Addressing nodal constraints on railway capacity

John Armstrong¹, John Preston¹, Chris Potts¹, Tolga Bektas¹, Dimitris Paraskevopoulos²,
¹University of Southampton, UK
²University of Bath, UK

Abstract

As demand for passenger and freight transport on Britain’s railways increases, providing additional capacity and making the best use of the existing infrastructure are priorities for the industry. Since the stations and junctions forming the nodes of the railway network tend to form the constraints on route and network capacity, improved understanding of their operation and capacity characteristics is particularly important.

This paper describes research undertaken to improve the understanding of nodal capacity and capacity utilisation, and to route and schedule trains more efficiently through nodes, thus improving service quality and/or releasing capacity for additional train services.

Keywords

Railways, Capacity, Capacity Utilisation, Nodes, Stations, Junctions, Trains, Re-scheduling, Re-routeing

1 INTRODUCTION

Railway capacity is an increasingly scarce and valuable commodity. The UK Department for Transport (DfT) summarised the situation in its 2011 High-Speed Rail Consultation (1) as follows:

Britain’s rail network is seeing a continuing pattern of steeply rising demand. As a result, rail capacity is under increasing strain and services are growing more crowded.
Capacity constraints at nodes (i.e. junctions and stations) in the railway network are a particular concern, since they tend to form the bottlenecks which limit the capacity of the overall network. As observed by the Engineering and Physical Sciences Research Council (EPSRC), the Rail Safety and Standards Board (RSSB) and DfT in their 2010 request for proposals for research into overcoming these nodal capacity constraints (2),

nodes in rail networks (stations and junctions) are vital elements of a railway infrastructure, but they are frequently also a major constraint on capacity.

This paper describes work undertaken for the OCCASION (Optimising Capacity Constraints: A Simulation Integrated with Optimisation of Nodes) project, one of five research projects funded and undertaken in response to the above request, to develop means of increasing nodal capacity without major investments in infrastructure enhancements. It also describes subsequent and ongoing work in this area, undertaken partly in collaboration with European railway Infrastructure Managers (IMs).

Following this introduction, measures of and issues relating to railway capacity and capacity utilisation are described and discussed, and the extension of an existing capacity utilisation measure from network links to nodes is described. Next, approaches to the re-routeing and re-scheduling of trains at individual network nodes to reduce capacity utilisation, and thus to increase potential capacity and/or service reliability, are described. Methods for extending the approach from individual nodes to a wider network are then considered. Finally, some conclusions are drawn and subsequent and ongoing work is described, followed by a list of cited references.

2 RAILWAY CAPACITY AND CAPACITY UTILISATION

Railway capacity is a somewhat abstract and elusive measure: as noted by UIC (Union International des Chemins de Fer, or International Union of Railways), it is difficult, if not impossible, to define a specific capacity value for any given railway track, route or network: “capacity as such does not exist [and] railway infrastructure capacity depends on the way it is utilised” (3). Achievable capacity therefore depends not only upon the characteristics of the infrastructure, including the signalling system used, but also upon the performance characteristics and mix of trains using the route, the timetable in operation, and also the target levels of reliability and punctuality to be achieved by the timetabled trains.

Two important railway capacity-related objectives are: (i) making the best use of the capacity that is potentially available, i.e. aiming to maximise the number of trains that can be operated, subject to other constraints (service mix, stopping patterns, etc.), and (ii) the related issue of maintaining quality (i.e. reliability and punctuality) of service by not operating so many trains that service quality is reduced below an acceptable level. Capacity utilisation (or capacity consumption) is a measure of the extent to which the available capacity under a given set of circumstances is being used or consumed, and thus
of the potential to operate additional services. Capacity utilisation assessments inform both of the above objectives, in that (i) the greater the number of trains being run for a given level of capacity utilisation, the greater the level of capacity being attained; and (ii) the greater the level of capacity utilisation, the less likely the service provided is to be robust and reliable, with particularly severe levels of service quality deterioration being seen at capacity utilisation levels in excess of 60%-70% over extended periods, and in excess of 75%-85% over shorter, peak periods (3).

As implied in the preceding paragraph, a wide range of capacity values may thus be provided for any given level of capacity utilisation; conversely, different levels of capacity utilisation may be achieved for a given level of capacity provision, depending upon how that capacity is provided. The situation is summarised as follows by Armstrong et al. (4):

where capacity is scarce, the objective will typically be to maximise the capacity for the maximum level of capacity utilisation that is consistent with the provision of a stable, reliable quality of service. In order to achieve this, it is necessary to have measures of both provided capacity and capacity utilisation for all parts of the route or network in question, i.e. at nodes (junctions and stations) as well as links (sections of plain line between nodes). Well-defined capacity utilisation measures are currently available for links, but not generally for nodes.

The first element of the development work undertaken for the OCCASION project was therefore the extension of capacity utilisation measurement techniques to railway network nodes.

2.1 Extending the CUI approach

Capacity utilisation can be taken into consideration and/or measured in various ways, from ensuring that timetable development is conducted in accordance with the relevant timetable planning rules, to the use of detailed operational simulation to assess the performance of a proposed timetable and infrastructure combination. Two of the best-known analytical approaches to the measurement of capacity utilisation are the international UIC 406 approach (3) and, in Britain, the Capacity Utilisation Index (CUI) method, as described by Gibson et al. (5). The two methods are similar in approach, both employing the technique of ‘timetable compression’, but vary in the level of detail at which they are applied, since the UIC 406 method is applied at the signal block level, whereas the CUI method is applied to longer route sections, and does not consider individual block sections. The CUI method is the more relevant to the British operating context, since it is based upon the timetable and capacity planning process employed, as set out in the Timetable Planning Rules (formerly known as the Rules of the Plan) established and maintained by Network Rail (6), the owner and operator of most of Britain’s heavy rail infrastructure. The CUI approach is illustrated below in Figures 2, 3, 5 and 6.

A limitation of the CUI and the original UIC 406 methods is that their application beyond simple track sections between nodes is not straightforward: it was recommended by UIC
that, for the application of the UIC 406 compression mechanism, “ideally, the line section used for compression should be reduced to the line section between two neighbouring stations (without overtaking or crossing possibilities)”, while the CUI methodology is also currently limited to such ‘plain line’ sections between nodes. While the UIC 406 method could, in principle, be applied to junctions and stations, the recommended levels of maximum utilisation for plain line sections were not valid for platform tracks and junctions. This limitation was recognised by UIC, and leaflet 406 was updated in 2013 (7) to include coverage of nodes. The CUI method remains unchanged, but it is recognised by industry practitioners that, for example, it is not particularly useful to have a methodology which indicates that the approaches to a station are working at an acceptable level of capacity utilisation, when the methodology provides no means of assessing the levels of capacity utilisation within the station itself, where the capacity constraints are likely to be located.

There is thus a need and an opportunity to extend the CUI methodology in a similar direction to that undertaken by UIC for Leaflet 406. Although the underlying issues may be complex, there is potential for significant benefits arising from the development of an outwardly simple, generic methodology and tool for the assessment of nodal capacity utilisation. Preliminary work conducted on this task is described by Armstrong et al. (4), and summarised below. Similar work, based on the UIC 406 approach, has been undertaken by others: while Lindner (8) expresses scepticism about the application of timetable compression to the assessment of railway nodes, the approach has been applied successfully by Landex (9) and by Libardo et al. (10).

As can be seen in Figures 2, 3, 5 and 6, the timetable compression approach entails the ‘squeezing together’, or compressing, of train paths, while maintaining their planned sequence, so that successive trains are separated by the minimum headway (or, in the cases shown, the minimum junction margin or platform reoccupation time) applicable to the route section in question. The CUI is then calculated as the ratio of the compressed time to the time interval under consideration (typically one or more hours), and usually expressed as a percentage. Capacity utilisation measures on railways are thus similar to the Ratio of Flow to Capacity (RFC) measure applied to highway traffic (11). As noted above, the UIC 406 and CUI methods vary chiefly in the level of detail at which they are applied.

The initial extension of the CUI approach from links to nodes was based on assessments of (i) Pirbright Junction, between Woking and Basingstoke on the South West Main Line (SWML) from London to the south-west of England, where the line from Alton and Aldershot joins the SWML (see Figure 1), and (ii) Southampton Airport (Parkway) station, a busy commuter station located on the SWML between Southampton Central and Eastleigh stations (see Figure 4), served by a mixture of local services and longer-distance trains operating between London and the south coast, and across the wider ‘cross country’ network.
Figure 1: Pirbright Junction

For both nodes, the ‘Up’ (i.e. London-bound) direction was assessed, focussing on train movements through the labelled Up Junction and Platform 1, respectively. The Timetable Planning Rules indicated a minimum junction margin of three minutes for Pirbright Junction; the weekday morning peak (08:00 – 09:00) timetable graph for the Up Slow line, including trains joining from the Up Alton line, is shown in Figure 2, and the compressed graph (for the junction only, ignoring headway constraints on the adjacent sections) is shown in Figure 3.

[Insert Figure 2]

[Insert Figure 3]

By reducing the interval between successive movements through Pirbright Junction to three minutes, the time between the first and last trains is 18 minutes (including the margin after the last train), representing a CUI of $18/60 = 30\%$ for the hour between 08:00 and 09:00.
Figure 4: Southampton Airport (Parkway) station

The Timetable Planning Rules did not specify a minimum platform re-occupation time for Southampton Airport (Parkway), but did specify minimum fast and slow headway values for different sections of the SWML: for the section of the route in question (Eastleigh - Redbridge), the slow value of 2½ minutes was applicable. The weekday timetable graph for the Up Main line between 08:00 and 09:00 is shown in Figure 5, with Southampton Airport (Parkway) highlighted, and the compressed graph (again for the station only, ignoring headway constraints on the adjacent sections) is shown in Figure 6.

Excluding the headway before the first train, and ignoring any effects on the adjacent line sections (note the reduction of the headway between the third and fourth trains at St Denys), it can be seen that, by reducing the interval between the departure and arrival of successive services at Southampton Airport (Parkway) station to 2½ minutes, the time required between the arrival of the first and the departure of the last trains is 20 minutes (= 17½ minutes + 2½ minutes, including the headway after the last train), representing a CUI of 20/60 = 33.3%.

This initial work was then ‘scaled up’ for application to the case study area chosen for the OCCASION project: the East Coast Main Line (ECML) between Huntingdon and Grantham (59 and 106 miles respectively north of London King’s Cross). A detailed node- and link-based network model was developed for the area, and the detailed routeings of trains through the modelled area were determined from electronic Common Interface Format (CIF) timetable data for the route, using a Perl script developed for this purpose. Having obtained the route, and thus the timings, of each train through the individual nodes (point ends, crossings and platforms) and intermediate links comprising the study area, the timings for each element of the network were aggregated, sorted and compressed to produce individual CUIs, from which overall minimum, maximum and average values were calculated, again using the script referred to above. The CUI calculation process was repeated for the optimised timetable with additional trains (as
described in the following section of this paper), and the results for both timetables are summarised in Table 1 for the weekday period between 07:00 and 09:00.

The 0% values occur on nodes and links that are unused during normal timetabled operation; it can be seen that the maximum and average values for both links and nodes increase when additional trains are included in the optimised timetable.

<table>
<thead>
<tr>
<th>Timetable</th>
<th>Network Element</th>
<th>CUIs</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 2011</td>
<td>Nodes</td>
<td>0%</td>
<td>51%</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Links</td>
<td>0%</td>
<td>64%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>Optimised</td>
<td>Nodes</td>
<td>0%</td>
<td>73%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Links</td>
<td>0%</td>
<td>81%</td>
<td>26%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: CUI summary results for existing and optimised ECML timetables

### 2.2 Subsequent and ongoing development work

Except in the cases of the simplest junctions and stations, overall average values are unlikely to provide a representative measure of capacity utilisation; maximum values may be more helpful (see Table 2 and the accompanying text, below), but are again unlikely to be sufficiently representative. Work is therefore ongoing to develop the methodology further. This work is being undertaken as part of the RSSB-sponsored DITTO Rail Systems (Developing Integrated Tools To Optimise Rail Systems) research project and the UIC-sponsored ACCVA (Assessment of Capacity Calculation Values) project, whose aim is to enhance the UIC 406 method’s application to the calculation of capacity utilisation values for nodes and to identify guideline maximum values, similar to those already available for links.

The ACCVA project is being undertaken in collaboration with IMs from the Austrian, German, French, Italian, Czech, Slovak and Hungarian national railway organisations, and the nodal updates to UIC 406 are being tested against existing methods employed by the German and Czech IMs, and compared with the results obtained from OCCASION and DITTO Rail Systems. To facilitate the comparison, the ECML study area is being used for the comparative exercise.

The ongoing research also draws upon experience gained and lessons learned from the application of the CUI calculation tools developed for the OCCASION project to the 2013 recalibration of Network Rail’s Capacity Charge by Arup and the University of Southampton (12,13). For the purposes of the recalibration, hourly and three-hourly CUIs were calculated across the national network by assigning weekday, Saturday and Sunday timetables, totalling in excess of 60,000 trains, to a ~6,800-link representation of the national network, and calculating the resulting levels of capacity utilisation.
Having established a measure of capacity utilisation for network nodes, the second stage of the OCCASION project entailed the development of methods for the re-routeing and re-sequencing of trains through the junctions and stations in the case study area, with the objective of minimising nodal capacity utilisation, and thus maximising reliability and potentially releasing capacity for additional services. This element of the project made use of well-established job shop scheduling techniques, whereby trains are treated as ‘jobs’ to be processed by ‘machines’ comprising the links and nodes of the railway network, train movements are treated as the ‘operations’ performed on jobs, and the sectional running times of trains through track sections are treated as the corresponding jobs’ processing times. The relationship between train scheduling and production scheduling has been known for some time, since the pioneering work conducted by Szpigel (14); however, its use for solving practical problems is comparatively recent, including, for example, its application by Liu and Kozan (15), in the form of a blocking parallel machine job shop scheduling problem, to the Train Timetabling Problem (TTP) on Australian railways.

‘Time windows’ were used in conjunction with the job shop model, to reflect the varying degrees of ‘flex’ available for application to different train services when adjusting the timetable. ‘Hard’ time windows were applied to long-distance passenger services, preventing their alteration; ‘regular’ time windows were applied to shorter-distance/local passenger services, enabling them to be modified, albeit at the expense of a penalty being applied to the objective function; and soft or no time windows were applied to freight services and ‘empty coaching stock’ moves, enabling them to be adjusted significantly, and reflecting their relative flexibility. When adding trains to the timetable, replicates of existing services (in terms of running times and stopping patterns) were used, both to limit the ‘search space’ and to reflect existing commercial service patterns – while a more comprehensive approach to re-scheduling would almost certainly reduce capacity utilisation further, and enable the introduction of a greater number of additional services, this would entail a more-or-less complete recast of the existing timetable, and would require significant liaison with operators and other stakeholders to ensure a commercially attractive and realistic outcome.

A construction heuristic was used to generate initial solutions, reducing/eliminating scheduled waiting time (i.e. ‘pathing’ and other allowances included in the timetable), and thus reducing capacity utilisation. Since this instance of the job shop scheduling problem is NP-hard, and exact solutions are not available, a two-phase metaheuristic was developed to provide a solution to the next two steps: in the first phase, a serial insertion heuristic was used to add replicate trains to the existing timetable, starting with the ‘hard time window services’, which have least flexibility, and then continuing to services that can be scheduled more flexibly (within each time window category, highest-priority services are given ‘scheduling precedence’), thus providing a set of initial solutions. In the second phase, a Tabu Search (16) algorithm was used to improve the initial solutions, using ‘relocate and exchange’ neighbourhood operators in conjunction with a perturbation strategy to enable the optimisation process to escape from local optima.
Finally, a CPLEX solver was used to minimise variations in service frequency, thus enhancing the regularity of the timetable pattern. The approach was applied to the two-hour weekday morning peak period and model area described above. The resulting model included 53 trains and 149 infrastructure component ‘machines’ to process the ‘jobs’ representing the trains. The overall solution framework is shown in Figure 7, and the results of the process are summarised in Table 2.

<table>
<thead>
<tr>
<th>Period</th>
<th>Initial Services</th>
<th>Total Services</th>
<th>Additional Services</th>
<th>% Increase in Services</th>
<th>CUI-type measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>07:00 – 09:00</td>
<td>53</td>
<td>67</td>
<td>14</td>
<td>26</td>
<td>79%</td>
</tr>
</tbody>
</table>

Table 2: ECML Re-Routeing and Re-Scheduling Results

The 24.5 minutes of scheduled waiting time were removed from the timetable by means of the re-routeing and re-scheduling process, and, as can be seen from Table 2, 14 additional services were accommodated, a 26% increase in service numbers, indicating a ‘CUI-equivalent’ measure for the existing timetable of 53/67 = 79%. This value is significantly higher than the average values shown in Table 1, but similar to the maximum values shown for the optimised timetable. Although the values are not strictly comparable, the similarity with the maximum values for the optimised timetable seems intuitively correct, in that the job shop analysis takes account of multiple constraints across the modelled network, and its results are therefore likely to reflect maximum, rather than average, CUI values on the network.

It should be noted that none of the 14 additional trains accommodated in the optimised timetable is a Main Line service to or from London: all are regional/cross-country passenger services operating on regional lines. However, this is a realistic outcome, reflecting capacity constraints between Peterborough and London, and high levels of capacity utilisation in the morning peak period. While there is unlikely to be much passenger demand for the additional identified train paths, they are of potential use for freight traffic.

4 CONSIDERATION OF ADJACENT NODES AND THE WIDER NETWORK

While optimising the routeing and scheduling of trains through a single junction or station may provide local performance and/or capacity benefits, it does not guarantee an overall optimal outcome, since an optimal solution for one node may not be feasible for one or more adjacent nodes, which may be subject to additional traffic demands or capacity constraints. The third and final major element of the OCCASION project was
therefore to extend the application of the techniques already developed from single to multiple, adjacent nodes, to ensure that maximising the throughput of one node does not cause problems elsewhere in the network. For the consideration of multiple network nodes, a Multi-commodity Capacitated Network Design Problem (MCNDP) approach (17) was used. The MCNDP can be defined on a graph of nodes and links (18): the links are typically directed, and represent the arcs of the network. The nodes represent the origins and/or destinations of a particular transport demand of single or multiple commodities (e.g. freight, passengers), and are connected by the intermediate transport links, each of which has a fixed cost, length and capacity. The fixed cost is considered when a particular link is used, while there are variable costs, coupled with each used link, related to the commodity flow. The capacity depicts the volume of the traffic that a link can accommodate. The objective is to select a subset of links of the network to satisfy all the transportation requests of the nodes and to determine how much and which commodity will travel via each link such that the total fixed and variable costs are minimised.

To model the MCNDP, a two-dimensional time-space representation was used, of the type shown in Figures 8 and 9, with the horizontal axis depicting the stations and/or junctions, and the vertical axis the time horizon of ‘time buckets’. Each node of the network thus reflects the origin and/or destination at each time bucket of the time horizon. The capacity of each node is considered to be one, i.e. only one train can arrive at the station/junction and only one can depart at a specific time bucket (it can be seen in Figure 8 that there is congestion at nodes B3, C4 and D5, with associated waiting times for the affected trains 2, 1 and 3 respectively). Solving the MCNDP for such a network is equivalent to selecting the appropriate flows to satisfy the passenger and freight demands, such that the total cost is minimised. In the context of the OCCASION project, the objective was to minimise the capacity utilisation of each station, which in turn minimises the potential for congestion and provides an opportunity to add capacity by introducing additional services (as indicated by trains X and Y in Figure 9).

To solve the model, an efficient metaheuristic algorithm was developed, which produces high-quality solutions for large-scale problem instances; a Scatter Search approach was used to generate initial solutions, which were then improved by means of an Iterated Local Search. The algorithm’s performance was tested on a benchmark set of problem instances (17), and the computational results indicated that the performance of the algorithm compared favourably with the current state of the art. The MCNDP model was integrated with the job shop scheduling model to provide a unified solution framework.

[Insert Figure 8]

[Insert Figure 9]
5. Changes to Timetable Planning Rules and Infrastructure Technology

Investigations were undertaken of the potential capacity benefits arising from changes to the Timetable Planning Rules (specifically, the reduction of planning headways) and infrastructure technology (the introduction of higher-speed turnouts at critical nodes, i.e., those nodes with the highest levels of capacity utilisation).

However, it was found that no additional Main Line services to/from London could be accommodated without reducing the headway to a value of 90 seconds, due primarily to the two-track section of line between Peterborough and Huntingdon. While this finding is in some ways disappointing, it is not particularly surprising to find that there is in fact limited spare capacity in the morning peak on a busy inter-city and commuter route serving London. It is also supported by subsequent capacity analysis work undertaken by Network Rail (19). In respect of the improved turnouts, it was found that, where they could provide significant capacity benefits, there was insufficient space within the constraints of the existing infrastructure to accommodate them, and, where they could be accommodated, they would not provide significant capacity benefits.

6. Conclusions and Further Work

This paper describes novel approaches to (i) the assessment of railway station and junction capacity utilisation and (ii) the re-routeing and re-scheduling of train services to reduce capacity utilisation and thus provide improved service reliability and/or additional capacity at individual station and junction nodes, and also across wider networks. Capacity utilisation (CUI) analysis of the individual nodes (point ends, crossings and station platforms) and intermediate links comprising the OCCASION study area indicated relatively low (23%-26%) average values, but maximum values in the range of 51%-81%. The subsequent optimisation process suggests that these maxima constrain the scope for the introduction of additional services. It was nonetheless found that the re-scheduling of services could both eliminate scheduled waiting time for trains, and enable the accommodation of additional timetabled services, although not on the ECML between the study area and London, where the potential benefits are likely to be greatest. Modest changes to operational rules (in the form of reduced headways) or to infrastructure technology (through the use of higher-speed turnouts) were not found to substantively change this conclusion.

The work undertaken provides a basis for improvements to operational resilience and/or increases in capacity at network nodes, and thus across the railway network. It is being further developed in collaboration with the Universities of Leeds and Swansea in the course of the DITTO Rail Systems project, funded by RSSB under the FuTRO (Future Traffic Regulation Optimisation) research programme. DITTO Rail Systems aims to (i) enable increased levels of capacity utilisation while maintaining standards of operational reliability, (ii) consider real-time movements of individual trains in the context of system-wide traffic management to produce optimal outcomes, (iii) enable the production of
optimal timetables that can be amended in real time in response to operational perturbations, and (iv) to underpin the preceding three objectives by ensuring operational safety.

Work is ongoing, as part of the DITTO Rail systems project, and also in parallel collaboration with Arup and the ACCVA project partners, to further refine the CUI and UIC 406 calculation processes, and to calibrate both measures in order to establish suitable overall upper limits for station and junction nodes. For the CUI development work, it is planned to build upon the experience gained in the Capacity Charge Recalibration work, and to use a combination of simulation techniques and historic timetable and performance data, to test the robustness and sensitivity of the models developed, including their application to a larger set of case studies.

References

1. DfT, High Speed Rail: Investing in Britain’s Future,  

2. RSSB. EPSRC/RSSB/DfT Strategic Partnership – Capacity at nodes,  


6. Network Rail. Timetable Planning Rules,  


16. Glover F. Future paths for integer programming and links to artificial intelligence. 


Figure 2: Uncompressed Timetable Graph for Pirbright Junction

Figure 3: Compressed Timetable Graph for Pirbright Junction
Figure 5: Uncompressed Timetable Graph for Southampton Airport (Parkway)

Figure 6: Compressed Timetable Graph for Southampton Airport (Parkway)
Figure 7: Timetable Optimisation Framework for a Single Node

Figure 8: Cyclic Time-Space Network with Congestion
Figure 9: Possible Solution