, we follow a methodical approach, established by the Swansea Railway Verification Group.

As a first modelling step, we systematically identify the entities of ERTMS; describe their abstract behaviour; and determine the abstract information flow between them, all in line with the design by Siemens Rail UK, see Figure 9.

.

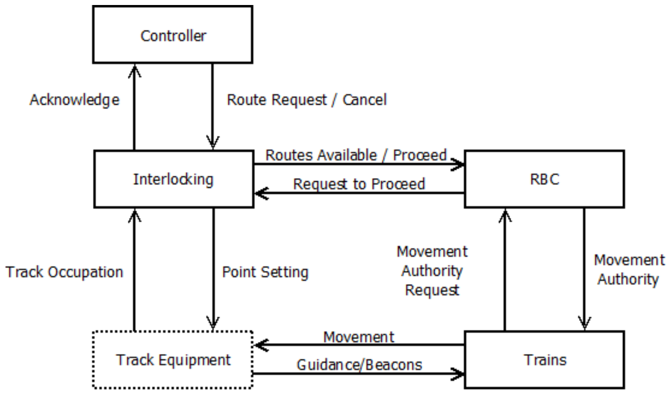


Figure 9: ERTMS Architecture

Tables 4 and 5 show a series of verification results that have been achieved via modelling. They highlight the number of rewrite (or verification) steps needed for three rail-yards against two different control strategies: a round-robin controller, which follows a given timetable for route requests, and a random controller that can choose to make any route requests at any time. For further details see James et al. 2015c.

Table 4: Verification Results of Model Checking with Restricted Control Strategy

Scheme Plan	**Round Robin Controller Unbounded**	
	No Crash Tracks	No Crash Distance
Pass-through	0.22s / 429,601 rewrites	0.25s / 585,862 rewrites
Cross	0.22s / 403,997 rewrites	0.25s / 514,958 rewrites
Twist	0.22s / 639,841 rewrites	0.48s / 972,169 rewrites

Table 5: Verification Results of Model Checking with Random Control Strategy

Scheme Plan	Random Controller in Time 300	
	No Crash Tracks	No Crash Distance
Pass- through	181.22s / 190,680,755 rewrites	212.26s/ 297,058,224 rewrites
Cross	891.50s / 503,331,780 rewrites	841.28s/ 723,639,655 rewrites
Twist	1,222.79s / 652,668,124 rewrites	1,340.09s/1,104,718,343 rewrites

The results show that unbounded model checking is successful when control is restricted, e.g., to our round-robin controller. This is due to the restrictions that such a timetable puts on train movements through the scheme plan. However, when using our random controller, the state space increases. Moreover, there are infinite traces possible, e.g., by the controller choosing the same route over and over again. Thus, we provide results for up to a given time bound of 300 seconds. Note that this time is enough to ensure that both trains can travel completely through each of the scheme plans. Another phenomenon is that model checking for the logical safety condition “No Crash Track” requires fewer rewrites (approximately 20%) than for the physical safety condition “No Crash Distance”. This follows one's intuition.

As expected, model checking times increase with the complexity of the scheme plans. One naive complexity measure would be the number of routes available in a scheme plan. We note that there are five routes in the Pass-through station; six routes in the Cross; and eight routes in the Twist. This again follows intuition, as the random controller has more freedom in more complex track plans. Note that this observation does not necessarily carry over to the round robin controller: here, the order in which the routes are requested plays a role as well and can possibly overshadow this effect. Finally, it is future work to consider more varied rail-yards, and also how the frequency of controller requests affects model checking results.

2.3 Capacity Assessment

Overcoming the constraints on railway capacity caused by nodes (stations and junctions) on the rail network is one of the most pressing challenges to the rail industry. In 2007, the UK governmental White Paper “Delivering a Sustainable Railway” stated: “Rail’s biggest contribution to tackling global warming comes from increasing its capacity” (DfT, 2007, page 10). High capacity, however, is but one design aim within the railway domain. Railways are safety-critical systems. Their malfunction could lead to death or serious injury to people, loss or severe damage to equipment, or to environmental harm. To this end, we aim to develop an integrated view of rail networks, within which capacity can be investigated without compromising safety.

Good Practice III: Capacity Analysis in Timed CSP

The process algebra CSP has successfully been applied to modelling, analysing and verifying railways for safety aspects, see Section 2.1 above. Solely concerned with safety, this approach has ignored the aspect of time. Yet the capacity of a rail network node is highly dependent on time: moving a point or moving a train through a node takes time, sighting and braking distance are functions of time. Thus, rather than using CSP, we apply Timed CSP, building on earlier work (e.g. Roberts et al., 2014), in order to achieve an integrated view on safety and capacity. To the best of our knowledge we are the first to consider railway capacity in Timed CSP or a related formalism. Timed CSP extends CSP by a number of operators, of which we use mostly the process "Wait d", which waits for d time units, and the delayed event prefix operator a →d P, which first performs a, then waits for d time units, before it behaves as P. Timed CSP speaks always about minimal delays; i.e. Timed CSP guarantees that a process is inactive for d time units, the process however can be inactive for longer.

The Wait d process allows us to model the time that a train needs to travel from one end of a track to the other by setting d = track length / max speed, i.e., provided that the train driver does not exceed the speed limit, it will take at least time d between entering a track and leaving a track. (In the current modelling, we ignore the length of a train.) Should the train driver decide to drive slower, this is covered by this modelling as well, since d is a minimal delay.

Of the various capacity notions within the railway domain, we deal here with so-called theoretical network capacity. "Theoretical" capacity as we look at the capacity that in principle can be scheduled -- as opposed to the capacity actually used. Capacity is often regarded as an elusive concept, which is not easy to define and measure. In general, it can be described as below:

"Capacity determines the maximum number of trains that would be able to operate on a given railway infrastructure, during a specific time interval, given the operational conditions." (Isobe et al., 2012, page 57).

We illustrate our approach to capacity on an example given to us by Siemens Rail engineers. The track plan below (Figure 10) consists of two lines: a main line from A to C, and a side line from A to B. The speed limit on the main line is 90mph, on the side line it is 70mph.

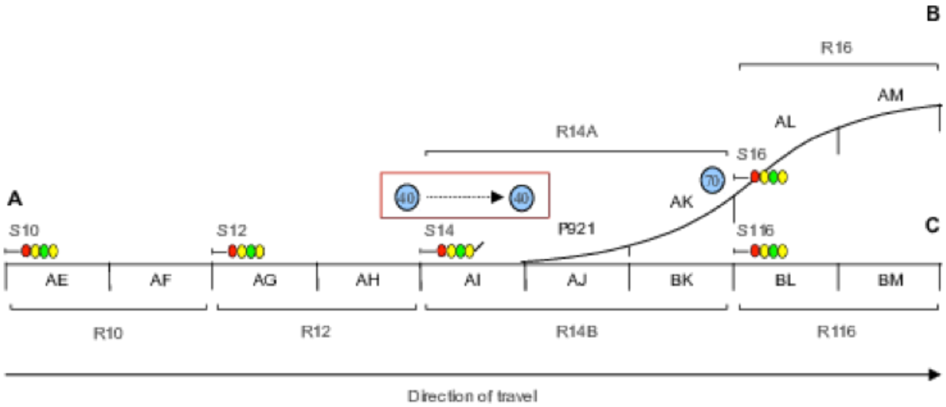


Figure 10: Illustrative Scheme Plan

In order to travel from the main line to the side line, a train has to pass the point on track AJ at a speed which at most can be 40mph. Here, we consider two scenarios. In scenario 1, there is a speed limit sign at the end of track AH that forces trains to slow down well before AJ. In scenario 2, this speed limit sign has been moved to the end of track AI, indicated by the dashed arrow in the picture above. The question is: how does moving this speed limit sign affect capacity? The answer is given by Figure 11.

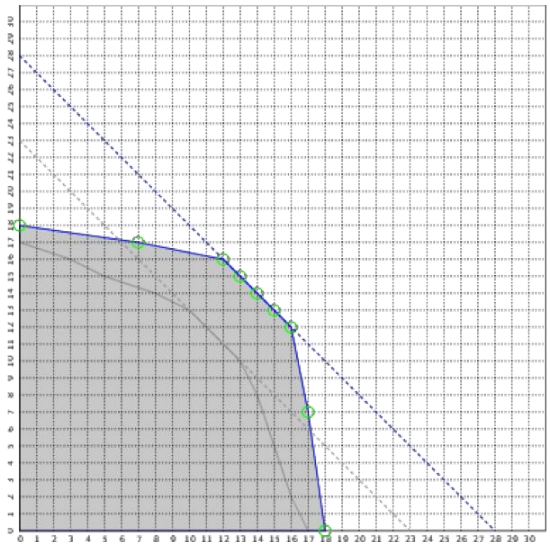


Figure 11: Capacity Enhancement from Moving a Speed Limit Sign.

Figure 11 shows the predictions that we obtain with our modelling. Given a period of time, on the Y axis we have the number of trains that can be scheduled on the side line, on the X axis we have the number of trains that can be scheduled on the main line. The solid blue line represents the maximal schedules for scenario 2. We see that we can schedule, e.g., 18 trains on the side line and 0 trains on the main line; 17 trains on the side line and 7 trains on the main line; etc. The grey area below the blue line represents the set of all possible schedules in scenario 2. The maximal schedules for scenario 1 are given by the grey line, i.e., 17 trains on the side line, 0 trains on the main line etc.

Speaking of capacity, we interpret the blue line (the grey line) as the theoretical network capacity of scenario 2 (scenario 1). The utilised capacity will be any point below this line, for scenario 2 the grey area indicates which choices are possible for the utilised capacity. We see these lines as the *characteristic curves* of the scheme plans under consideration. Note that – although not shown here – also the control tables influence these curves. In previous work we have demonstrated that control tables without overlap lead to higher capacity than control tables without overlap (Isobe et al., 2012).

Our modelling and analysis confirms the expectation that the Siemens rail engineers had with respect to the given example: moving the 40mph sign further down yields a capacity gain. It will be future work to further calibrate the numbers.

Beyond computing the characteristic curve of a rail node, it is also possible to check if a given schedule is possible or not. Take for example the following two schedules:

Table 6: Comparison of two schedules

Schedule 1: Possible			Schedule 2: Impossible		
Train ID	Time	Destination	Train ID	Time	Destination
1	0	Line 1	1	0	Line 1
2	100	Line 2	2	100	Line 2
3	200	Line 1	3	200	Line 1
4	300	Line 1	**4**	**220**	**Line 1**
5	400	Line 2	5	400	Line 2
6	500	Line 1	6	500	Line 1

Table 6 shows the times when a train enters the network by moving on to track AE. Schedule 1 and Schedule 2 differ on train 4, in schedule 1 it shall enter at time 300, in schedule 2 it shall enter at time 220. Our tool says that the first schedule is possible, while the second one is not.

It is also possible to produce possible schedules as shown by Table 7.

Table 7: Possible Schedule

Train ID	Time	Destination
1	0	Line 2
2	27	Line 1
3	54	Line 2
4	87	Line 1
5	120	Line 2
6	153	Line 1
"	"	"

Table 7 gives the beginning of a possible schedule for capacity (12, 16) – here, the first number, 12, denotes the number of trains on the side line, and the second number, 16, denotes the number of trains on the side line. Taking such a maximal schedule, we can produce from it a smaller schedule, say for capacity (6, 11), by leaving out 6 trains with a destination of Line 2 and 5 trains with a destination of Line 1.

While we see these results as promising, future work is needed in order to address

- Calibration – are the predicted numbers of trains realistic?
- Scalability – how can we treat example of realistic size?

3. Capacity Utilisation and Performance

In this section we examine the relationship between capacity utilisation and service performance, as delays are a key performance indicator of the system. Hence, this relationship is at the crux of the monitoring and evaluation of rail system performance. Therefore, we will define the key terms and provide some analysis of the relationships between capacity utilisation at nodes and secondary delays, which we highlight as an area of good practice.

3.1 Capacity Definitions

Many definitions of rail capacity are available and are applied for different purposes and in different contexts (Kontaxi and Ricci, 2012). For the purposes of this work and documentation, the most useful definition is the number of trains using a section of infrastructure per unit time (usually per hour or day, or sometimes per three-hour peak period). Similarly, as noted by UIC (2004) it is difficult, if not impossible, to identify a unique maximum theoretical capacity value for a railway system or sub-section, but this is not necessary for the calculation methods used to evaluate capacity utilisation, as described below.

Capacity Utilisation is a measure of the extent to which the theoretical capacity of a section of a railway system is being utilised, or consumed, and is expressed as the percentage of the time period under consideration during which the infrastructure is occupied.

3.2 Capacity Utilisation at Nodes

As the capacity bottlenecks of the railway system, nodes (i.e. stations and junctions) tend to limit overall capacity, and an understanding of their practical capacity utilisation limits is therefore particularly valuable. However, because of their variability and – in some cases – complexity, both in terms of layout and train operations, they are difficult and time-consuming to model and assess. For these reasons, and in contrast to the 'plain line' links between nodes, standard methodologies have not been available until comparatively recently, and advisory upper limit capacity utilisation values have not yet been established.

The updated UIC 406 'Capacity' leaflet (UIC, 2013) extended the assessment methodology from links to nodes, but did not specify any recommended upper limit values – again, this partly reflects the variability and potential complexity of station and junction layouts and operations. The follow-on ACCVA (Assessment of Capacity Calculation Values) project included among its objectives the identification of such upper limits, and this is part of the DITTO work, but the results to date have confirmed the difficulty of identifying unique values, independent of location and layout.

3.3 Primary and Secondary Delays and their attribution

Chief measures of performance on the railway network in Britain are train punctuality and reliability, i.e. lateness (caused by delays) and cancellations, the focus of this work being on delays. Delays are categorised as primary or secondary, primary delays being attributed to trains suffering initial delays, such as mechanical failures, and secondary delays to other trains that are in turn delayed as knock-on effects of the primary delays.

On Britain's railways, delays of three minutes or more are recorded and attributed in the TRUST (Train Running Systems on TOPS (Total Operations Processing System)) database. The attribution includes details of the service and Operator affected, the party responsible for the delay (it could be the affected Operator, another Operator, Network Rail or an External cause), and the date, time, duration and location of the delay. As well as recording whether the delay is Primary or Secondary, additional, more specific, cause codes are also recorded, together with other relevant data. Historic delay data is available, together with explanatory notes, on Network Rail's website, listed under 'Historic delay attribution', at

https://www.networkrail.co.uk/who-we-are/transparency-and-ethics/transparency/datasets/#H

Note: the location information in the dataset takes the form of numeric STANOX (Station Number) codes, and it will normally be necessary to map these to the corresponding location names or TIPLOC (Timing Point Location) codes, using the mapping included with the dataset. This mapping is not necessarily one-to-one.

3.4 The Relationship between Capacity Utilisation and Secondary Delays

As capacity utilisation levels increase, the system becomes more vulnerable to secondary delays, which cannot easily be absorbed by the system, and can instead spread quickly and widely across the network. An illustration of the typical, theoretical relationship between capacity utilisation and secondary delay is shown in Figure 12, where secondary delay increases exponentially with capacity utilisation (note: the relationship shown is indicative only, and is not intended to show a suitable upper limit for capacity utilisation).

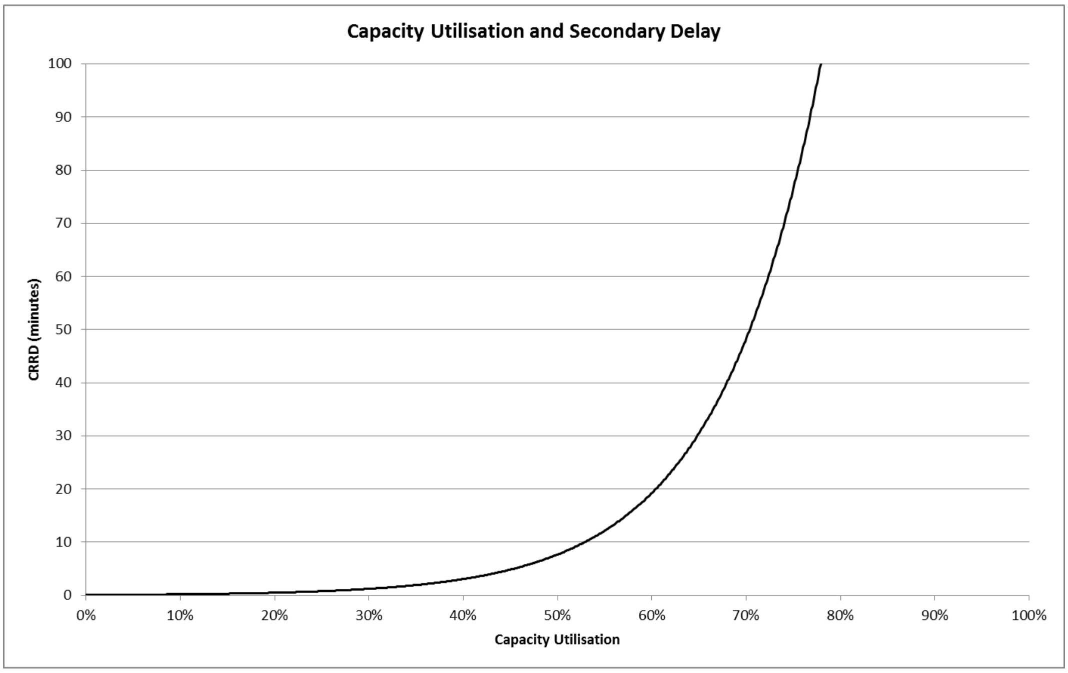


Figure 12: Theoretical Relationship between Capacity Utilisation and Secondary Delay

A more generalised representation of the interdependencies involved is shown in Figure 3.2 (UIC, 2004). Performance and secondary delays are a reflection of timetable stability, while capacity utilisation increases with the number of trains and their heterogeneity, and is also affected, less directly, by their average speed. Figure 13 illustrates two service types: mixed train working with high average speeds and service heterogeneity, but with modest capacity and stability (this will be akin to our East Coast Main Line case study) and metro train working with lower average speeds and heterogeneity but higher capacity and stability. It should be noted that low levels of capacity utilisation are not a guarantee of a stable, reliable timetable, since this also depends upon the detailed planning of and interactions between services; however, all things being equal, higher levels of capacity utilisation are likely to result in reductions in timetable stability.

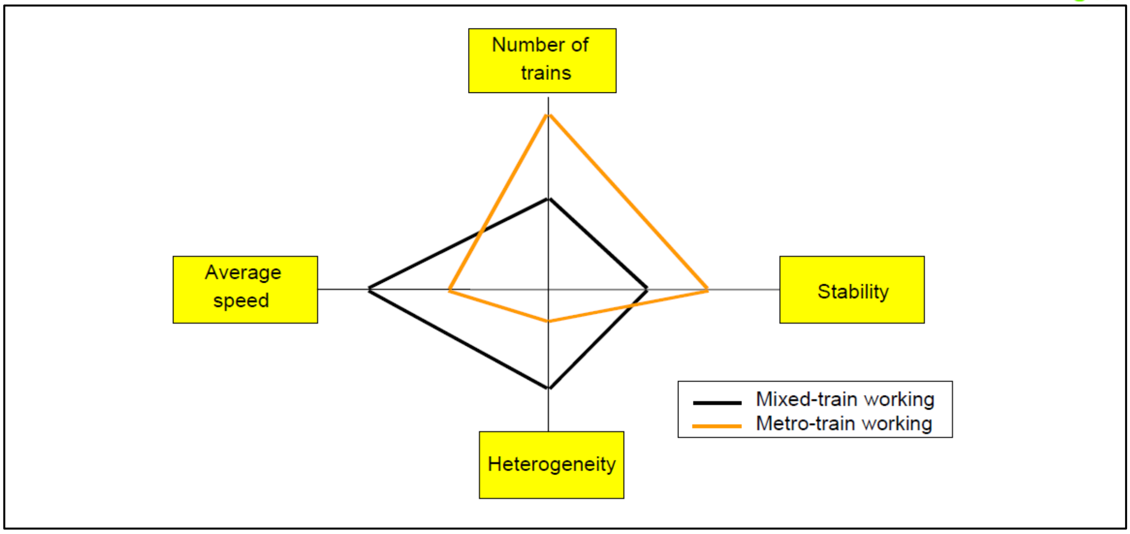


Figure 13: Interdependencies between Operating Characteristics and Timetable Stability

Good Practice IV: Analysis of the Relationship between Nodal Delays and Capacity Utilisation

Nodal delays can be identified in the TRUST datasets as records with the same start and end locations, and can be extracted for locations of interest by selecting the appropriate STANOX codes or, once the appropriate mapping has been done, the corresponding TIPLOCs or location names. Secondary delay records and different causes of delay can similarly be selected; if the focus (as was the case for DITTO Railway Systems) is on congestion-related reactionary delay (CRRD), the following cause codes should be used: YA, YB, YC, YD, YE, YF, YG, YO[2]. An example of the attribution process which emphasises the role of nodes, and particularly stations, in delay propagation is shown by Table 8.

Our starting point was to use the Capacity Charge Recalibration dataset (2012) used by Arup in work for ORR. This contained 458,000 records of nodal delay – some 26% of the total of over 1.74 million delay incidents. From this dataset, 57,958 records were extracted for the London and North Eastern route which included 755 TIPLOCs. It was found that 146 nodes accounted for over 90% of nodal delays and all these nodes related to passenger station or freight terminals. The top six categories, accounting for 83 nodes and 41,612 delay incidents (over 70% of the route total) are shown overleaf.

[2] See http://nrodwiki.rockshore.net/index.php/Delay_Attribution_Guide

Table 8: Example of Attribution of Delays to Nodes
(See also Armstrong and Preston, 2015)

Node Classification	No. of Nodes	No. of CRRD Incidents
Complex, Major Station	11	18,887
Freight Terminal	35	6,847
Complex, Medium Station	10	5,734
Complex, Minor Station	8	5,721
2-track through Station	15	2,393
2-track Terminus	4	2,030

As indicated above, only limited guidelines are available for capacity utilisation calculations at junctions and stations: no formal guidelines are available for the Capacity Utilisation Index (CUI) approach used in Britain (Gibson et al., 2002), and, while updated UIC 406 provides an outline methodology, it does not include any guidance for capacity utilisation calculations for trains calling (i.e. arriving, stopping and then departing) at stations, or for trains arriving and terminating their journeys, and then going out of service or forming subsequent originating departures.

It will typically be impractical to identify a single level of capacity utilisation for a station or junction, unless it is formed of a single track and platform, or a single switch or set of points. For more complex locations, it will be necessary to subdivide the layout into separate tracks, switches and platforms, and assess their individual levels of capacity utilisation, paying particular attention to the busiest and most critical infrastructure elements. Depending upon the nature of the elements in question, individual levels of capacity utilisation can be calculated by means of the standard timetable compression approach, using minimum headways, junction margins, dwell times, platform reoccupation times and turnaround times as appropriate (location-specific or general values for these can be found in Network Rail's Timetable Planning Rules – see http://archive.nr.co.uk/browse%20documents/Rules%20Of%20The%20Route/Viewable%20copy/roprhome.pdf).

For the purposes of investigating relationships between capacity utilisation and performance at nodes, capacity utilisation values should be calculated for, and delay data records assigned to and aggregated for, common time bands (the delay records are assigned to time bands on the basis of their recorded start time). These time bands will typically be of one hour, but users may choose their own to suit their circumstances and needs: previous work, including DITTO Railway Systems, has found that one-hour time bands can produce quite 'noisy' results, and three-hour time bands were found to produce improved levels of correlation. The use of three-hour time bands has the additional advantage of mapping onto the typical morning and evening peak travel periods of 07:00 - 10:00 and 16:00 - 19:00 and also fitting the three-hour intervals between the peaks. Separate capacity utilisation calculations and delay allocations will typically need to be undertaken for weekday, Saturday and Sunday timetables, to reflect their different characteristics. In cases where occupancy of an infrastructure element ‘straddles’ multiple chosen time periods (e.g. a train arrives in a platform at 06:58, and departs at 07:03), the occupancy time should be split between the two time periods, and the resulting occupancies for both periods compressed and assessed in the usual manner.

For simple, two-track, two-platform stations, delay data can be separated and assigned by direction, and the relationship with capacity utilisation plotted, as shown in Figure 14 for Platform 1 (westbound) of Knaresborough station.

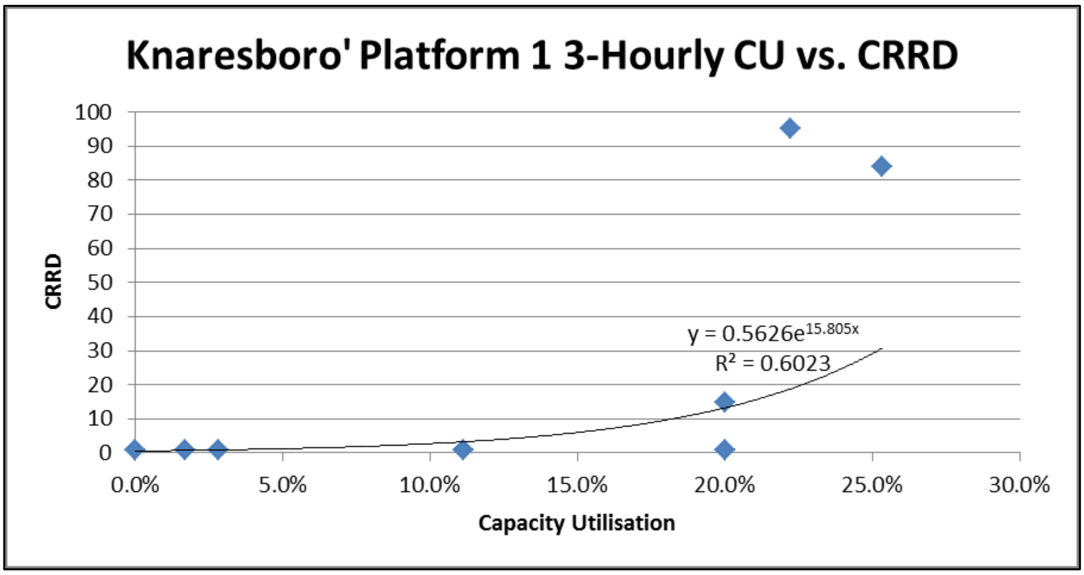


Figure 14: 3-Hourly CRRD vs. Capacity Utilisation for Knaresborough Platform 1

For more complex stations, the assignment of delay records requires further consideration, and simply assigning them by direction may not be sufficient, and could produce misleading results. For example, Grantham station, on the East Coast Main Line is shown in Figure 15, with Platform 1 on the southbound main line, and Platform 4 on the branch. The relationships between capacity utilisation and aggregate southbound delay for both platforms are shown in Figures 16 and 17.

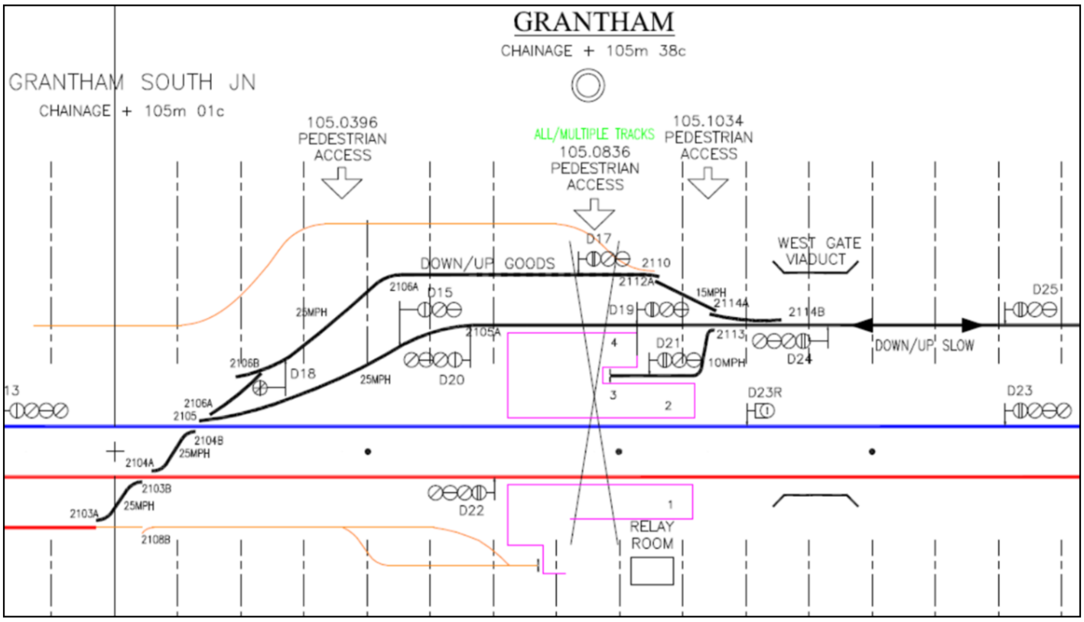


Figure 15: Grantham Station Layout

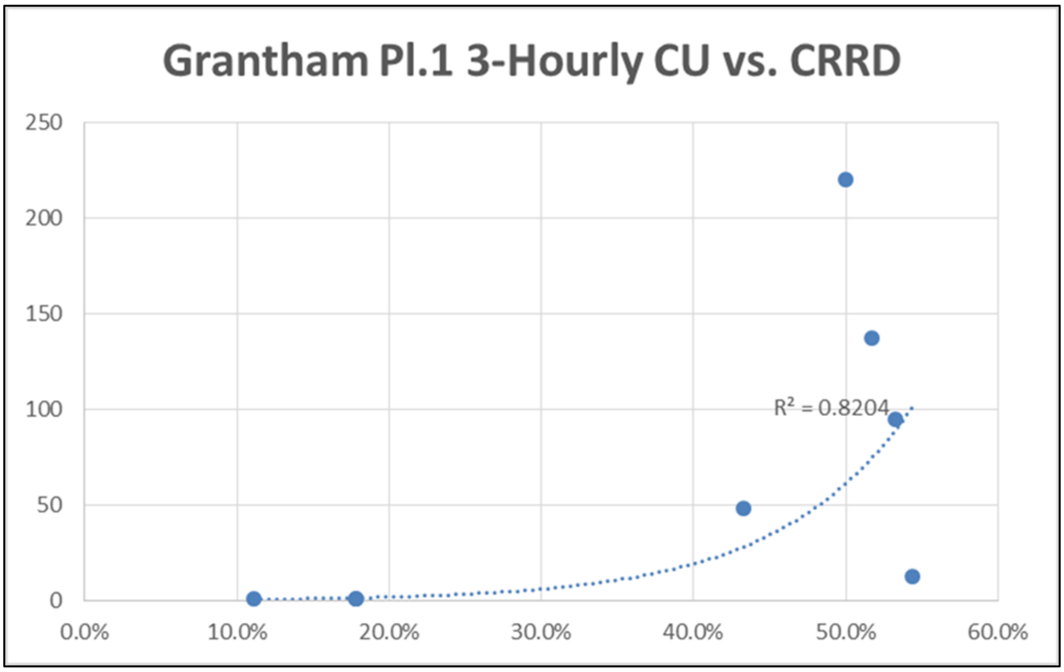


Figure 16: Capacity Utilisation and Aggregate Southbound Delay at Grantham Platform 1

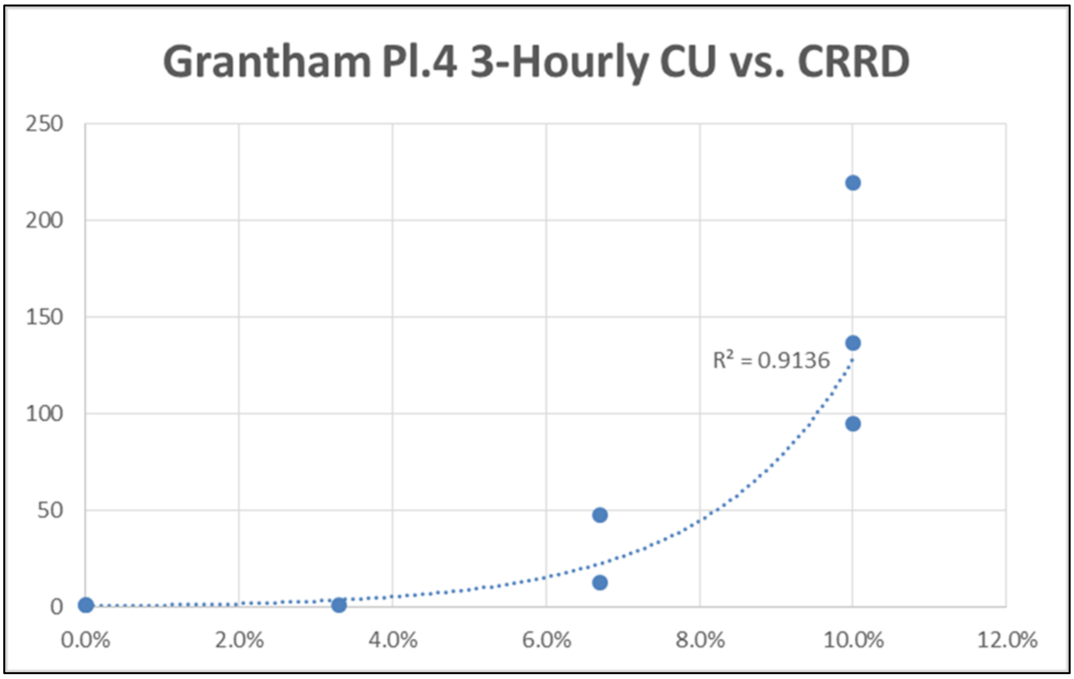


Figure 17: Capacity Utilisation and Aggregate Southbound Delay at Grantham Platform 4

The correlation shown for Platform 1 is quite high, at 82.04%, and delay is seen to increase quite sharply at what seem to be sensible levels of capacity utilisation, i.e. 50% - 60%. For Platform 4, the correlation is greater, but the capacity utilisation levels are quite low (reflecting traffic levels at and past the platform), and delay is seen to increase markedly at very low capacity utilisation levels of 8% - 10%, which is both pessimistic and misleading. Although this might be reflecting a capacity constraint upstream or downstream of Platform 4, e.g. the crossover at the southern throat of the station, the more likely explanation is that the attribution of delays to Platform 4 leads to a spurious correlation. This, in turn, confirms the requirement for more detailed assignment of delay data for even slightly complex infrastructure layouts. Such an approach was applied to the assessment of Peterborough station, whose layout is shown in Figure 18.

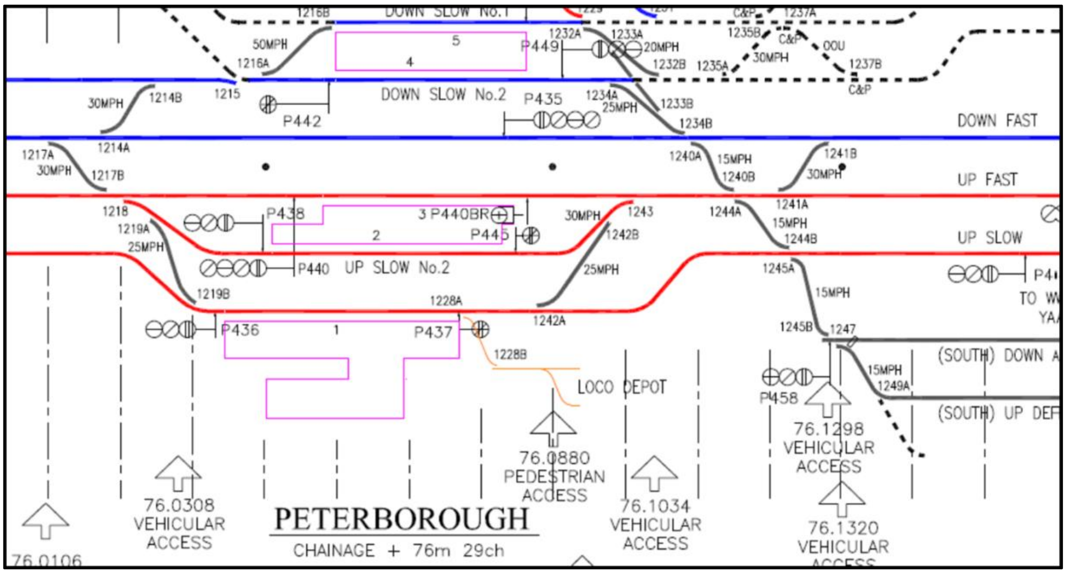


Figure 18: Peterborough Station Layout

The results of the initial analysis of switch 1218 (on the Up Fast line, to the left in Figure 18) are shown in Figure 19, based upon aggregate southbound delay data.

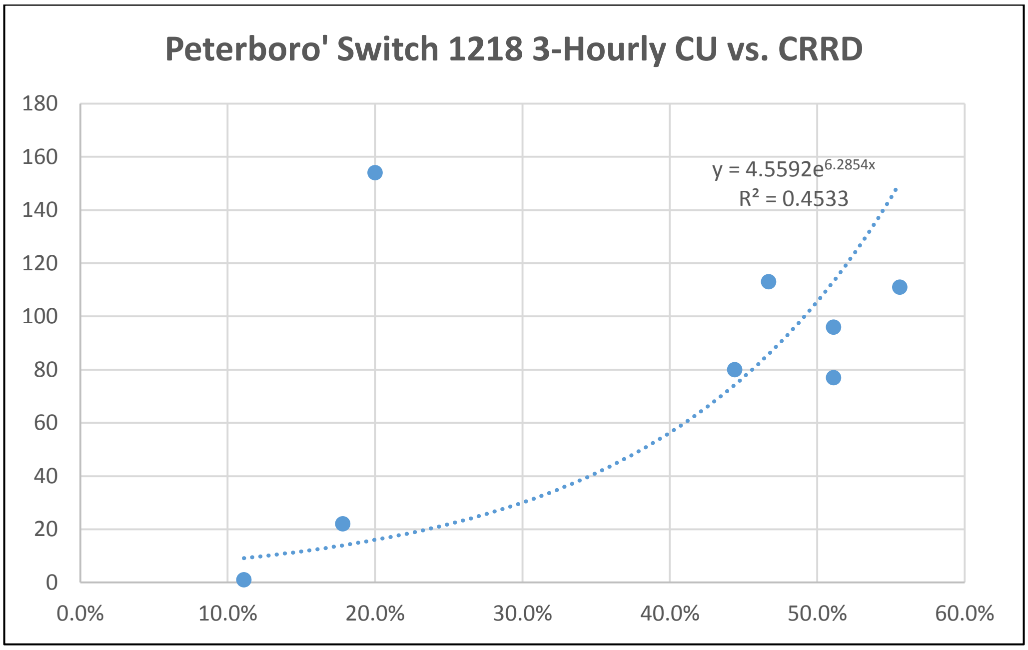


Figure 19: Capacity Utilisation and Aggregate Southbound delay at Peterborough Switch 1218

It can be seen that the correlation between capacity utilisation at the switch and aggregate delay is quite poor, partly because of the outlying data point at 20% CUI and 154 delay minutes. The equivalent results for the disaggregated delay data are shown in Figure 20.

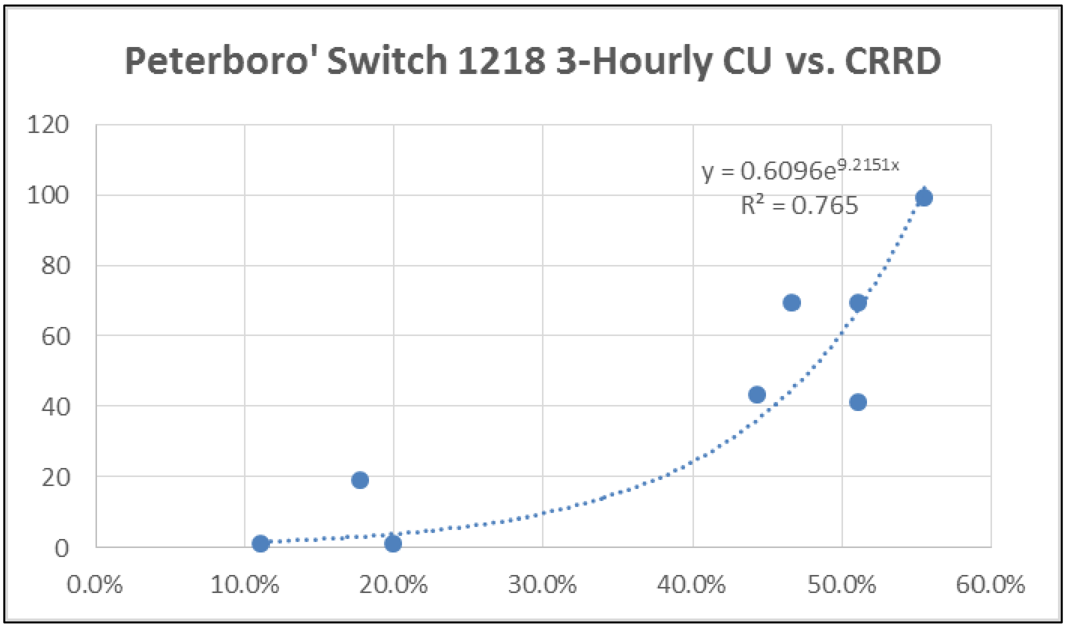


Figure 20: Capacity Utilisation and Disaggregate Southbound Delay at Peterborough Switch 1218

The revised relationship is based on delays associated with train movements through platforms 2 and 3, all of which use switch 1218. It can be seen that a higher level of correlation is achieved, partly due to the removal of the outlying observation, which was associated with a freight movement on the west side (platforms 4-7) of the station. In order to obtain meaningful results, it is therefore important to disaggregate delay data for even slightly complex station layouts, and assign the records to the relevant infrastructure elements. However, this process is not straightforward, and manual assignment is quite time-consuming; work is underway to automate the process.

4. Timetable Optimisation at Nodes

Our optimisation problem is based on the job shop scheduling formulation associated with Liu and Kozan, 2009 – see also Bektas et al., 2015. Our initial work in the pre-cursor OCCASION project was with deterministic formulations (Paraskevopoulos et al., 2015). The associated problem of scheduling trains in a stochastic environment is complex, although a standard method of overcoming this complexity is to use a sample average approximation (SAA) approach (Kleywegt et al., 2002). The stochastic optimisation that we develop has been applied at a meso-level to individual stations (which are themselves assemblages of nodes and links). At a macro-level, we model the network as a Multi-Commodity Network Design Problem (MCNDP) (Bektas et al., 2010). The job shop scheduling and the MCNDP are iterated until an overall feasible and robust timetable is found.

4.1 Stochastic Optimisation at One Node

Our work on the interrelationships between rail service performance and capacity utilisation has been outlined in the previous section. A viable approach to keeping up with increasing numbers of railway passengers is to run more services at peak times; that is, add more services to the timetable. However, more traffic means more conflicts amongst trains; the tighter the capacity constraints, the more conflicts. Without sufficient buffer times to absorb uncertain delays, the delay of one train might propagate over the entire network.

Given this, we address a realistic timetabling problem by considering the number of services offered along with their reliability. The approach adopted is detailed in Kovacs et al. (2016a) but a summary is provided here. A two-stage stochastic programming model has been developed for generating timetables with the required number of services at the tactical level. Different recourse actions to recover from delays are taken into account at the operational level (e.g., speeding up trains). The model considers conflicts among different types of trains (e.g., express and freight trains) at different locations (e.g. points, junctions, and platforms). In our representation, we define junctions as complex assemblages of points, typically where passenger and/or freight routes join/separate. Points involve simpler layouts, typically where the number of tracks on a given route changes.

Small instances can be solved by commercial solvers; however, for solving large instances, we developed a large neighbourhood search algorithm (LNS). In each iteration, the algorithm executes two phases: in the first phase, a feasible order among trains is determined; given this order, the reliability of the timetable is optimised in the second phase.

Train services are scheduled by a recursive job shop algorithm that is guaranteed to insert a service into a given timetable if a feasible insertion position exists. Appropriate buffer times are incorporated into the timetable by a greedy algorithm and linear programming in order to absorb uncertain delays.

More complicated recourse actions have been tested which include changing the platform assignments if a platform is blocked, and allowing trains to overtake if an express train is stuck behind a regular train. However, our results suggest that considering complicated recourse actions can be avoided in the timetabling phase. This result remains to be verified on railway systems with large-scale delays.

The LNS has been tested extensively on benchmark instances. The results show that the algorithm is able to generate feasible timetables even when capacity constraints are tight. The solution quality increases with a larger number of iterations. The generated results are on average 6.6% worse than the best known solution; the average computation time is 4.1 hours. (Kovacs et al., 2016b)

The results of a case study appear to indicate, at a first-cut, that there is room for increasing the operational capacity at our one node case study - Peterborough. However, as the availability of rolling stock and staff, as well as shunting movements within the station, have not been considered here, the results should be interpreted as a best case situation. Nevertheless, they suggest that it is possible to increase the capacity utilisation of the existing infrastructure by using state-of-the-art optimisation techniques, as opposed to alternative strategies that are significantly more expensive and involve reducing headway times (e.g., by updating the signalling system and/or improving the braking performance of trains) or laying new tracks.

For our timetable optimisation modelling, our model consists of a network layout, a set of trains, and a set of delay scenarios. An illustration of the network layout used to examine Peterborough is given in Figure 21. We consider several stations with different numbers of platforms, points, junctions, double track lines, quadruple track lines with fast and slow tracks in each direction and single track lines that are traversed in both directions.

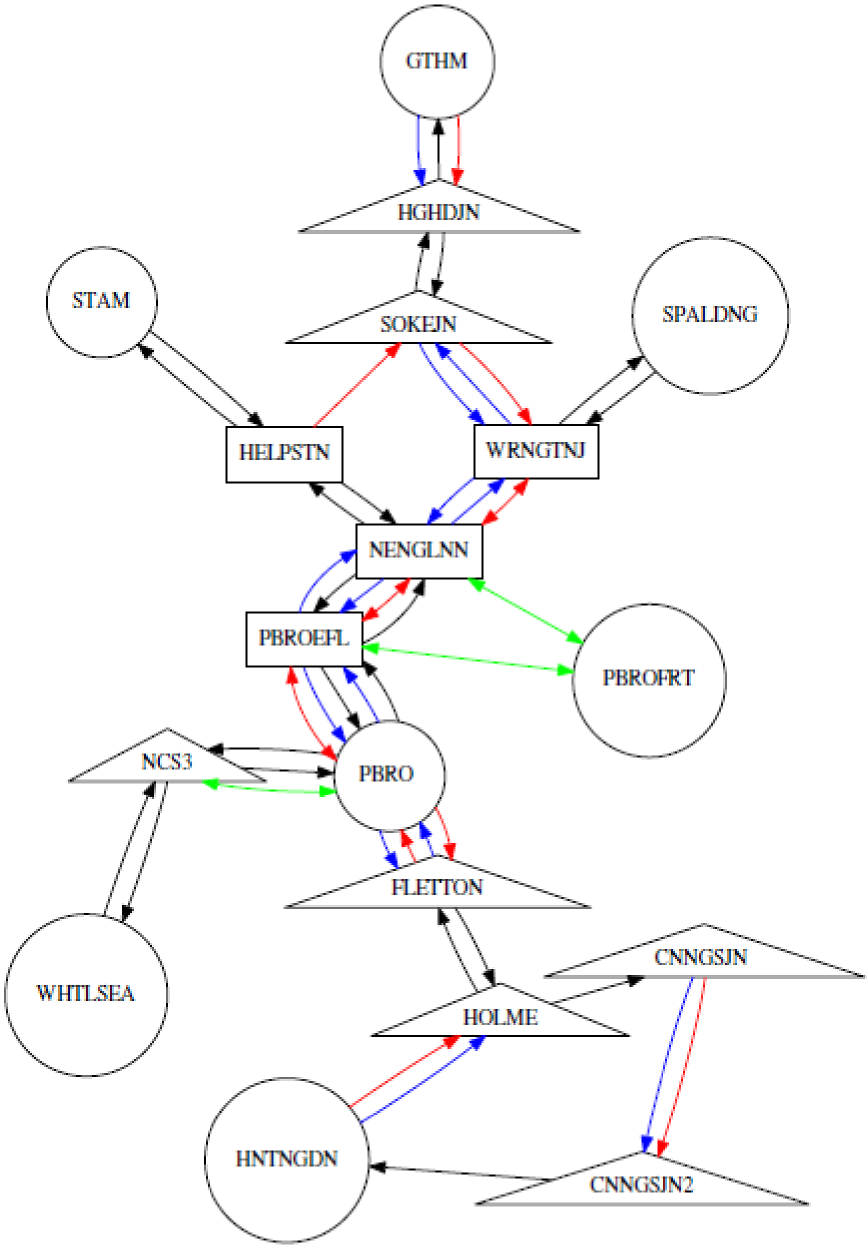


Figure 21: Example of the Peterborough Network Layout.

The set of trains travelling along the network is provided in the form of a timetable. For each train, we are given the route (i.e., a sequence of stations, junction, and points); preferred arrival and departure times at different locations; and the type of train (e.g., freight, express, or regular). Furthermore, a delay scenarios list is provided, where for each train, the duration and location of the delay is specified.

Our case study focuses at the rail network surrounding Peterborough station. The wider network layout involves seven stations (circles), four junctions (rectangles), and seven points (triangles). The network comprises 47 arcs, each arc representing a track segment that can either be a fast, slow, main, or freight track. Freight trains can be assigned to slow, main, and freight tracks; regular trains to slow and main tracks; and express trains to fast, slow and main tracks.

The set of trains is selected from a representative weekday (Wednesday, 4/11/2015). From the national timetable, we select all passenger and freight trains (including empty movements) that visit Peterborough between 7am and 9am. In total, we consider 55 services in the reference timetable. The average speed of express, regular and freight trains is assumed to be 125, 100, and 75mph, respectively. The time required for acceleration and deceleration is considered by decreasing the average speed by 7% if a given train has to stop once in our model, by 14% with two stops, and by 21% with at least three stops.

Delay information is gathered from historical delay data provided by Network Rail. More than 6 million delays were recorded between 1/12/2013 and 18/04/2015 (i.e., over 503 days). As primary delays are the model input, we filter out irrelevant information on secondary delays. Almost 800,000 delay records remain. The efficiency of the model algorithm is measured by its ability to mitigate delay propagation by incorporating proper buffer times into the timetable. The smaller the secondary delays, the better the objective value, and the better the solution.

In a second step, we match filtered trains (T) with trains in the delay data (D). There is no unique identifier that unambiguously links trains in the two sources of data. Therefore, we apply the following strategy. We take the set of relevant trains, the delay data, and a time margin TM. We then match T with D if: (i) T is a passenger train (delays of freight trains are not considered); (ii) T and D have the same origin and destination; and (iii) T departs within the departure time of D +/- TM.

Delay scenarios are sampled in a Monte-Carlo fashion. In each scenario, and for each train, we decide by Bernoulli trial whether or not it is delayed; if yes, we associate the location and duration of the delay. The length of the delay is modelled by a Gamma distribution.

Good Practice V: Stochastic Job Shop Scheduling at One Node

The results of applying stochastic job shop scheduling to Peterborough are shown by Figure 22. Out of 200 timetables tested, 184 were feasible and this suggests that an additional 40 trains in the morning peak hour at Peterborough is feasible, although not necessarily desirable. This preliminary result would represent an increase in service of around 73% but an increase in an index of delays (as measured by the objective function (OV), which is a combination of maximum and mean delays) of around 144%. This represents a 3.64% increase in delays with each train.

Of these 40 additional trains, 18 will run to/from Grantham. Grantham is modelled as having 46 trains in the morning peak, so this would represent an increase in service of 39% at this node. Figure 22 involved over half a year of computing time (ran in parallel on the University of Southampton's Iridis 4 supercomputer). The pattern of delay increases appears to follow a linear rather than an exponential function.

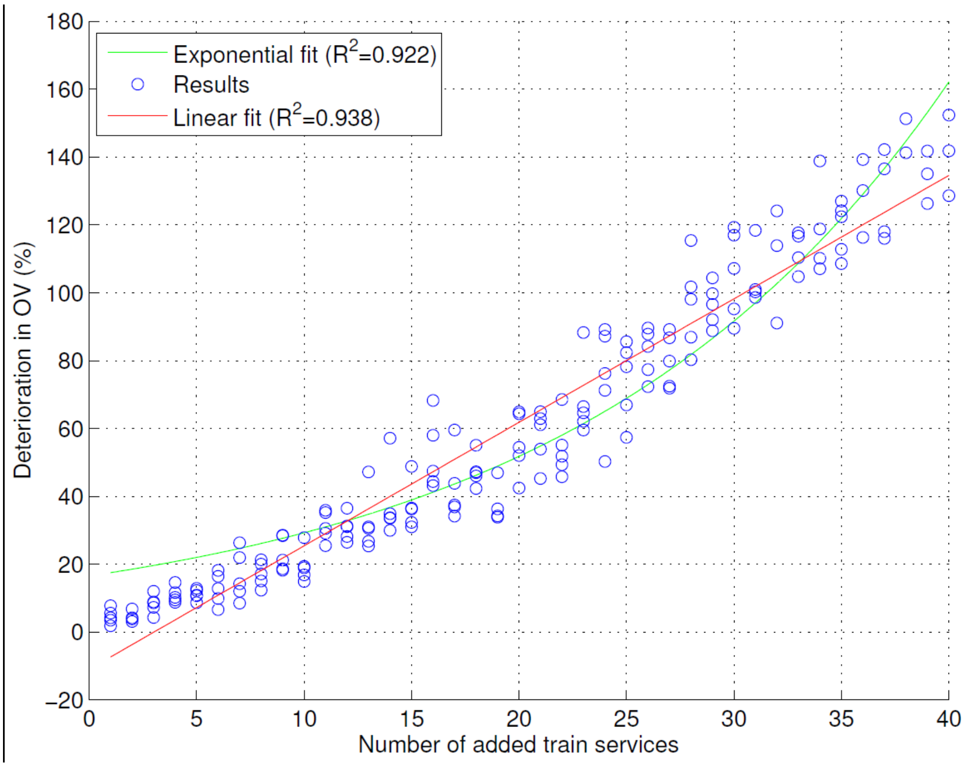


Figure 22: The Relationship between Additional Train Services and the Delay-based Objective Function at Peterborough station

4.2 Multi Commodity Network Design Problem Type Applications

Figure 22 suggests that Peterborough station (which was remodelled in 2014) is not a bottleneck. However, to confirm that these services are feasible we need to examine a larger area but a larger network will lead to higher complexity. Our first attempt to examine this issue was to treat it as a Multi-Commodity Network Design Problem (MCNDP), with the commodity being an individual service and the network being a time-space diagram where space refers to rail network layout. The design is then the trajectory of each service. This approach is widely used in logistics to make strategic decisions (Goetschalckx et al., 2002).

Complexity is reduced by aggregating data and simplifying constraints. However, some important aspects of railway operations have not yet been considered in the MCNDP including the impossibility of overtaking on the same track, the possible availability of multiple tracks and the imposition by current signalling systems of consistent headways.

In a second attempt a mathematical programming approach was adopted based on column generation (see, for example, Desrosiers and Lübbecke, 2005). Column generation algorithms are convenient when the number of variables in the linear programme is great. In our model, the variables are the possible trajectories that a service can be assigned to where each trajectory is a path through the time-space network. The goal is to schedule as many services as possible by assigning at most one trajectory per service. A prototype has been developed but the algorithm didn't work well enough to tackle the desired scope of the network and as a result the column generation approach, which has some novelty, was abandoned.

Our third and final attempt used a variant of the job shop algorithm as described above but the algorithm was simplified in order to solve a larger case study. In particular, uncertain delays and spread constraints are ignored and conflicts at large junctions are simplified. The focus is exclusively on generating a feasible timetable and delay propagation is ignored.

The network involves the main trunk of the East Coast Main Line between Alexandra Palace and Doncaster via Peterborough, but also includes branch lines to, for example, Cambridge, Leicester, Lincoln and Nottingham. In total, the modelled network consists of 35 stations, 17 junctions, 21 points, and 215 arcs and covers all trains from the national timetable on an average weekday between 7am and 9am. We additionally consider 16 services that run as required (denoted by Q paths in the CIF files). In total, there are 326 services in our reference timetable. This results in 4,968 operations in the job-shop model (see Kovacs et al., 2016a). In contrast to the earlier approach where the travel times were calculated by dividing the distance by the average speed[3], we use the travel times as indicated in the timetable that additionally include scheduled waiting times[4]. The layout is presented in Figure 23, in which arcs represent track segments that are either fast (blue), slow (red), main (black), or freight (green) tracks. It is assumed that a given train can be assigned to any track and to any platform as this approach is currently applied on heavily utilised sections. However, no changes are allowed on directions, even on bi-directional tracks.

[3] Working with the average speed was appropriate for express trains that do not stop within the network. That is, increasing the travel times by assuming lower velocities would make the original timetable infeasible. However, it is unclear whether or not slow trains can achieve the assumed speed.
[4] Scheduled waiting times were minimised in the OCCASION project. As this may have unexpected consequences on the reliability of the timetable, we model the buffer times as given to prevent any knock-on effects.

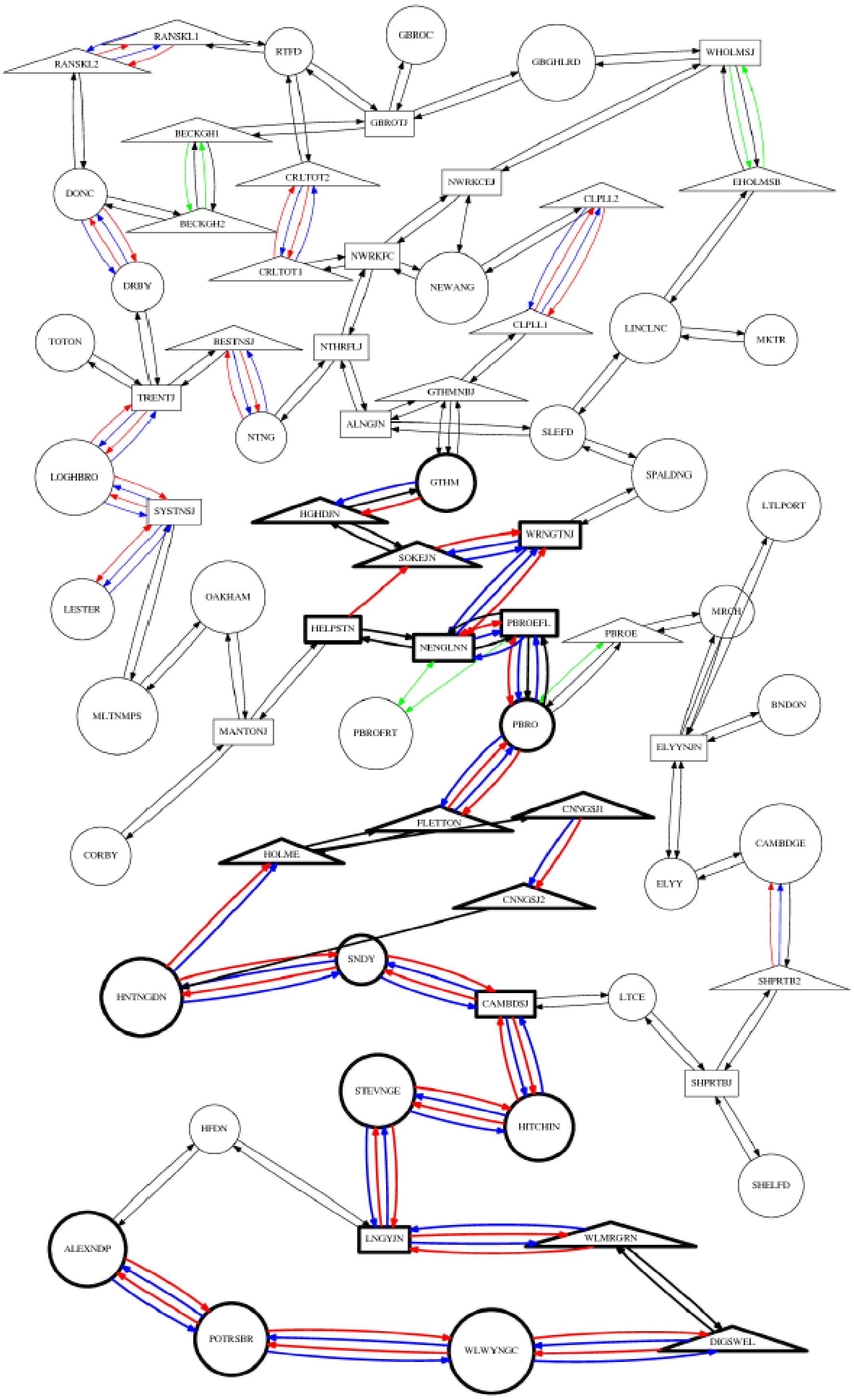


Figure 23: Network Used in the Case Study.

Stations are indicated by circles, points by triangles, and junctions by rectangles. The East Coast Main Line is highlighted.

Good Practice VI: Optimisation at the Network Level

For the wider network optimisation we use deterministic job-shop scheduling, as explained above. For each number of additional services from 1 to 10, we generate five instances by replicating randomly chosen services from the reference timetable to account for the variability of random choice. Four scenarios are examined: adding services that either pass through (i) Peterborough, (ii) Peterborough and Alexandra Palace, (iii) Peterborough and Grantham, or (iv) Peterborough, Grantham, and Alexandra Palace. The LNS is aborted either after 90,000 iterations, or 30 hours, or when a feasible timetable is found, whichever is reached first. Each instance is solved three times to account for the stochastic nature of the solution algorithm. Despite the generous computational resources made available, it is still possible for a timetable with additional services to be declared infeasible even though it may not be. This is due to the nature of the algorithm used, which does not guarantee optimality of the solutions identified.

Figure 24 summarises the results. The horizontal axis shows the number of additional services. The vertical axis shows the number of feasible timetables identified out of the five tested, for varying numbers of additional services. The routes of the new services are distinguished by colour. In particular, the services that pass thorough Peterborough are indicated in blue, services that pass through Peterborough and Alexandra Palace in red, services that pass through Peterborough and Grantham in yellow and services that pass through all three stations in green.

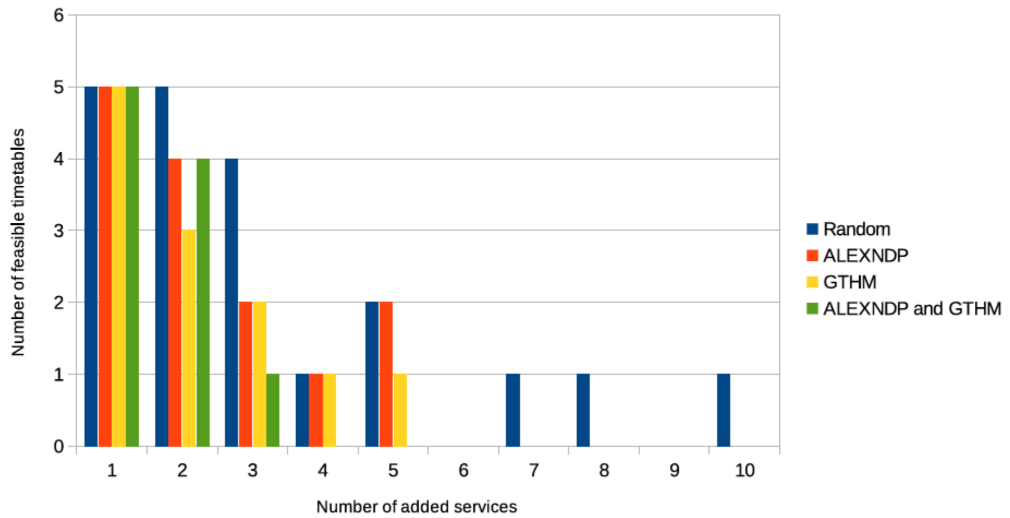


Figure 24: Number of Feasible Timetables for Different Numbers of Added Services with Different Routes.

The results shown in Figure 24 indicate a highly utilised network. The results also indicate that it is possible to insert up to ten services into the reference timetable[5]. However, the capacity drops significantly if the additional services are those running along prominent routes. In particular, only five services could be added to the timetable between Peterborough and Grantham or between Peterborough and Alexandra Palace. It turns out that the services that run between Alexandra Palace and Grantham (through Peterborough) are the hardest to schedule – at most three of such services could be replicated.

[5] The random selection of the new services added is the main reason for failing to identify a feasible timetable with nine additional services. There might be more services on busy routes or services with long routes.

5. Dynamic Simulation and Advanced Train Control

The ERTMS (European Rail Traffic Management System) is expected to significantly increase the railway capacity, reliability and punctuality. However, the increased capacity means shorter train headways and stronger interference among trains using the same infrastructure, which may lead to longer delay and higher energy consumption. Therefore, it is important to investigate, in a practical train operating environment, whether the claimed performance improvement can be achieved. A railway traffic simulation platform can provide convenience for such investigation before the ERTMS is widely implemented in practice. As a part of the DITTO project, the simulation platform could help to generate detailed train operation data for the capacity analysis of the railway network, and the analysis of time- and energy-efficiency of timetables generated by other scheduling tools.

The European Train Control System (ETCS) is the component of ERTMS for signalling, train control and train protection. ETCS Levels 2 and 3 are the two most advanced application levels under the existing categorisation of ETCS, and they are distinguished from other lower levels in the way that they are radio-based (SUBSET-026). The most obvious difference between ETCS Level 2 and Level 3 is that the former is a fixed-block system while the latter is a moving-block system.

A microsimulation platform was established for simulating the key functions of ETCS Level 2/3, which works based on the interaction of the train, the radio block centre (RBCs), the control centre, and other trackside equipment such as interlocking. Various advanced train operation and control rules are also developed and investigated, such as train following control and energy-efficient train control.

5.1. Principles of TrackULA

The TrackULA (standing for Track Unified simuLation Algorithms) model is developed for the rail traffic simulation. It is adapted from the existing road-traffic microsimulation model DRACULA (standing for Dynamic Route Assignment Combining User Learning and microsimulAtion) (Liu, 2005, 2010; Liu et al., 2006).

TrackULA is a microscopic simulation model which represents the movement of individual trains. It is based on discrete-time simulation where the train status is updated at a fixed time interval. It can model stochastic travel times (as opposed to deterministic, scheduled times) and disruption. It also allows heterogeneous train characteristics, train operating and train drivers behaviour, as well as variations in drivers' experience and driving behaviour with a given probability distribution.

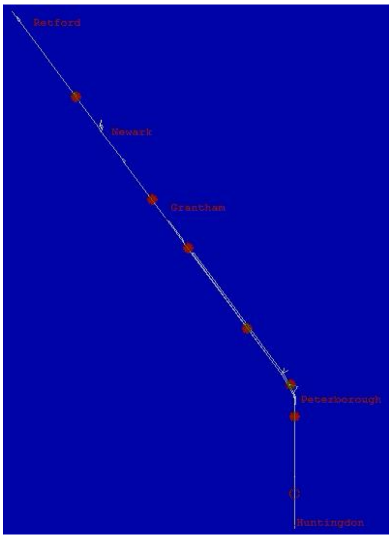
The core functions of TrackULA include:

- Simulation loop based on fixed time increments;

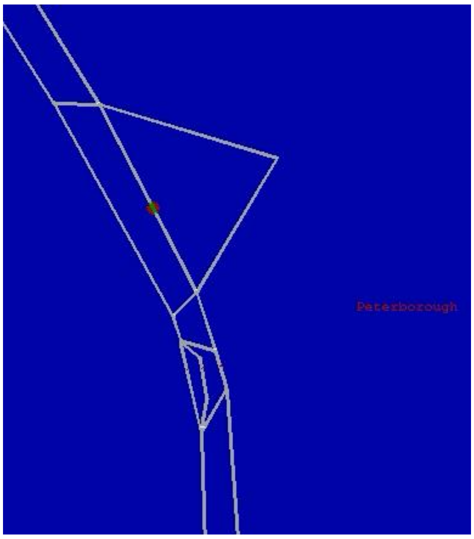
- Railway network representation;
- Railway timetable and train route representations;
- Train and driver behaviour representations;
- Train movement simulation;
- Control command simulation;
- Simulation outputs, including individual trains’ second-by-second space-time trajectories as well as route/link-based and network-wide statistics (means, variances and distributions when involving stochasticity).

Good Practice VII: Microsimulation Applied to Rail

The resultant microsimulation model of rail operations, based on the principles of road traffic simulation, is illustrated by the snapshots in Figure 25 that show the simulations for the section of East Coast Main Line (ECML) from Retford up to Huntingdon.



(a) ECML: Retford up to Huntingdon



(b) Around Peterborough Station

Figure 25: Snapshots of TrackULA Simulation for the ECML Section (Retford up to Huntingdon): (a) the Whole Section; (b) around Peterborough Station.

5.2. Simulation of ERTMS/ETCS

The trackside systems of ETCS Levels 2 and 3 share some common components such as RBC and balise. RBC is a centralised unit in charge of the safe movement of all trains under its supervision, delivering moving authorities (MAs) to all supervised trains based on information received from trains and other trackside systems. The balises (or balise groups) are mainly for location referencing. Some trackside equipment (such as track circuits or axle counters) are indispensable for ETCS Level 2 for train detection and train integrity supervision, but not for ETCS Level 3 since such functions will be done by the train itself.

The on-board ETCS system is responsible for the train protection including speed supervision and prevention of overrunning the MA, as well as displaying cab signalling to the train driver.

The RBCs' rules for determining the MA, as well as the quality of train-RBC communication, will affect the operation of each train and thus the performance of the whole railway system. Therefore, in our simulation platform, flexibility is provided for adopting different MA generation rules, and the non-ideal train-RBC communication situations are considered. Furthermore, the RBCs are allowed to exchange information from other traffic management functions such as the control centres, to receive real-time operation commands such as temporary speed regulation.

The following flow chart (Figure 26) illustrates the simulation, based on simplifications of the features of ETCS Levels 2 and 3 provided in the Functional/System Requirements Specification (ERA/ERTMS/003204, SUBSET-026).

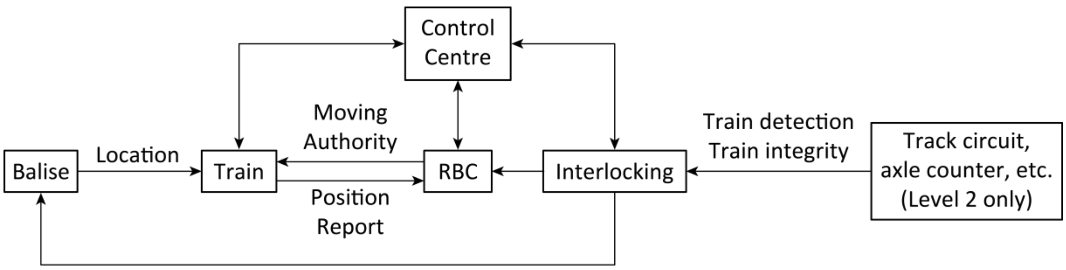


Figure 26: Flow Chart of the Simulation Platform for ETCS Levels 2 and 3

The train calculates time-efficient or energy-efficient control profiles based on MA and train condition, and moves forward given the control profile. The RBC determines the MA, while the control centre calculates the scheduled arrival time at the end of authority. The time between the two adjacent attempts of the RBC to recalculate the MA could be deterministic or random.

Figure 27 illustrates a four station simulator of ETCS Level 2 developed for both a single line (left) and a line with passing loops at intermediate stations (right). The distance between stations is set at 30 km and is an approximation of the Retford-Newark-Grantham-Peterborough section of the East Coast Main Line (100 km). For both simulations a mix of fast trains (blue solid) and slower trains (red dash) are operated. Compared with the case without loops, in the case with loops, the time for all 10 trains to travel from origin to destination decreases from 2 hours and 15 minutes to 2 hours, which is equivalent to a 12% time saving.

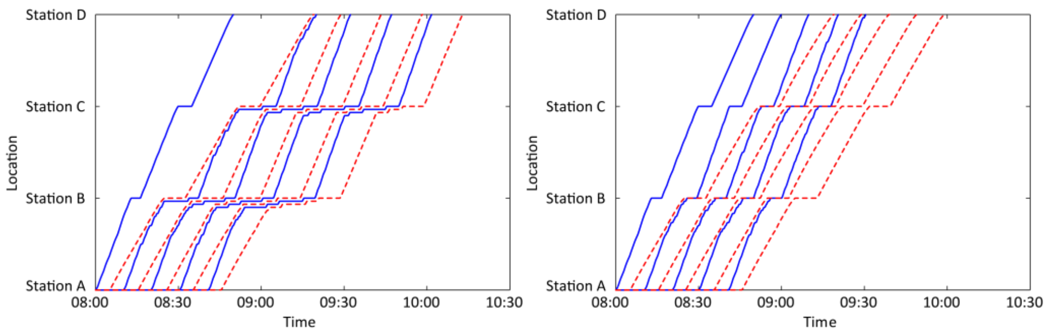


Figure 27: Four Station Simulator for a Single Track without Passing Loops (Left) and a Single Track with Passing Loops (Right).

5.3. Advanced Train Operation and Control Rules

The capacity increase of ERTMS relies on not only the new radio communication system, but also the intelligent traffic management and control strategies for, e.g., train following, train trajectory optimisation, moving authority generation, and rescheduling. Meanwhile, with the real-time and detailed train running information, the MAs and schedules can be adjusted in a more sophisticated way, which introduces challenges for the algorithms of real-time scheduling and train control.

Besides the establishment of simulation platforms for ETCS Levels 2 and 3, we made progresses in the development of advanced train operation and control rules. We proposed new models and algorithms for generating energy-efficient speed profiles for driver advisory systems, in both fixed-block system like ETCS Level 2 and moving-block system like ETCS Level 3 (Ye and Liu, 2016, 2017). We considered multiple trains running simultaneously on a same railway track segment and allowed possible rescheduling at the intermediate stations on their journeys. We also developed a controlled train-following model for ETCS Level 3 to maintain both the optimal speed and the desired following headway in a train platoon, and discussed the impact of the parameters of such train-following model on the stability of the train following, where such impact would determine how a disturbance propagates along the train platoon (Chen et al., 2016).

Good Practice VIII: Energy-Efficient Control for Single and Multiple Trains

To keep the railway lines operating smoothly with short headways, the train control command/advice should be obtained in short time once the train running condition changes, e.g. when the MA is extended or the schedule is adjusted.

The classic train speed profile design problem is described as minimising the energy consumption for a train running freely from one location to the next without obstruction from another train (Albrecht et al., 2016a, 2016b). However, this assumption can be easily violated when the trains are operated in very short headway in ERTMS Levels 2 and 3; in such cases, it is preferable to optimise the speed profiles of all these trains together. Meanwhile, more accurate train speed/location estimation and real-time telecommunication bring a chance to better coordinate the train controls in real time to further reduce energy consumption and improve sustainability of the railway system.

In Milestone 3, we presented a multiphase optimal control formulation (Preston et al., 2016 – see also Ye and Liu, 2016) to simultaneously optimise the speed profiles of multiple trains passing multiple stations in a railway network. The case study showed how the method works in a moving block system (like ETCS Level 3) to both reschedule the trains and provide energy-efficient speed profiles, as an unexpected incident happened in the leading train and reduced the engine power. In Milestone 6 (Liu et al., 2015), another example illustrated how the coordinated control on a leading-following train pair can help guarantee punctuality, stabilise the speed profile and reduce energy consumption (Figure 28).

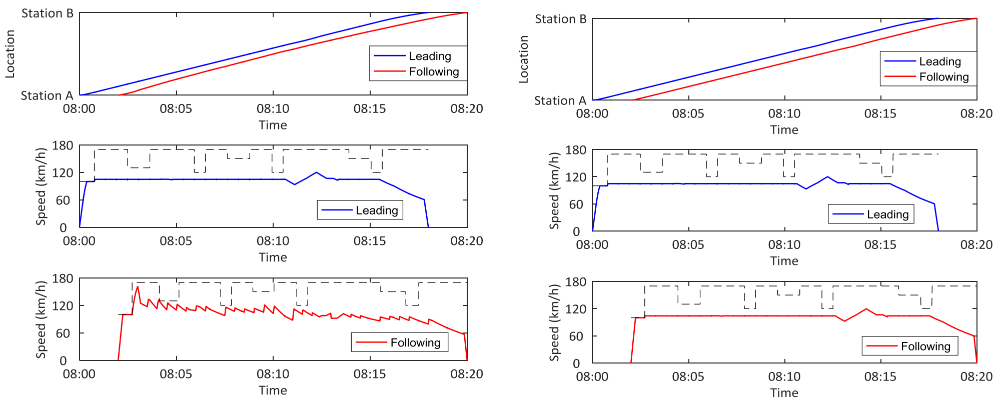


Figure 28: Trajectories and Speed Profiles of Both Trains when Their Speed Profiles are Optimised Separately (Left) and Simultaneously (Right).

To solve such train speed profile design problems, we then proposed new algorithms by converting the optimal control problem into the nonlinear programming problem in a novel way based on closed-form expressions (Ye and Liu, 2017). By dividing the track into subsections where each particular subsection has constant speed limit and constant gradient, the following two methods are used.

Method 1. The train is assumed to sequentially apply 'max traction – cruising – coasting – max braking' on each subsection. The nonlinear functions of maximum tractive force and maximum braking force are approximated by piecewise-linear functions. The decision variables are the time durations of all operations on all subsections.

Method 2. Constant tractive/braking force is applied on each particular subsection to allow general nonlinear forms of maximum tractive force and maximum braking force without using approximation. The decision variables are the constant forces applied. Each subsection can be further subdivided to improve the energy saving.

Merits and advantages of our methods over the existing methods lie in that:

(i) They can solve complex train control problems that cannot be solved by the indirect methods based on Pontryagin's maximum principle (Pontryagin et al., 1962) such as the train control problems with train following and/or multiple stations.

(ii) They allow (more) realistic train and track conditions, so the resultant control/speed profile is feasible and close to the true optimum;

(iii) Better energy saving can be achieved compared with coasting control, which is not unexpected since coasting control is highly dependent on the number of coasting operations;

(iv) Unrealistic fluctuation can be avoided, which is extremely difficult, if not impossible, to avoid in the direct methods such as the pseudospectral method (Ye and Liu, 2016, 2017);

(v) The closed-form expressions of speed, distance and energy can help accelerate the solution process.

Good Practice IX: Advanced Control Based on Train Following Rules

The train-following model describes how one train follows another along a track under ERTMS Level 3. The acceleration and speed of a train is calculated based on its own desired movements and the relative speed and distance to the train ahead. The train-following model consists of three scenarios: (1) a free-flow acceleration scenario, (2) a controlled train-following scenario, and (3) a deceleration scenario.

- The free-flow acceleration scenario applies when the train is not influenced by the train ahead and thus will simply follow its own desired driving cycle. The 'jerk' was considered to restrain the rate of change in acceleration and deceleration in Milestone 6 (Liu et al., 2015), which leads to a discontinuous acceleration profile. Milestone 3 (Preston et al., 2015) further proposed a continuous acceleration model by introducing a smooth function to represent the concave relationship between acceleration and speed.
- The controlled train-following scenario applies when a train's movement is constrained by the train ahead. The train is then controlled to reach some optimal speed and to keep to its desired separation to the train ahead. The optimal velocity is a function of space gap to the train ahead, while the desired distance headway is a function of train speed and the desired time headway. Such a control mechanism is adopted from Chen and Liu (2016) and formulated in Milestone 3.
- The deceleration scenario applies when the train needs to stop. The coasting and braking will be applied in this scenario.

In Milestone 3, we also analysed the stability property of the proposed train-following control mechanism for ETCS Level 3. The findings include:

- The parameter values used in the train-following procedure can have a significant impact on safety and capacity;
- The stable region of a following train is larger with longer desired time gap and shorter response time.

6. Conclusions

DITTO has been developing tools to increase the capacity of rail systems. We have shown how formal methods from computer science can be used to determine theoretical capacity limits (Good Practices I to III). We have shown how analysis of the relationship between capacity utilisation and service reliability, in combination with optimisation techniques, can determine practical capacity limits (Good Practices IV to VI). We have also shown how microsimulation can be used for dynamic scheduling and can be used to control running times (Good Practices VII to IX).

Our Peterborough case study has provided some illustrative results. In the earlier OCCASION project deterministic job shop optimisation was used to insert services into the 2011 timetable (53 services between 0700 and 0900 hours). It was found that 14 additional services could be added in theory - a 26% uplift.

In DITTO stochastic combinatorial job shop optimisation has been used with the 2015 timetable (55 services between 0700 and 0900 – the additional two paths having been facilitated by remodelling in 2014). This work found that an additional 40 services could be inserted, in theory, although with large deteriorations in reliability. However, it was found that wider network constraints reduce this number to 10 (18% uplift). Additional work has found that the benefits of these additional paths (in terms of reduced service interval penalties and reduced overcrowding) do not always exceed the costs (in terms of increased delays) (Vong, 2017). The introduction of moving block signalling would not seem to lead to the claimed capacity enhancements of 40% for existing infrastructure (Wendler, 2007)[6], unless this would lead to substantial reductions in headways.

In future work, we intend to use our simulation tools to examine this issue in more detail by further examining the scope for inserting additional trains on the East Coast Main Line with a particular focus on Newark to Huntingdon. This will involve simulating the additional services identified by the job shop optimisation and examining patterns of disruption. The scope for using TrackULA to reschedule services will be examined. We will also compare our results with those of others.

Good Practice X: Integration of Safety Analysis with BRaSS

The development of railway systems is usually supported by a range of tools, each addressing individual, but overlapping concerns such as, e.g., performance or safety analysis. However, it is a challenge for users to organise work-flows; results are often in different, non-aligning data formats. Furthermore, tools work on different levels of abstraction from macro to microscopic. Thus, tool integration would be beneficial, and

[6] This study found the ETCS level 3 increased capacity compared to level 1 by around 42%. However, it could be argued that for practical purposes the more appropriate comparison is between ETCS level 3 and level 1 with limited supervision. In this case, the capacity uplift is only 28%.

also allow for more playful, experimental prototyping and design. However, such an endeavour needs close collaboration. How to use the various tools and software needs to be demonstrated and documented, especially the relation between the rail system and its model, how this model is encoded into an input format, what commands are used for the analysis, and what the expected result shall look like.

As part of the DITTO goal towards integration, this section reports on lessons learned from the integration of BRaSS – the Birmingham Railway Simulation Suite, see http://www. BRaSSsim.org - and OnTrack. These tools are both being developed as part of the wider FuTRO project. We present first steps towards an approach that bridges the gap that occurs from varying details in data sources through automated transformations. This integration provides a seamless environment for prototyping, concept development, and safety analysis under "one roof".

BRaSS is an easy-to-use railway simulation software package for development, modelling and flow analysis. It is developed by the Birmingham Centre for Railway Research and Education and forms a core part of their research and teaching environment. Modelling the operation of train services with all of the complexities of signalling, interlocking, timetabling and train performance is a challenge. BRaSS simplifies the process, using an intuitive graphical interface to design and develop models of routes, signalling systems and timetables. BRaSS is currently being used to validate signalling design and simulate the Communication Based Train Control system for the new Hefei metro in China. A full account of the features of BRaSS is available at http://www.BRaSSsim.org. BRaSS is written in Java and uses a proprietary format for storing data models. However, it does provide a number of useful data export features that will form the basis of the integration we present.

Overall, both the BRaSS and OnTrack datasets were found to align at a fairly similar level with regards to data abstraction. However several key messages and lessons should be taken away from the integration process:

• Sharing Data: The idea of sharing data between different platforms within rail simulation, verification and design is possible and provides a number of advantages, including reduced time and effort in data entry, guaranteed consistency between models, and a seamless and automatic workflow.

• Granularity of Data: Even models on the same level of abstraction can differ in granularity. This provides a challenge in creating an automated model transformation. Naturally, one has to start with the more detailed model (or model element) and construct a suitable abstraction.

• Modelling assumptions: The model transformation has to moderate between different sets of assumptions used within the different toolsets. Good documentation and understanding of these assumptions is necessary for success.

• Adaptability of Involved Tools: In tool design it would be beneficial to strive for flexibility with regards to modelling assumptions in order to deal with small changes enabling integration. Thanks to careful consideration of the different modelling assumptions made by both the tools, we have successfully provided a first integration of BRaSS and OnTrack.

The results of this integration provide a seamless environment for prototyping, concept development, and safety analysis without the need to re-enter and develop models in different software packages. The usefulness of our approach is illustrated by example, through considering the East Coast main line (ECML). The ECML has been selected as the area of interest for a common case study within the DITTO project, with a particular focus on the stretch of railway running from Huntingdon to Retford. This region of railway has already been encoded by the Birmingham Railway centre within the BRaSS tool. The data captured within BRaSS is based on accurate information from Network Rail. As a first exemplar towards integration, the remainder of the section is based on data that has been exported from the BRaSS toolset and then transferred into OnTrack. The data originally comes from simulation models provided to Birmingham Rail Centre by Network Rail. We have demonstrated the usefulness of our approach by giving integrated simulation and verification results for elements of the ECML model area - see Preston et al., 2016. BRaSS allows users to simulate and design the overall model for the whole East Coast main line, and several sections of this line, namely Barkston South Junction and Werrington Junction have been translated into OnTrack and formally verified against various safety properties. It should be noted that a prerequisite for BRaSS is that the scheme plans to be simulated are safe. If this is not the case, BRaSS simulations will not produce reliable data. Integration with OnTrack allows users to perform this necessary safety check.

Acknowledgements

This research was funded by RSSB and undertaken as project T1071-02**,** as part of the FuTRO (Future Traffic Regulation Optimisation) programme. Preliminary findings were presented to the FuTRO Project Management Board on 18 January 2017 and to the RSSB Advisory Board on 14 February 2017. This paper partly draws on a presentation made to the 7th International Conference on Railway Operation Modelling and Analysis (RailLille2017 – see Preston, et al., 2017).

References

Abril M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P. and Lova, A. (2008). An assessment of railway capacity, *Transportation Research Part E,* 44 774–806.

Albrecht, A., Howlett, P., Pudney, P., Vu, X. and Zhou, P. (2016a). The key principles of optimal train control-Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points. *Transportation Research Part B* 94, 482-508.

Albrecht, A., Howlett, P., Pudney, P., Vu, X. and Zhou, P. (2016b). The key principles of optimal train control-Part 2: Existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques. *Transportation Research Part B* 94, 509-538.

Armstrong, J. and Preston, J. (2015) Node-based Capacity and Congestion Estimator. DITTO Milestone 4. September.

Bektas, T., Chouman, M. and Crainic, T.G. (2010) Lagrangean-based decomposition algorithms for multicommodity network design problems with penalized constraints. *Networks*, 55, (3), 171-180.

Bektas, T., Kovacs, A. and Potts, C. (2015) *Scenario Generation for Train Timetabling and Train Scheduling under Uncertainty.* DITTO Milestone 5. November.

Chen, J., Liu, R., Ngoduy, D. and Shi, Z. (2016). A new multi-anticipative car-following model with consideration of the desired following distance. *Nonlinear Dynamics* 85 (4), 2705-2717.

Department for Transport (DfT) (2007) *Delivering a Sustainable Railway*. Cm7176. July.

Desrosiers, J. and Lübbecke, M. E. (2005) A Primer in Column Generation. In Desaulniers, G.; Desrosiers, J. & Solomon, M. M. (Ed.) *Column Generation,* Springer US, 1-32

ERA/ERTMS/003204. *ERTMS/ETCS Functional Requirements Specification*, Version 5.0. Accessed on 27/09/2016. http://www.era.europa.eu/Document-Register/Pages/Set-3-SystemRequirements-Specification.aspx.

Gibson, S., Cooper, G., Ball, B. (2002), Developments in Transport Policy: The Evolution of Capacity Charges on the UK Rail Network, *Journal of Transport Economics and Policy*, 36 (2), 341-354.

Goetschalckx, M.; Vidal, C. J. & Dogan, K. (2002) Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms, *European Journal of Operational Research,* 143, 1 – 18.

Goodall, R., Dixon, R., Bemment, S., Fujiyama, T., Mei, T. Preston, J., Romanovsky, A. and Roggenbach, M. (2013) *The Future of Train Control Systems.* Technical Paper, Institution of Railway Signal Engineers, London, 13 February.

Hewitt, C., Bishop, P. and Steiger, R. (1973) *A Universal Modular ACTOR Formalism for Artificial Intelligence.* Proceedings of the 3rd International Joint Conference on Artificial Intelligence, 234-243.

Hoare, C.A.R. (1985) *Communicating Sequential Processes.* Prentice Hall International.

Isobe, Y., Moller, F., Nguyen, H.N. and Roggenbach, M. (2012) Safety and line capacity in railways. *Lecture Notes in Computer Science* 7321:54-68.

James, P., Moller, F. and Roggenbach, M. (2015a) *DITTO Tool Integration: Basic Prototypes for Interfacing.* DITTO Milestone 2. September.

James, P., Roggenbach, M. and Moller, F. (2015b) *OnTrack Documentation.* DITTO Milestone 1. June.

James, P., Lawrence, A., Roggenbach, M. and Seisenberger, M. (2015c) Towards Safety Analysis of ERTMS/ETCS Level 2 in Real-Time Maude. *Communications in Computer and Information Science*, 596, 103-120.

James, P., Trumble, M., Treharne, H., Roggenbach, M., and Schneider, S. (2013) OnTrack: An Open Tooling Environment for Railway Verification. *NASA Formal Methods.* 435-440

James, P., Moller, F., Nguyen, H.N, Roggenbach, M., Treharne, H. and Wang, H. (2016) OnTrack: The Railway Verification Toolset -- Extended Abstract. *ISoLA* (2), 294-296.

Kleywegt, A., Shapiro, A., and Homem-de-Mello, T. (2002) The Sample Average Approximation Method for Stochastic Discrete Optimization, *SIAM Journal on Optimization*, 12, 2, 479-502.

Kontaxi, E. and Ricci, S. (2012) Railway Capacity Handbook: A Systematic Approach to Methodologies. *Procedia - Social and Behavioral Sciences*, 48, 2689-2696.

Kovacs, A., Bektas, T. and Potts, C. (2016a) *Stochastic Optimisation.* DITTO Milestone 9, Deliverable 2.3. October.

Kovacs, A., Bektas, T. and Potts, C. (2016b) *Train timetabling and scheduling under uncertainty.* Technical Report, University of Southampton.

Liu, R., Van Vliet, D. and Watling, D. (2006). Microsimulation models incorporating both demand and supply dynamics. *Transportation Research Part A* 40 (2), 125-150.

Liu, S.Q. and Kozan, E. (2009) Scheduling trains as a blocking parallel-machine job shop scheduling problem, *Computers & Operations Research* 36, 2840-2852.

Liu, R., Ye, H. and Whiteing, T. (2015) *Dynamic Simulation for Real-time Operations of ERTMS Level 3.* DITTO Milestone 6, Deliverable 3.1. September.

Liu, R. (2005). The DRACULA dynamic network microsimulation model. In: Kitamura, R. Kuwahara, M. (Eds), *Simulation Approaches in Transportation Analysis: Recent Advances and Challenges*, Springer, pp. 23-56.

Liu, R. (2010). Traffic simulation with DRACULA. In Barcelo, J. (Ed) *Fundamentals of Traffic Simulation*, Springer, pp. 295-322.

Milner, R. (1980) *Calculus of Communicating Systems.* Lecture Notes in Computer Science, Volume 92. Springer-Verlag, Berlin.

Network Rail. (2017a). *Our information and data: Historic delay attribution.* Available from https://www.networkrail.co.uk/who-we-are/transparency-and-ethics/transparency/datasets/#H <Accessed 9 March 2017>

Network Rail, (2017b). *Timetable Planning Rules, Operational Access Statement.* Available from http://archive.nr.co.uk/browse%20documents/Rules%20Of%20The%20Route/Viewable%20copy/roprhome.pdf <Accessed 9 March 2017>

ORR (Office of Rail and Road). (2016) *National Rail Trends.* http://dataportal.orr.gov.uk/ <Accessed 25 November 2016>

Paraskevopoulos, D., Bektas, T. and Potts, C. (2015) *Train Scheduling under Nodal Capacity Constraints.* Technical Report, University of Bath.

Paulson, L. (2012). *Isabelle user-list*. In Moller, F., Nguyen, H.N., Roggenbach, M., Schneider, S. and Treharne, H. (2012) Combining event-based and state-based modelling for railway verification. Computing Science Report CS-12-02. University of Surrey.

Petri, C. A. (1962). *Kommunikation mit Automaten (Communication with Automata).* PhD Thesis, University of Bonn.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V. and Mishchenko, E. F. (1962). *The Mathematical Theory of Optimal Processes.* English translation. Interscience, New York.

Preston, J., Armstrong, J., Kovacs, A., Liu, R., Ye, H., James, P., Moller, F. and Roggenbach, M. (2015). *DITTO Interim Report and Prototype.* Milestone 3, Deliverable 4.1. April

Preston, J., Kovacs, A., James, P. and Ye, H. (2017) Developing Integrated Tools to Optimise Railway Systems: An Overview. *RailLille 2017.* April.

Roberts, L., Smith, M., Moller F. and Roggenbach, M. (2014) *Visualising Timed CSP Train Simulations for Capacity.* Proceedings of TPCG'14: Theory and Practice of Computer Graphics, pp91-93, Leeds, September.

RSSB (2012) *Rail Technical Strategy.* RSSB, London. http://www.rssb.co.uk/library/future%20railway/innovation-in-rail-rail-technical-strategy-2012.pdf <Accessed 20 November 2016>

Smith, M. (2017) *A Structured Approach to Modelling Safety with a Perspective on Capacity in the Railway*. MRes Thesis, Swansea University.

SUBSET-026. *System Requirements Specification, Version 3.6.0.* http://www.era.europa.eu/Document-Register/Pages/Set-3-System-Requirements-Specification.aspx accessed on 27/09/2016.

UIC. (2004). *Leaflet 406: Capacity*, Union Internationale Des Chemins De Fer (UIC), Paris.

UIC. (2013). *Leaflet 406: Capacity (2nd Edition)*, Union Internationale Des Chemins De Fer (UIC), Paris.

Vong, C. H. H. (2017) *Determining the Relationship between Capacity Utilisation and Rail Performance.* Individual Project. BEng Civil Engineering, University of Southampton.

Wendler, E. (2007) Influence of ETCS on Line Capacity. *UIC ERTMS World Conference* 11-13 September. Berne, Switzerland. http://www.uic.org/cdrom/2007/14_ERTMS_conference07/docs/presentations/G4_wendler.pdf <Accessed 25 February 2016>

Ye, H. and Liu, R. (2016). A multiphase optimal control method for multi-train control and scheduling on railway lines. *Transportation Research Part B* 93, 377-393.

Ye, H. and Liu, R., 2017. *Nonlinear programming methods based on closed-form expressions for optimal train control.* Working Paper, Institute for Transport Studies University of Leeds.