APPENDIX 2: Issues in defining and measuring railway capacity

A2.1 Introduction

A2.1.1 The purpose of this appendix is to explain some of the issues involved in defining and measuring railway capacity. It draws heavily on work undertaken by Jonathan Tyler for GNER as part of its franchise bid. As an illustration we use the East Coast Main Line between Doncaster and London. The next section discusses the physical characteristics of the East Coast Main Line that can lead to bottlenecks. We then consider in turn the issues involved in developing a commercially attractive timetable for ECML inter city services, the problems involved in accommodating freight on the route and the problems posed by the necessity to share capacity with inner suburban services on the southern end of the route. Finally we present our conclusions.

A2.2 Infrastructure issues

A2.2.1 The East Coast Main Line between Doncaster and London King's Cross illustrates the difficulty of defining 'bottlenecks'. In effect, discussion of practical timetabling leads to the conclusion that there is no such thing, *prima facie*, as ‘a bottleneck’. There are instead multiple constraints, and relief is needed for those that cause real pain, severally or collectively, to an optimum-seeking timetabler.

A2.2.2 Significant pathing problems arise from shortcomings in the infrastructure at a number of locations.

- Most of the available routes for coal trains between the Port of Immingham and the power stations cross the path of ECML trains.
- South of Doncaster several flyovers enable freight movements and some minor passenger flows to be kept clear of the main line, but the principal passenger flows all cause conflicts at flat junctions: up Leeds trains and southbound trains between York and Sheffield have to cross the down line, and all trains between Humberside and South Yorkshire must cross the ECML.
- At Newark Crossing passenger and freight trains on the Humberside … Lincoln … East Midlands corridor cross the ECML.
- At Grantham trains from Nottingham conflict with northbound ECML trains.
- The transverse route between the East Midlands and East Anglia via Peterborough is grade-separated south of Peterborough station, but the arrangement of tracks and platforms involves conflicts, includes under-utilised through lines and is unsatisfactory for the station's interchange functions.

- Cambridge Junction at Hitchin, where the ECML is joined by the line from Cambridge, is flat and requires wide margins for certain sequences of movements.
- South of Stevenage is the classic bottleneck, where a four-track railway is reduced to two tracks over Welwyn Viaduct.
- Platform space at Kings Cross and conflicting movements on the approach can constrain planning, although difficulties may be compounded by inefficient operating practices such as lengthy turnrounds.

A2.2.3 Each of these imposes its own restrictions (and note that some affect both directions, some only one), each involves different mixes of services and each may interact with others to affect the movement of a train through the network. Counts of the number of trains using open sections of track between the problem locations might appear to suggest an under-utilised railway, but unless measures can be taken to relieve or remove the constraints that is merely notional.

A2.2.4 Some problems could be resolved by a single and relatively simple project (as has been done recently by building the Allington Chord near Grantham. Others, however, may well only be justified as a package and may therefore become huge projects. A flyover to take down Cambridge trains over the up main line at Hitchin would undoubtedly make pathing more flexible and operations more robust, but it could not increase the number of services run because that is determined by Welwyn Viaduct. And in turn, suppose doubling the latter really were feasible in environmental and cost terms, it is not proven that capacity exists for many, if any, additional trains at the southernmost end of the route, between Alexandra Palace, Finsbury Park and King's Cross, without some very expensive remodelling. In other words, there may here be an acute example of indivisibility.

A2.3 Market and service-planning issues

A2.3.1 The Capacity Utilisation Index is calculated by taking the existing set of services and squeezing them together at minimum headways. The proportion of the relevant time period they occupy is then calculated. the index was originally developed for the specific purpose of examining the impact of capacity utilisation on reactionary delay, where the margin available for recovery from delays in the key issues. It is much less relevant to the measurement of timetabling bottlenecks, where such an arrangement of services makes little commercial sense.

A2.3.2 In respect of passenger services, a host of requirements mean that trains cannot sensibly be pathed simply on the basis of 'efficient' capacity utilisation. Indeed, in many circumstances in a complex network like Britain’s (as distinct from a railway with more homogeneous trains, few junctions and simple routeing) it is the relationship between services that really determines the use of the route rather than theoretical models based on parameters such as differential speeds.

- For the great majority of flows it is important that trains run at as close to even intervals as is possible.
- There are considerable marketing and operational advantages in a timetable that repeats itself in a clear cyclic pattern every hour (or in some cases every half-hour or every two hours). Downward variation in demand can be accommodated by removing selected runs, but that is the limit of

adjustment without disrupting the scheme, since deviation from the normal path usually has a disproportionate ripple effect on other paths and undermines the simplicity of the marketing message. Upward variation, i.e. the peak periods, is ideally accommodated by superimposing extra trains on the standard arrangement of services, but this is not always possible for infrastructure reasons and it does mean that the basic pattern has to be planned around the peak requirements.

- Maximising capacity might suggest some standardisation of paths and stops, but this clashes with market requirements. The Edinburgh ↔ London trains cannot compete with air for travellers whose high value of time governs their decisions (or those interlining at the London airports), but equally there is a market segment which is willing to trade speed for the attractive characteristics of rail. For these customers, nonetheless, the journey must not be too long nor involve too many intermediate stops (which appear to have a detrimental effect on perceptions), and the corollary is thus that flows not served directly must be provided with good connections (e.g. between Scotland and Durham or Grantham). Similar considerations apply to the West Riding ↔ London market relative to the road alternative, though perhaps less strongly. This might lead to dividing the four basic Inter-City East Coast [ICEC] paths into two pairs, one with few stops and one with rather more.
- The 'fifth' path (which is undoubtedly available, given a standard-hour approach) presents its own need for tactical decisions. One argument is that it should provide the main service for the three stations between Doncaster and Peterborough, which all generate much less demand than those two key nodes. Its timings would be closely related to those of the four core services to secure good journey times, albeit with a change, and it might take on supplementary functions, such as providing Huntingdon with an inter-city service. This is however not readily compatible with the desire of open-access operators to use the fifth path for fast services between northern centres and London, and if their trains did not serve all the intermediate stations it would be necessary to insert stops in the ICEC expresses. That too is not straightforward, because while it may deal with the dominant London traffic it can lead to anomalies in connections for other places. The outcome could be messy, especially with less than hourly services, unless the timetable for both, or all three, operators is designed in an integrated fashion.
- It is important to plan interchanges between services in order to enhance the connectivity of the network. (Arithmetically it can be shown that a prerequisite of optimisation is that all timings in a network should be centred on a chosen 'base-minute' and that the timetable for each direction should be a mirror-image of that for the other, a rule that introduces its own constraints.) There is evidence that the historical neglect of this factor, compounded by the planning procedures initiated at privatisation, has created a situation in which considerable revenue is being forgone because lengthy waits are unappealing in themselves and prolong overall journey times. (Johnson, Nash and Tyler, 2004).

- This is not just a matter concerning obvious junctions, such as York for traffic from and to Scarborough, or Peterborough for virtually all movements between Scotland, North East England and Yorkshire & Humber on the one hand and northern East Anglia on the other. It can also be a function of stopping patterns. Consider for instance the issue of stops at Stevenage. In a standard-hour timetable comprising, say, four trains it is unlikely to be appropriate to stop all four or none at all, but stopping one, two or three requires careful structuring of the sequence of services if there is to be a sensible offer between the range of places on the ECML and Stevenage.
- Decisions on routeing can also interact with the ordering of stops. At present the Liverpool ↔ Norwich service uses the ECML between Grantham and Peterborough. By good timetabling this can create the opportunity to serve the Grantham ↔ London business partly by one direct train in each hour and partly by a change at Peterborough. If the Liverpool ↔ Norwich were diverted to run via Melton Mowbray in order to release capacity this might lead to an additional stop in an ICEC express, with its attendant consequences.
- The configuration of stations can impose restrictions on a timetabler's freedom of manoeuvre. Peterborough is particularly difficult. In the up (southbound) direction two platforms (2 and 3) can be used by express services. This could allow trains to run at the line headway of 4 minutes by occupying each platform alternately (although that might then create passenger-handling problems by concentrating the flow in a short space of time). However, because the bay platform (1) requires down terminating trains from London to cross the up line and because it is not long enough for a train composed of two 4-car sets, the WAGN services mostly run into platform 3. The consequence is that expresses must follow each other through platform 2, which implies a minimum headway of 7 minutes for trains from Doncaster (or an intermediate stop). Northbound, a different set of constraints arises from the fact that the island (platforms 4 and 5) is shared with the East Anglia trains.
- Any revision of the timetable for the ECML would have repercussions on many other routes. Constraints on the extent to which their timetables can be altered may affect the success or otherwise of attempts to optimise the ECML, but the logistical challenge of widespread recasting imposes its own costs. Even if one starts with the luxury of a clean sheet there can be difficult interactions. For example, a requirement for the two Sheffield ↔ Birmingham services in each hour to be evenly-spaced determines their rather close relative timings at York so long as one runs via Doncaster and the other via Leeds. Similarly, it is desirable that the fast services between York and Leeds should be at even intervals in order to offer a ‘turn-up-and-go’ facility for one of the largest inter-urban flows outside the South East, but the routes concerned extend to Edinburgh, Newcastle upon Tyne, Middlesbrough, Scarborough, Liverpool, Manchester, Manchester Airport, Plymouth, Bristol and Bournemouth. And the point can extend to secondary and local lines too, ranging from the problematic Edinburgh Park & Ride service to Newcraighall through questions about minor rural

stations on the Newcastle … Carlisle line to how best to address the degraded condition of the Bishop Auckland branch.

- Nor can turnrounds be ignored in most timetabling exercises. Under some railway administrations the timetable is planned independently and the cost of underutilised rolling stock, if that is the outcome, is carried by the general support payments (and sometimes justified too by the cover it gives for peak periods). In Britain, however, that is not a feasible approach, and the diagramming of stock is a material consideration. Moreover, it may be affected by issues such as the suitability of station layouts for turning trains, margins to ensure that late arrivals do not cause late departures and requirements for cleaning and other servicing.
- Finally, it is not always appreciated how tight much timetabling can be on a busy main line. The convention in Britain is to work in half-minutes (some European systems use tenths of minutes), but those half-minutes can make a difference and thus call into question the advisability of high-level bidding that may be negated when it comes to detailed analysis. This feature arises from such matters as signalling overlaps, platform reoccupation rules and the distribution of the additional time inserted to allow recovery from everyday perturbations[1]. And to complicate matters further, it is possible to rearrange the latter in order to manage difficult circumstances: notably, the imperative of scheduling trains as close to each other as possible through the Welwyn Viaduct bottleneck can only be achieved if consequential sub-standard headways beyond are removed by judicious relocation of allowances, that is by slowing trains down.

A2.4 Freight trains

A2.4.1. On the ECML theoretically a maximum of eight ICEC trains (and no freights) could in theory be pathed every hour, given the existing limitations (this would only be compatible with the constraints discussed in the next section if some significant changes were made there). However, such a scheme would probably depend on similar running times and hence on distributing intermediate stops in a manner that would be detrimental to journey-times and connectional arrangements (running eight trains might be appropriate during the peak periods where additional limited-stop services are superimposed on a basic pattern of say five or six trains).

A2.4.2 Six trains/hour is likely to ensure a more attractive timetable, including a mix of few-stop and several-stop trains, unless moving sheer numbers of people were to become the dominant consideration. 'Clean-sheet' exercises have shown that a freight path would take out one of these, and two freights probably two[2]. Arranging for five ICEC expresses and one fast freight permits

---

[1] Practice differs on this. At present in Britain there are three separate types of allowance that are not clearly differentiated from each other nor always sensibly distributed along a train's route. The former Southern Region and some mainland European railways prefer to make a percentage addition when calculating each sectional running time.

[2] Even a high-speed *Freightliner* is of course slower than the passenger trains and therefore might be expected to absorb at least two paths. That it may in practice take only the one path is because ICEC expresses are best grouped in pairs as part of the overall scheme. The need to allow for crossing movements at Newark is also a factor, as is the timing of the Liverpool ↔ Norwich service.

a satisfactory timetable for the former, but introducing a second (perhaps slower) freight would probably lead to a deleterious redistribution of stops. This is of course complicated by the presence and aspirations of open-access operators and by GNER's obligations under the franchise agreement. Scarcity pricing on a marginal basis could usefully clarify the priorities.

A2.4.3 These judgements are based on analysis (not yet completed) of the interactions between the ideal ICEC scheme, the London suburban services (see the next section) and the configuration of the route between Doncaster and Peterborough. It should be stressed that they assume a 'clean sheet' for timetable design. Their basis is therefore entirely different from that of recent studies by Network Rail, which have related to the identification of additional paths in a timetable notorious for the inefficiency of its non-regular pattern.

A.2.5 The influence of the London suburban services

A2.5.1 It is also important to understand the inter-relationship between groups of services where the structure of one can profoundly affect the structure of another. A particularly striking example concerns the WAGN suburban services at the London end of the East Coast Main Line that largely determine the availability of ICEC paths. It is worth examining this case in depth.

A2.5.2 WAGN presently operates an off-peak standard pattern comprising seven routes:

- Cambridge ↔ London King's Cross Fast : half-hourly, non-stop
- Cambridge ↔ London King's Cross Semi-Fast : hourly, calling at Royston, Baldock, Letchworth, Hitchin, Stevenage and Finsbury Park
- Cambridge ↔ London King's Cross Slow: hourly, calling at all stations to Hatfield, then Potters Bar and Finsbury Park
- Peterborough ↔ London King's Cross Semi-Fast : hourly, calling at all stations to Biggleswade, then Hitchin, Stevenage and Finsbury Park
- Peterborough ↔ London King's Cross Slow: hourly, calling at all stations to Hatfield, then Potters Bar and Finsbury Park

[the four Semi and Slow services collectively form the Outer Suburban Group]

- Welwyn Garden City ↔ London Moorgate Inner-Suburban : all stations, every 20 minutes
- Hertford ↔ London Moorgate Inner-Suburban : all stations, every 20 minutes.

A2.5.3 The associations between these services are important.

- The structure ensures that each station has a service appropriate to its relative importance, particularly in respect of frequency and journey-time to London.
- In each half-hourly cycle the Semi precedes the Slow at Stevenage southbound (and *vice versa* northbound), with the Semi coming from Cambridge and the Slow from Peterborough in one half-hour and the reverse scheme in the other – this arrangement gives passengers on certain

relations not served directly every half-hour a with-change option in a similar journey-time to the through time at the other half-hour.

- The two Semi paths form a 30-minute interval south of Hitchin, and likewise the two Slow paths.
- The two Inner services have a common section between Alexandra Palace and Moorgate, with an even 10-minute interval.
- One Hertford train in each hour starts from / terminates at Letchworth, in order to link Stevenage and north thereof with stations on the Hertford Loop.

A2.5.4 The one significant weakness in this scheme is that the 30-minute cycle of Slows at Welwyn does not fit comfortably with the 20-minute cycle of Inners. Apart from that the scheme appears to be an elegant solution. No other solution that would be as fit for purpose has been identified, although some scope may exist for minor modifications. In discussion with WAGN we have established that no variant has ever been demonstrated to be commercially or financially preferable, and WAGN has no proposals for any radical change, at least pending *Thameslink 2000*.

A.2.5.5 A detailed analysis of options for this stretch of the route (available from the author) suggests the following. The pattern of suburban services restricts the number of paths available to ICEC trains (assuming no freight paths over this stretch of the route). Paradoxically adjusting the inner suburban frequency to run every 15 minutes would allow one, or possibly even two, extra ICEC trains per hour. This illustrates the complexity of the inter relationships between different parts of the network.

A2.6 CONCLUSION

A2.6.1 This appendix has demonstrated a number of key issues regarding the measurement of railway capacity. Firstly, the physical characteristics of the infrastructure impose multiple constraints and attempts to overcome one of them may simply activate another. Secondly, the commercial requirements of passenger services themselves impose constraints, which prevent the most 'efficient' use of capacity in terms of maximising the number of trains run; this means that the capacity utilisation index may overstate practical capacity availability. Thirdly, the relationships between different groups of services impose their own constraints, with sometimes paradoxical consequences in which increasing the frequency of one set of services may actually permit higher frequency on others.