Challenging established rules for train control through a fault tolerance approach: Applications at a Classic Railway Junction

Ronghui Liu*, Institute for Transport Studies University of Leeds, Leeds, LS2 9JT
Anthony Whiteing, Institute for Transport Studies University of Leeds, Leeds, LS2 9JT
Andrew Koh, Institute for Transport Studies University of Leeds, Leeds, LS2 9JT

Abstract—Current rules of railway operations employed in the UK can be regarded as being overly conservative. While the safety of passengers is a clearly an overriding paramount consideration, overly conservative rules may represent a drawback in that they reduce capacity. At present, with high and growing demand there is a clear need to increase railway capacity. Whilst construction of new lines or significant augmentation of existing ones offer ways of providing increased capacity, there is a long gestation period and it is extremely expensive at a time of tight governmental budgetary constraints. A viable alternative is to investigate methods of increasing the productivity of the current infrastructure by understanding how the established rules of railway operation can be challenged without compromising the safety of the passengers. To accomplish this stated goal, a fault-tolerance based approach to train controls at nodes (junctions and stations) is proposed. This paper introduces the concept of fault-tolerance in the context of train operations at a classic railway junction, and a dynamic, meso-scopic simulation model developed to represent the new control rules. It discusses some initial results from the application of this approach on capacity gains.

Keywords—Fault Tolerance, Meso-scopic Simulation, Junction Control, Capacity Improvements

I. INTRODUCTION

The operation of a rail network is safeguarded through the use of train control and protection systems, traditionally in the forms of ‘block signalling’ and ‘interlocking’ ([1]-[3]). Current rules of railway operations employed in the UK can be regarded as being overly conservative. While the safety of passengers is a clearly an overriding paramount consideration, overly conservative rules may represent a drawback in that they reduce capacity. At present, with high and growing demand there is a clear need to increase railway capacity. Whilst construction of new lines or significant augmentation of existing ones offer ways of providing increased capacity, there is a long gestation period and it is extremely expensive at a time of tight governmental budgetary constraints.

A viable alternative is to investigate methods of increasing the productivity of the current infrastructure by understanding how the established rules of railway operation can be challenged without compromising the safety of the passengers utilising the system. To accomplish this stated goal, a fault-tolerance based approach to train controls at nodes (junctions and stations) is proposed, as part of an ongoing research project.

This paper introduces the concept of Fault-Tolerance in the context of train operations at a classic railway junction, and a dynamic, meso-scopic simulation model developed to represent the new control rules. It discusses some initial results from the application of this approach on capacity gains.

II. PROBLEM DESCRIPTION AND THE CONCEPT OF FAULT TOLERANCE

Taking a classic railway junction as our example, Fig. 1 illustrates (with reference to British 3-aspect colour-light signalling systems) how conflicts between trains are managed. Four categories of train are relevant as indicated in Fig. 1. A1 trains travel from left to right on the straight route, whilst A2 trains diverge from this route to the right at the junction. B trains take the straight route, but from right to left, whilst C trains merge into this route at the junction.

Fig 1. A classic railway junction. The signals controlling trains are represented by s1 to s12, while the trains (A1, A2, B and C) are depicted as rectangles with their direction of travel indicated. Also illustrated are the two points (p1 and p2) that are used to set the routes for the relevant trains.

In this particular instance, the basic principle of the Fault Tolerance (FT) approach is that when we believe that train A2 will be slowed down unnecessarily, because train B that

R. Liu: r.liu@its.leeds.ac.uk
A. Whiteing: a.e.whiteing@its.leeds.ac.uk
A. Koh: a.koh@its.leeds.ac.uk

Supported by Engineering and Physical Sciences Research Council of the UK and the Rail Safety and Standards Board under Grant EP/B010637/1: “Challenging Established Rules for train control though a Fault Tolerance Approach”
is preventing signals from changing will actually have moved out of the way in time for A2 trains to proceed, then elements of the current rules can be relaxed.

To investigate the scope for increased capacity at this junction through the implementation of revised control rules, we consider three scenarios involving different approaches to the application of FT principles. The essential differences between these approaches relate to the deceleration profiles of A2 trains as they are driven towards a red aspect at signal s3.

- In the first scenario, the current conservative rules are maintained. Whilst the ‘rulebook’ [4] allows drivers to maintain normal speed until the yellow aspect is reached, in practice railway companies adopt more defensive driving styles and hence drivers frequently decelerate on first sight of a yellow aspect. This is the scenario we adopt.

- The second scenario represents the ‘ultimate’ or ‘extreme’ interpretation of FT principles. In this case, the driver (having been advised that FT rules are in force for his A2 train) does not decelerate at all, even though signal s3 is showing red. If the conflict is not resolved, the train proceeds straight ahead on an incorrect route.

- The final scenario is a ‘moderate’ interpretation of FT (and possibly more reasonable and more likely to be acceptable to industry). In this interpretation, the driver maintains normal running speed until the yellow aspect is passed, and then decelerates more rapidly than at present in order to stop at the red aspect.

Fig. 2 illustrates these three speed profiles.

---

### III. SIMULATION MODELS OF TRAIN OPERATIONS

Simulation models have been applied to a variety of situations in rail planning and operations [5]. In the UK, the then British Rail had been using a computer simulation package, GATTS, since the 1970s to aid planning changes to infrastructure, changes to timetables and for the design of train regulation strategies. In other applications such models have been used to estimate the effect of exogenous random delays on track occupation and consecutive delays of hindered trains [6].

Following Siefer [7], simulation models can be categorised as follows:

- **Microscopic vs Macroscopic**

  Microscopic models describe train movements in terms of the train performance and track conditions and aim to reproduce the actual operation of the rail system over a user defined time period [5]. Examples of models in this category include OpenTrack [8], RailSys [9],[10], SimMETRO [11] and VISIONS [12]. Generally speaking microscopic models take as input the infrastructure parameters, signalling systems, rolling stock parameters and the timetable (all in extensive detail) into account as it replicates performance over a given time period.

  In contrast to the microscopic models, macroscopic models do not model individual unit (e.g. train) operations nor do they consider how trains are impacted by other trains [8]. An example of a macroscopic model is NEMO [13]. Input data such as infrastructure is modelled with less detail resulting in improvements in reduced computational run times [14].

- **Synchronous vs Asynchronous**

  Synchronous models simulate all occurrences of events within the network concurrently in time steps while asynchronous models path the rolling stock in a user specified order [7] often via a series of inter-related model runs. VISIONS, RailSys and OpenTrack and SIMMetro are examples of synchronous simulation models. As synchronous models simulate all the trains operating in the modelled network at the same time, they provide a good way to simulate realistic operating conditions. For example, they can be used to determine the impact of delays and their propagation through the network [15].

  On the other hand, in asynchronous simulation models e.g. STRESI [16], the trains are allocated priorities depending on the class they belong to (e.g. long distance intercity trains may have higher priority local stopping ones). The allocation of train paths, in the creation of timetables, is then dependent on these priorities. In the first run the highest priority trains are modelled first. Their allocated paths (i.e. schedule) are then “locked” and the second class are modelled. The operation of this latter set is not allowed to impact the first but they may be impacted by the former. This process is repeated until all predefined priority classes are exhausted with the current class impacted by all higher proceeding priority classes. It is suggested that these models are often used for timetable construction since they can replicate ideal conditions [8].

- **Deterministic vs Stochastic**

  Deterministic models estimate the arrival, departure and running times according to the schedule [7]. Their primary use is for the preliminary design of a timetable. On the other hand, stochastic models utilise statistical distributions of
arrival, departure and running times. The simulation packages mentioned above (such as RailSys and OpenTrack) are both equally capable of performing in deterministic mode as well as stochastic mode. They are endowed with facilities to enable users to determine the robustness of timetables in the face of disruptions (e.g. rail vehicle breakdown on track) and incidents (e.g. inclement weather) [17].

IV. A MESO-SCOPIC SIMULATION MODEL FOR FT CONTROL OF TRAIN OPERATIONS

Most of the abovementioned commercially available railway simulation packages were originally developed with the intention of modelling railway operations under the existing rules. However our research aims primarily to assess the outcomes associated with the relaxation of existing rules through the use of FT principles.

To accomplish this stated goal, we have adapted an existing road-based simulation model [18-19], to the railway domain and new modelling features were developed to represent railway blocks, overlaps, points and the three-aspect signal signalling system described above. The model takes into account the line speed limits, and the normal operating speed of relevant rolling stock and its acceleration and deceleration rates.

The simulation model we have developed is a time based synchronous mesoscopic simulation model. Unlike OpenTrack[8] and RailSys [9] for example, we do not model the detailed traction of the individual trains but instead apply parameters sourced from the literature in consultation with the railway industry. Unlike STRESI [16] for example our simulation methodology employs time-based synchronous simulation whereby trains are loaded onto the network based on the schedule.

The model features a directed graph based representation of networks, where the signals and points are represented as nodes of the graph, and the blocks of tracks are represented as directed links. We use adjacency list to represent the connections of nodes.

The rolling stock is individually represented in the simulation, with their scheduled time of arrival to the junction, scheduled route and type of rolling stock. To model the movement of the rolling stock through the junction, we do not solve differential equations to calculate the acceleration of a train based on its load, running, grade and curvature resistance. Rather, we represent the motion of each individual train by calculating the equation of motion on its acceleration, speed and position at every time interval, with a given profile of its acceleration and deceleration.

The model represents the signal control at the junction based on the British (three-aspect colour-light) situation, though the principles are transferable to other regimes. Signals that are critical to managing junction and other conflicts are held at danger (a red aspect) unless a train is scheduled and can proceed without conflict. Signals preceding red aspects sequentially show yellow and green aspects. The model can be easily extended to handle four aspect (employing an additional “double-yellow” aspect) signalling systems [3] though this was not considered in the numerical simulations reported in the next section.

V. SIMULATION ANALYSIS AND DISCUSSION

Through simulating train operation under the three different rules established in the previous section of the paper, we can estimate the time taken to approach the junction under existing vis-à-vis both variants of FT rules to estimate the potential reductions in delays.

Using the following parameters as shown in Table I, we carried out a simulation under deterministic conditions (i.e. there are no other delays on the network) and all trains are operating according to the specified timetable. We assume the application of FT to trains of types A2 and B and assume other types (namely A1 and C) operate under existing rules.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Assumed Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running speed on green (V1)</td>
<td>100 km/h</td>
</tr>
<tr>
<td>Reduced running speed (V2)</td>
<td>40 km/h</td>
</tr>
<tr>
<td>Acceleration</td>
<td>0.4(m/s²)</td>
</tr>
<tr>
<td>Deceleration</td>
<td>-0.6(m/s²)</td>
</tr>
<tr>
<td>Block Length</td>
<td>1000 (metres)</td>
</tr>
</tbody>
</table>

The theoretical minimum time a train takes to travel past the junction (3 block lengths) can be computed to provide a benchmark to inform us of the potential time savings any application of FT rules can provide. At the same time, the simulation model reports the time taken for the A2 train to clear the junction and travel past it under current rules and both variants of FT (‘moderate’ and ‘ultimate’). A summary of these results is shown in Table II.

<table>
<thead>
<tr>
<th>Time (seconds)</th>
<th>Additional Time vis-à-vis Theoretical Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical Minimum Travel Time</td>
<td>126</td>
</tr>
<tr>
<td>Travel Time under Current Rules</td>
<td>185</td>
</tr>
<tr>
<td>Travel Time under Moderate FT</td>
<td>179</td>
</tr>
<tr>
<td>Travel Time under Ultimate FT</td>
<td>146</td>
</tr>
</tbody>
</table>

The first stage of this analysis is described with reference to Fig. 3. The starting point is to schedule a B train through the junction. This train is assumed to pass, unrestricted, through the junction at its normal running speed. The issue in
question relates to the earliest time we can schedule an A₂ train to pass through the junction after the B train has passed, but without the A₂ train having to decelerate.

Under current rules, the earliest that we can schedule an A₂ train is shown by the solid line, because if it arrived any earlier it would have passed the green aspect of signal s1 and its driver would respond to the yellow aspect of signal s2 and start to decelerate. In contrast, application of ‘moderate FT’ rules would allow the driver to maintain normal operating speed until signal s2 is reached. This means that the A₂ train could be scheduled earlier, but would still pass through the junction without deceleration. The ‘ultimate FT’ case allows the driver to maintain normal operating speed right up to signal s3, which will change from red to green precisely when the A₂ train reaches it, in which case the A₂ train could be scheduled earlier still.

Using the parameter settings given in Table I, we simulated the space-time trajectories of several different schedules for A₂ trains under these three control rules (existing, ‘moderate’ FT and ‘ultimate’ FT), and show in Figs. 4-6 the respective trajectories. It can be seen that, under the current rule and for a given trajectory of the B train, the earliest departure for an A₂ train would be at time 108 sec without the train having to slow down. Any earlier departures would incur some delays (due to the lost time incurred in slowing down upon sighting the yellow aspect). However, under the moderate FT rule, A₂ train could be scheduled to depart at time 72sec without incurring further delay, whilst under the ultimate FT rule we can bring forward that schedule even earlier to time 36sec. Fig. 7 summarizes the amount of delays to A₂ trains depending on rules in force.

It can be seen that, under the current rule and for a given trajectory of the B train, the earliest departure for an A₂ train would be at time 108 sec without the train having to slow down. Any earlier departures would incur some delays (due to the lost time incurred in slowing down upon sighting the yellow aspect). However, under the moderate FT rule, A₂ train could be scheduled to depart at time 72sec without incurring further delay, whilst under the ultimate FT rule we
can bring forward that schedule even earlier to time 36sec. Fig. 5 summarizes the amount of delays to A2 trains depending on rules in force.

![Delay to earlier departures](image)

Fig 3. Delays to earlier departures of A2 train under the different control rules.

VI. SUMMARY AND CONCLUSIONS

This paper has discussed a single example of the potential capacity gains by relaxing the rules of railway operation in the single situation of a classic railway junction at which a double-track route diverges into two alternative double-track routes. The meso-scope simulation model developed is demonstrated to offer the flexibility to test new railway operation control principles that challenge the existing rules with the aim of increasing the productivity of the infrastructure.

Whilst the analysis set out above clearly identifies potential capacity improvement through the adoption of FT principles, there are significant hurdles to overcome before such principles (and most particularly the ‘ultimate FT’ principles) can be adopted. One obvious issue relates to the need to change the setting of points as trains approach, although ongoing research and development work into more resilient point control and locking mechanisms may address this issue.

Perhaps a more realistic view is that ‘moderate FT’ rules are much more acceptable, as they retain a train’s ability to stop if required (albeit in a manner possibly less comfortable for the passengers on board). In such situations, ‘moderate FT’ should not be regarded as a totally radical change, because it is actually an amalgam of three possibly much more acceptable changes – firstly that deceleration rates are significant greater than those typically used under present practices, secondly that speed instructions are changed from standard to ‘moving block’ systems such as the European Rail Traffic Management System (ERTMS) which entails removal of physical signals in any case, and this will also facilitate the third change. As our earlier analysis shows, implementation of ‘moderate FT’ rather than ‘ultimate FT’ still retains around half of the full capacity benefits.

We note that there are other important aspects of railway operation in which similar examples of much more dynamic real-time decision-making may lead to worthwhile capacity gains. These might include for example a dynamic version of the train platforming problem [20] or the problem of specifying route choices through complex junctions with multiple routes.

ACKNOWLEDGEMENT

The work reported is part of a collaborative research project between Control and Systems Engineering Research Centre (CASE)-University of Salford led by Professor TX Mei and Institute for Transport Studies (ITS)-University of Leeds led by Dr Ronghui Liu. The project is funded jointly by the Engineering and Physical Sciences Research Council of the UK and the Rail Safety and Standards Board under Grant EP/I010637/1: “Challenging Established Rules for train control though a Fault Tolerance Approach”. We would like to thank our collaboration partners at University of Salford as well as members of the project’s Industrial Advisory Group for their insightful comments when an initial version of the results were presented.

REFERENCES


