

The practicalities of scarcity charging

Final Report

March 2014

Prepared for:
Office of Rail Regulation
One Kemble Street
London
WC2B 4AN

Prepared by:
Steer Davies Gleave
28-32 Upper Ground
London SE1 9PD

+44 (0)20 7910 5000
www.steerdaviesgleave.com

CONTENTS

EXECUTIVE SUMMARY	I
Background (Section 1)	i
Railway capacity (Section 2)	i
Economic approaches (Section 3)	i
International experience (Section 4)	ii
Design of scarcity charging (Section 5)	ii
Primary effects on operators of scarcity charging (Section 6)	iii
Secondary effects on operators of scarcity charging (Section 7)	iv
Effects on capacity allocation by Network Rail (Section 8)	iv
Role in investment decision-making (Section 9)	v
Conclusions and possible next steps (Section 10)	v
1 BACKGROUND AND OBJECTIVES	1
Context to this study	1
The economic rationale for scarcity charging	1
The legal context	2
The practicalities of scarcity charging	3
The structure of this Final Report	3
2 THE CONSUMPTION AND CREATION OF RAILWAY CAPACITY	4
Introduction	4
The consumption of railway capacity	4
The creation of railway capacity	7
Summary	10
3 ECONOMIC APPROACHES TO SCARCITY CHARGING	11
Introduction	11
Opportunity cost	11
Bilateral trading or barter	12
Market-clearing prices	12
Lowest value in use	14
Summary	15
4 INTERNATIONAL EXPERIENCE OF SCARCITY-RELATED CHARGES	16
Introduction	16
Comparison of charging structures	16

Final Report

Scarcity charges as defined in EU law.....	17
“Scarcity-related” charges.....	17
The incentives to operators provided by scarcity-related charges	18
The ability of charges to allocate capacity in periods of congestion.....	19
The effect of charges on operator decisions.....	19
The effect of charges on infrastructure manager decisions.....	20
Summary.....	20
5 DESIGN OF SCARCITY CHARGING.....	21
Introduction.....	21
The location of congestion	23
The period(s) of congestion	26
The measure of scarcity	27
The measure of value	27
The calculation process.....	28
Practical issues	29
Summary.....	33
6 PRIMARY EFFECTS OF SCARCITY CHARGING.....	35
Introduction.....	35
The scale and timing of primary effects	35
The effect on non-timetabled services	37
The effect on freight operators	37
The effect on charter passenger operators	38
The effect on open access passenger operators.....	38
The effect on franchised passenger operators	38
The effect on workload	41
Summary.....	42
7 SECONDARY EFFECTS ON OPERATORS OF SCARCITY CHARGING.....	43
Introduction.....	43
The mismatch between market-clearing prices and optimal use	43
Distribution between operators	44
The reserve price effect.....	44
Behaviour at the boundary of charging locations and periods	45
Changing the balance of risk between operators and funders.....	46
Distribution between customers	47

	Peak services withdrawn and not replaced	47
	Distorted investment incentives	48
	Summary of mitigation measures	48
	Summary	50
8	EFFECTS ON CAPACITY ALLOCATION BY NETWORK RAIL	51
	Introduction	51
	Railway timetabling	51
	Allocation of capacity	53
	Implications for scarcity charging	55
	Shadow scarcity charging	57
	Priority systems	57
	Summary	58
9	ROLE IN INVESTMENT DECISION-MAKING	59
	Introduction	59
	The investment planning process	59
	International precedents	60
	Comparison of existing processes and use of a scarcity charge	60
	Summary	62
10	CONCLUSIONS AND POSSIBLE NEXT STEPS	63
	Introduction	63
	International precedents for scarcity charging	63
	Could scarcity charging reflect opportunity cost?	63
	Could scarcity charging improve operator use of capacity?	64
	Could scarcity charging improve allocation of capacity?	65
	Could scarcity charging improve investment decision-making?	65
	Possible next steps	65

FIGURES

Figure 2.1	Train services, round trips and daily diagrams	6
Figure 3.1	Background to this study	11
Figure 5.1	Reading to Gatwick Airport “congested infrastructure”	25

TABLES

Table 2.1	Elements of consumption or use of infrastructure	5
Table 2.2	Means of creation of railway capacity	8
Table 3.1	Market-clearing may not maximise value	13
Table 4.1	Current charging structures compared	16
Table 4.2	Means of increasing capacity by operators unilaterally	18
Table 5.1	Long list of options for scarcity charging	21
Table 5.2	Comparison of airline slot trading and scarcity charging	23
Table 5.3	Network Rail infrastructure declared to be congested	24
Table 5.4	Precedents for administratively set charges	29
Table 6.1	Determinants of the effects of charges on operators.....	35
Table 6.2	Infrastructure with only one credible operator	40
Table 6.3	Possible operator responses to scarcity charging.....	42
Table 7.1	Unintended effects and mitigations	49
Table 8.1	Objectives and criteria in the Network Code	52
Table 10.1	Possible operator responses to scarcity charging.....	64

APPENDICES

A	BIBLIOGRAPHY
B	SCARCITY CHARGING AND CAPACITY ALLOCATION AND INVESTMENT
C	INTERNATIONAL EXPERIENCE
D	AIRPORT SLOTS AND RAILWAY PATHS AND DIAGRAMS

Executive Summary

Background (Section 1)

1. The economic rationale for scarcity charging is well-established. It can in principle send signals to operators and infrastructure managers for efficient allocation of existing infrastructure capacity and efficient investment in new capacity. However, a number of features of the railway industry may influence how effectively it can be implemented in practice.
2. The focus of this study is therefore to examine what practical issues would need to be addressed, and the extent to which they could be addressed, in order to secure the theoretical benefits of scarcity charging.

Railway capacity (Section 2)

3. Railway capacity cannot be subdivided into, or measured in, units. Different operators require it to be packaged in different ways, from single station-to-station journeys (arcs), through single and round trip journeys, and daily and weekly train workings or “diagrams”, to complete regular interval services.
4. An individual train path may be usable by more than one operator but not, on its own, allow it to provide a viable service. A complete daily train diagram, or a set of diagrams, may allow one operator to provide a viable service but may not be of use to another operator.
5. Capacity can be made available with and without investment, and over various timescales, in a number of ways by operators acting alone, Network Rail acting alone, or by coordinated action. Scarcity charges might in principle be used to incentivise at least some of these means of making more capacity available.
6. As activity on the network in Great Britain has increased, and some existing bottlenecks have been eased, scarcity of infrastructure capacity may be becoming increasingly systemic, and not necessarily attributable to a single constraint at a particular location.

Economic approaches (Section 3)

7. Railways have historically used administrative processes to allocate capacity, but from an economic perspective this approach does not consider the relative value of competing applications for capacity.
8. Economic theory establishes the concept of an opportunity cost, the value that an excluded operator could secure were it able to operate a service using currently occupied capacity.
9. As a proxy for opportunity cost, prices in some markets can be set at market-clearing levels, but there are practical and theoretical grounds for assuming that this approach may not be applicable or efficient for railway infrastructure.
10. As a proxy for market-clearing prices, it would theoretically be possible to estimate the value in use of existing operators, and to set scarcity charges at or below this level with the aim of providing at least some incentive to users to reroute, retime or withdraw their services.

11. Alternatively, scarcity charges could be based solely on a measure, such as Network Rail's Capacity Utilisation Index (CUI), or the UIC 406 measure, of the degree of utilisation of the infrastructure at a particular location and time.

International experience (Section 4)

12. We examined five networks in Europe and found that none applies scarcity charges as defined in EU legislation. However, we found examples of "scarcity-related" charges in Austria (ÖBB), Belgium (Infrabel) and Sweden (Trafikverket).
13. All of the scarcity-related charges could incentivise the withdrawal, rerouting and retiming of low value train services, and Infrabel's approach could also provide incentives for trains to operate at speeds which consume least additional capacity.
14. All the charges have been set administratively, are relatively low, and may have had little effect on operator behaviour, although Infrabel informed us that their charges had reduced the demand for freight train paths in peak periods.
15. None of the infrastructure managers reported any linkage between the scarcity-related charges and the planning of investment.

Design of scarcity charging (Section 5)

16. A scarcity charge would, as a minimum, need to be defined by location, time, a measure of scarcity and a measure of value. We examined a "long list" of options for scarcity charging before focusing on the most practicable.
17. The location of scarcity could be based on infrastructure declared to be congested under the Regulations which implement Directive 2001/14 Article 22. The location and period of scarcity could also be based on measures of capacity utilisation, but an administered review process might also be necessary, to avoid the inclusion of locations where there was no credible need for further capacity.
18. The measure of scarcity could also be a function of capacity utilisation which fell from a peak on fully-utilised infrastructure to zero below a utilisation threshold.
19. The measure of value could be based on the value in use of existing passenger services, using assumptions about the minimum service which an operator could credibly reroute, retime or withdraw. However, different assumptions about the scale of this minimum change, and hence the level of charge at which operators would make it, could result in a wide range of values for the scarcity charge at any one location.
20. Alternatively, a scarcity charge could be set without attempting to estimate value, as has happened in the other networks examined, and its effects monitored. An illustrative charge of £10 per peak train path in one-hour weekday AM and PM peaks could cost some commuter franchisees £200,000 per annum.
21. Scarcity charges would probably need to be recalculated at least annually to reflect changes in utilisation resulting from changes in capacity and services.
22. Scarcity charges could in principle also be applied to freight services.
23. In some cases capacity might also still need to be reserved for prospective but unidentified users: in such cases scarcity charges could only be recovered where the capacity was actually used.
24. To avoid creating incentives to Network Rail to restrict capacity, which would be inconsistent with the existing volume incentive, it might be desirable to make scarcity charges revenue-neutral through a rebate on the Fixed Track Access Charge (FTAC).

25. This might, however, dilute the incentives to the Department for Transport (DfT) to specify services which avoid locations and periods where capacity utilisation is high, unless charges were borne by franchisees. Franchisees might in turn request greater flexibility to reroute, retime or withdraw services, particularly at peak times.

Primary effects on operators of scarcity charging (Section 6)

26. We provisionally conclude that one possible outcome of scarcity pricing would be the emergence of two dominant effects:

- Rebalancing services from peak to off-peak
- Rebalancing the mix of services offered

27. The table below sets out the possible responses of each type of operator. We stress that the table is illustrative and that the extent and nature of responses would depend on the level and structure of scarcity charging in each location and on any other constraints to operators rerouting, retiming or withdrawing services.

Market	Possible withdrawal or retiming	Possible new services
Non-timetabled	Take-up of reserved capacity declines, so less capacity needs to be reserved	Replacement or retimed services outside charging periods
Freight	Take-up of reserved capacity declines, so less capacity needs to be reserved Less demand during charging periods	Replacement or retimed services outside charging periods
Charter	Likely to avoid scarcity charging, given the nature of charter services	
Open access	Uncertain effect, depending on business models of open access operators	
Franchised	Removal or retiming of lower value services during charging periods	Higher value services, or no replacement services

28. Network Rail must keep capacity free for a number of non-timetabled services such as empty stock movements. If users of this reserved capacity were required to pay a scarcity charge, we would expect that the take-up of the capacity reserved would fall and the need for it might be reduced.
29. Freight services do not always use all the capacity set aside for them. If they were required to pay a scarcity charge, we would expect that the take-up of the capacity reserved would fall and the need for it might be reduced.
30. Charter operators would not generally need to operate at specific times or even over specific routes, and we would expect that they would avoid locations and periods with scarcity charging.
31. Open access operators have distinct business models, and might respond to scarcity charging in a number of different ways. We conclude that their response to scarcity charging is uncertain.
32. Franchised passenger operators might, if permitted to do so, operate fewer services at charging locations in peak periods. However, dominant or sole operators might have unintentionally high effective marginal scarcity charges, which may over-incentivise franchised passenger operators to withdraw peak services, depending on the terms of their franchise agreements.

Secondary effects on operators of scarcity charging (Section 7)

33. Scarcity charging might have a number of secondary and unintended effects, some of which are inherent in the nature of the industry and could not readily be mitigated. We focus here on five effects which we consider likely to be most important and difficult to mitigate.
34. First, a posted price scarcity charge based on current value in use, or any other measure, may be a poor estimate of the opportunity cost of the capacity consumed by any individual existing service.
35. Second, careful design would be required to ensure that scarcity charging did not distort incentives near the boundaries of the locations and periods of charging.
36. Third, and as noted above, dominant or sole operators, which will often be franchised passenger operators, might have unintentionally high effective marginal scarcity charges, which may over-incentivise the withdrawal of peak services and distort their behaviour.
37. Fourth, optimisation of the allocation of capacity will probably involve redistribution of the coverage of passenger services between geographical markets. Some locations and communities will gain but others will lose.
38. Fifth, a potentially inefficient outcome is that scarcity charging, coupled with more flexible franchise specifications, would result in the withdrawal of peak services without any immediate or subsequent introduction of services to other destinations being either proposed or possible.

Effects on capacity allocation by Network Rail (Section 8)

39. Current administered procedures are designed to ensure that the timetable is operable, efficient and economical.
40. Railway timetabling cannot be reduced to a series of procedures which lead to an optimal timetable. The Network Code recognises this by setting Network Rail objectives and criteria for devising a timetable but not prescribing the processes to be followed.
41. The passenger timetable is a joint product to the final customer, and the value of any given train path or diagram, to an operator or to society, can only be defined in the context of the remainder of the timetable.
42. Previous studies have suggested that the development of the passenger timetable could not be based on pricing.
43. Reference to actual or shadow scarcity charges could in principle be included in the Network Code criteria.
44. Scarcity charging would not remove the need for the current administered timetabling process.
45. Scarcity charging based on value would require a valuation of each existing service. These valuations could inform capacity allocation decisions better than the resulting scarcity charge.
46. Scarcity charging on a shadow basis would avoid any of the secondary effects to operators described above, but would not address any of these issues.
47. Alternatively, priority systems may be able to achieve some of the benefits of scarcity charging. However, they may be inefficient if the priority rules are only be a poor guide to the actual value of individual train services.

Role in investment decision-making (Section 9)

48. The original vision for the privatised rail industry was that the infrastructure manager would plan investment on the basis of price signals from operators, but this proved unworkable without models and forecasts of demand in final markets.
49. In practice, investment decision-making is a diffuse process. Investment in capacity expansion can be initiated, optimised, appraised, authorised, funded and financed by a number of industry stakeholders with different objectives and requirements. Decision-making must, in some cases, be supported by a business case supported by cost-benefit analysis (CBA).
50. A range of evidence suggests that scarcity charging cannot be as effective as CBA in supporting investment decisions. The principal difficulties are that:
- As indicated above, and unlike deriving the price at which two airlines agree to exchange slots, deriving a scarcity charge would require an administered process, a number of assumptions and potentially a number of judgements.
 - Scarcity charges could not identify whether the most efficient solution required action by train operators, by Network Rail, or on a coordinated basis.
 - Scarcity charges based on the notional location of a constraint might prove no indication of where and when investment in infrastructure would be a cost-effective intervention.
 - Scarcity charges could not reflect both short term “spot” levels of utilisation and long term forecasts taking into account market growth, planned changes to the infrastructure and planned changes to services.
51. Additionally, the use of scarcity charges, analogous to liquidated damages as a “rule of thumb” for addressing minor and recurring issues, seems prima facie unlikely to be appropriately robust for the appraisal of large, infrequent and irreversible investments in infrastructure capacity.

Conclusions and possible next steps (Section 10)

52. The international precedents for scarcity charging are limited and scarcity-related charges are generally low, although they may reduce demand for peak period freight paths.
53. Scarcity charging could not directly reflect opportunity cost but could be based on value in use, as a proxy for market-clearing prices, as a proxy for opportunity cost. Scarcity charges calculated on this basis could vary widely depending on the process and assumptions used.
54. If scarcity charges, however calculated, were sufficiently high, they might influence operator use of capacity, especially if franchisees had more flexibility to withdraw services. Effects might include a rebalancing of services from peak to off-peak and of the mix of services offered. However, where there was a dominant or sole operator the effective marginal scarcity charge might be higher than intended, and over-incentivise the withdrawal of low value services.
55. One possible outcome would be that peak services were withdrawn but no new commercial services replaced them.
56. We found no evidence that scarcity charging could make a practical contribution to the processes of capacity allocation or investment decision-making.

Final Report

57. If ORR wished to investigate scarcity charging further, we suggest carrying out a case study of a route on which one or more applications for capacity have not been satisfied, with the aim of:
- Estimating what scarcity charge would cause the rerouting, retiming or removal of a service
 - Identifying whether this would make it possible to include an unsatisfied application
 - Comparing the overall value of the original and revised timetables

1 Background and objectives

Context to this study

- 1.1 In July 2013 ORR published its long-term regulatory statement, which set out its vision of a simpler, more efficient charging structure (see Appendix B10.1). It stated that:

ORR's vision is of a simpler, more efficient charging structure. It is possible to conceive of a charging structure which reflects the full cost of providing capacity including both short-run wear and tear costs of running trains on the network and the costs of scarcity and congestion. These costs may vary by geography and time of day but we understand the need to balance the pursuit of cost reflectivity with simplicity and practicability.

- 1.2 In February 2014 ORR appointed Steer Davies Gleave to carry out a study of the practicalities of scarcity charging, to report by 31 March 2014.

The economic rationale for scarcity charging

- 1.3 The economic rationale for making use of pricing signals is well-established and has been detailed at length in a number of earlier studies (see Appendix B). We have not repeated these arguments in depth, but following the “Scoping study for scarcity charges” (ITS, 2006) note that:

- According to economic theory, the most efficient allocation of resources in an economy arises from the pricing of all commodities at marginal social cost
- When capacity is optimally adjusted to demand, the scarcity price is exactly equal to the incremental cost of additional capacity

- 1.4 From an economic perspective, therefore, the lack of use of pricing signals to inform decision-making in capacity allocation and investment planning is a potential missed opportunity. It is also possible to identify a number of objectives for scarcity charges, each of which reflects a possible criticism of the current system:

- Sending signals for the efficient allocation of capacity to the system operator
- Encouraging the system operator to accommodate additional capacity
- Sending signals and incentivising operators for the efficient use of capacity
- Encouraging informed decisions about subsidy and expansion of the network

Sending signals for the efficient allocation of capacity to the system operator

- 1.5 The current system for allocating capacity does not rely on any price information. It focuses first on meeting existing contractual rights, which may not be those which would generate the highest value, and then on meeting requests for additional or changed capacity within operational constraints, while ensuring that acceptable levels of punctuality and reliability are maintained.

Encouraging the system operator to accommodate additional capacity

- 1.6 The system operator is not incentivised to accommodate applications for scarce capacity as a result of the price offered for access reflecting its scarcity value, and the decision about whether to accept an application is an administrative one.

Sending signals and incentivising operators for the efficient use of capacity

- 1.7 Similarly, the behaviour of train operators seeking access to scarce capacity does not currently depend on any information on its social value at the margin signalled through price: they are neither deterred from making nor encouraged to make applications by considering the impact of a scarcity charge on the value derived from proposed new services.

Encouraging informed decisions about subsidy and expansion of the network

- 1.8 Proposals for expanding capacity, which are largely driven by the Department for Transport's and Transport Scotland's respective High Level Output Specifications (HLOSs), are based on investigations of network constraints and Cost Benefit Analysis (CBA) of specific schemes rather than on operators signalling the value of capacity enhancements through the price they are prepared to pay for access.

- 1.9 In principle, these objectives could be met through the introduction of a scarcity charge. However, a number of conditions are necessary before pricing signals can be effective:

- A means of defining units to be priced and traded, either as produced by the system operator or as packaged and consumed by the system users
- A smooth, or at least well-defined, demand curve effectively determining the rank order of different demands for capacity in terms of market (and ideally social) value
- Lack of dominance or monopsony among buyers in any location or period such that prices are not distorted
- Timely updating of prices to send efficient signals reflecting market conditions: either prices must be able to change to reflect changing circumstances, or circumstances must remain relatively stable so that a price calculated at a point in time continues to convey reasonably accurate information about the value of capacity over an extended period

- 1.10 We set out in Appendix B how studies undertaken since the late 1980s in the two most liberal railway environments in Europe, Great Britain and Sweden, have identified challenges of meeting these conditions in the case of rail services.

The legal context

- 1.11 The rail industry of Great Britain operates within the framework of European Union (EU) Law and charges for the use of the infrastructure must be consistent with the relevant Directives and Regulations. The most relevant legislation is Directive 2001/14 which, inter alia, refers to scarcity charges in Article 7.4:

The infrastructure charge may include a charge which reflects the scarcity of capacity of the identifiable segment of the infrastructure during periods of congestion.

- 1.12 In our review of international experience in Appendix C we examined the legal basis of charging in other countries, whether based on Article 7.4 or other Articles. We understand that ORR is taking legal advice on the implications of the legislation, including the meaning of “periods of congestion”.
- 1.13 Nothing in this report should be regarded as legal advice on whether any particular approach to scarcity charging should be considered compliant with EU law.

The practicalities of scarcity charging

- 1.14 In this study, we have investigated the practical issues that would need to be addressed in order to introduce a more economic, price-based approach to the use, allocation and enhancement of scarce capacity. Note that this would not necessarily involve a wholesale replacement of the current systems and processes with a market-based mechanism. Rather, we have considered whether the setting of a scarcity charge could enhance or complement the current arrangements such that decision-making would result in outcomes closer to the economic optimum.
- 1.15 We also note that for the scarcity price to be equal to the cost of additional capacity, existing capacity must be optimally adjusted to demand, which in turn means that it must be possible to expand capacity in incremental units. It is well established that this condition is rarely if ever met in the rail sector, in which investment in new infrastructure and rolling stock is typically “lumpy”, with the result that a single charge cannot ensure both the efficient allocation of existing capacity and optimal levels of expansion.
- 1.16 We have therefore focused on the question of whether a scarcity charge might usefully inform short term capacity allocation and long term investment by signalling potential value in use (in economic terms, Short Run Marginal Cost), and have not considered whether the charge could in practice reflect the long run cost of expansion (Long Run Marginal Cost).

The structure of this Final Report

- 1.17 In this Final Report we therefore include in turn:
- A summary of how capacity on railways is consumed and can be expanded
 - A discussion, from an economic perspective, of approaches to setting charges
 - International experience of charges which appear to reflect scarcity
 - The design of possible scarcity charging
 - The primary and secondary effects on capacity use by operators
 - The effects on capacity allocation by Network Rail
 - The role of scarcity charging in investment decision-making
 - Conclusions and possible next steps

2 The consumption and creation of railway capacity

Introduction

2.1 We set out in paragraph 1.9 that, before pricing signals can become effective, it is necessary to have a means of defining units to be priced and traded, either:

- As produced by the system operator
- As packaged and consumed by the system users

2.2 In practice, it is difficult to define a unit or increment of railway infrastructure capacity, and way in which capacity is used by operators varies and is complex.

2.3 In this Section we therefore provide brief overviews of:

- How train operators consume infrastructure capacity
- How train operators and infrastructure managers can make additional capacity available within the constraints of a given rail network or through investment

The consumption of railway capacity

2.4 Any approach to charging for capacity will require some measure of consumption or use to which the charge can be applied. A number of studies refer to a “slot”, implying an analogy with an airport runway slot but, as we set out in Appendix D, the analogy is limited.

Standard paths

2.5 One approach which can be used on individual sections of infrastructure is the “standard path” based on the principle that trains operating at the same speed require less spacing.

2.6 The standard path approach has been used by Eurotunnel, whose 50 kilometre long tunnels are used by 100kph and 120 kph freight trains, its own 140 kph shuttles, and by 160kph Eurostar trains. Eurotunnel has defined a standard path as 140kph, which means that its own shuttles consume one path but other services, whether slower or faster, consume more than one. While freight trains may not be able to operate at 140kph, Eurostar has the option of reducing its consumption of capacity by slowing its services from 160kph to 140kph.

2.7 The “standard path” can be an effective measure of **consumption of capacity** over a single extended constraint between points of convergence and divergence of trains travelling through the constraint at different speeds. However, it is not a **unit of capacity** which is of use, in isolation, to any given train operator.

Capacity utilisation indices

2.8 Another approach which can also be used on individual sections of infrastructure is a Capacity Utilisation Index (CUI), which can be calculated by a standard method such as UIC 406. This identifies the proportion of time unused in “gaps” between trains. Making existing trains operate at more similar speeds will reduce the CUI and, in principle, allow more trains to be timetabled over an individual section.

- 2.9 Railtrack investigated charging individual trains on the basis of their contribution to CUI in the early 1990s, with the aim of incentivising operation at more similar speeds. However, the studies identified a number of practical difficulties with such an approach, as we discuss in Appendix B7.10.

Capacity as required by train operators

- 2.10 In practice, train operators combine individual station-to-station movements or arcs into a complete pattern of service over their entire service network, rather than an individual section of infrastructure. In Table 2.1 we summarise the different building blocks of how an individual passenger train makes use of infrastructure within a day or over a number of days.

TABLE 2.1 ELEMENTS OF CONSUMPTION OR USE OF INFRASTRUCTURE

	Short distance services	Long distance services
Arc	From one station call to the next, up to 100-200 per day	From one station call to the next, typically 20-30 per day
Train service (A to B)	Up to 20 services per day	As few as 2 services per day (1 for sleeper services)
Round trip (A to B to A)	Up to 10 services per day	May not return to starting point within a day
Closed diagram (Return to starting point at end of day)	Common, but some morning and afternoon working are effectively independent	Common
Open diagram (Return to starting point after several days)	Common, but at least one point on the diagram will be passed through regularly	Common (We are aware of a “circuit” only completed over 5 days)
Regular interval service	Common, requires multiple diagrams	Common, requires multiple diagrams

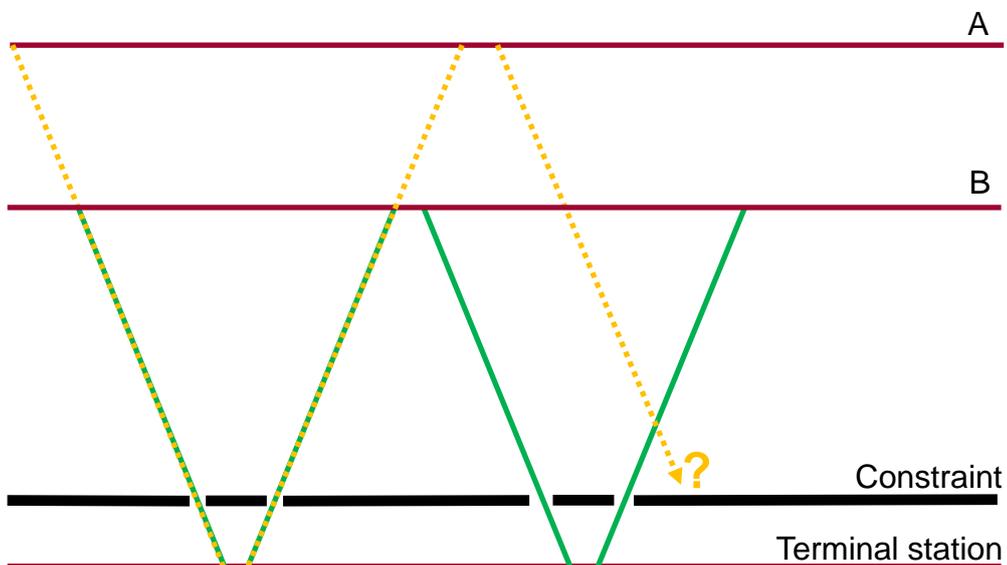
Source: Steer Davies Gleave analysis, numbers are illustrative

- 2.11 Providing an individual passenger with a point-to-point journey going more than one stop is a joint product of a number of arcs or a whole end-to-end train service.
- 2.12 Providing a regular interval service, which both operator and funder may see as the minimum credible level of service, is in contrast likely to be a joint product of a number of diagrams operated by different trains throughout the day.
- 2.13 The distinctions are important because any discussion of the cost or the value associated with a train must relate to a measure of what minimum package of elements of service an operator might remove or add in response to a scarcity charge. For an operator wishing, or required by a funder, to operate a day-long half-hourly service, a scarcity charge at particular times might be treated as an unavoidable cost to the service as a whole.
- 2.14 For an aircraft, as we discuss in Appendix D, the minimum service which could be added or removed would nearly always be an out-and-back service from its base airport. For a train, however, even the concept of out-and-back is problematic:

- There is no operational requirement for trains to make only out-and-back journeys, and in many cases they do not.
- Long distance trains may only make one or two full journeys in a daily diagram, and not return to their starting point, leaving no identifiable out-and-back service.
- Short distance trains may make a number of round trip journeys between the same two points, but there may be no clear rationale for identifying one direction as “out” and the other as “back”. On Thameslink, the future Crossrail, or on services between Glasgow and Edinburgh, for example, there is no basis on which to identify one direction as “out” and the other as “back”. This means that any one journey subject to a scarcity charge might be removed or modified in conjunction with the preceding one or the following one, but that there would be no prima facie reason to identify which.

- 2.15 In practice, operators rarely apply for capacity to provide a single arc or even a single train service from an origin to a destination. Rather, they require a number of efficient daily or weekly diagrams, the patterns of use of their fleet over the whole day, taking into account the location of activities such as stabling, fuelling and maintenance.
- 2.16 Figure 2.1 illustrates in greater detail the differences between a train service, a round trip and a daily diagram.

FIGURE 2.1 TRAIN SERVICES, ROUND TRIPS AND DAILY DIAGRAMS



- 2.17 In the Figure a route extends from station A to station B, on through a constraint, and into a terminal station. The solid green operator currently operates a train diagrammed to make two round trips per day from station B to the terminal station, in the process passing through the constraint four times.
- 2.18 If the solid green operator were to withdraw services in response to a scarcity charge, and a dotted orange operator wished to operate a potentially higher value service from A, it might be possible to time its first round trip to pass through the constraint at the times now available. However, the longer journey time from and

to station A would mean that it would not also be possible to time its second round trip to pass through the constraint at the times now available.

- 2.19 The services of the two operators are not interchangeable, and removing the solid green service would not necessarily make possible either the dotted orange service or even any other service. The dotted orange operator's services could only be accommodated if capacity could be made available at different times suited to its requirements, through the administrative process of devising a new timetable.
- 2.20 This lack of substitutability between operators means that it is not possible for train operators to trade capacity like airline slots. Even if this were permitted, which it is not, removal of one operator's service would normally only allow another operator to provide an identical service between the same stations with the same speeds and stopping patterns. We are not aware of any examples of a rail service by one operator being directly replaced by another operator, except through the franchising process.
- 2.21 Figure 2.1 also illustrates one of the key features of the rail industry:
- A train path may be usable by more than one operator but would not, on its own, allow any of them to provide a viable pattern of services
 - A train diagram or set of diagrams may allow an operator to create a viable pattern of service but would not, on its own, allow another operator to provide anything other than the identical pattern of service

The creation of railway capacity

- 2.22 We set out as possible objectives of scarcity charging (in paragraph 1.4) sending signals not only to operators for the efficient use of capacity but also to Network Rail to accommodate additional capacity.
- 2.23 In practice, the provision of capacity to ultimate passenger and freight customers is a complex function of decisions made by both operators and Network Rail, both individually and on a coordinated basis.
- 2.24 Table 2.2 summarises a number of ways in which the capacity of services operated on a railway can be made available, noting the typical timescales over which they can be delivered and whether investment is required. The table, which is not exhaustive, distinguishes in turn:
- Changes achievable unilaterally by operators
 - Changes achievable by industry consensus but not requiring major investment
 - Changes requiring investment by Network Rail
 - Changes requiring coordinated investment by Network Rail and all operators
- 2.25 Note also that the even the direction change required to increase capacity may depends on the circumstances:
- By making trains slower (to fit better among slower ones) or by making them faster (to fit better among faster ones)
 - By making trains longer (to provide more space) or by making them shorter (to fit into platforms)

TABLE 2.2 MEANS OF CREATION OF RAILWAY CAPACITY

Means	Issues	Delivery timescale	Investment	
			Operator	Infrastructure
Changes achievable unilaterally by operators				
Reroute train	Assumes acceptable alternative available	Short	x	x
Retime train	Assumes alternative timing acceptable	Short	x	x
Withdraw train	Make way for more valuable service	Short	x	x
Slow trains down	Reduces speed variations	Short	x	x
Add stops	Improves pathing, subject to access rights	Short	x	x
High density trains	Refurbishment may allow more seating or standing passengers but add to dwell time	Short	✓	x
More/spare trains	May reduce the need to use trains with partial power and hence poor performance	Medium	✓	x
Longer trains	Trains may be slower, en route or passing junctions, or require more platforms	Medium	✓	x
High performance	Shorter journey times, particularly for stopping trains	Medium	✓	x
Electric trains	May deliver higher performance	Medium	✓	x
Fixed formations	Reduces need for marshalling and stabling	Medium	✓	x
Shorter trains	May allow two trains in one platform	Medium	✓	x
Changes achievable by industry consensus but not requiring major investment				
Recast timetable	May create winners and losers	Short	x	x
Changes requiring investment by Network Rail				
Longer platforms	May be necessary for longer trains	Medium	x	✓
Remove conflicts	At junctions, increases flexibility of timing	Medium	x	✓
More platforms	Trains can dwell, be passed or be stored	Medium	x	✓
Electrify lines	Provides for higher performance	Medium	x	✓
Upgrade power	Limited power reduces performance	Medium	x	✓
Denser signalling	May let trains operate closer together	Medium	x	✓
Diversionsary routes	Diverts some trains, eases pathing	Medium	x	✓

Means	Issues	Delivery timescale	Investment	
			Operator	Infrastructure
More loops	More flexible passing on single track	Medium	×	✓
More double track	Removes constraint of single track	Medium	×	✓
More loops	More flexible passing on multiple track	Medium	×	✓
More tracks	More flexible passing on multiple track	Long	×	✓
New lines	May also require new rolling stock	Long	×	✓
Changes requiring coordinated investment by Network Rail and all operators				
ATO (and ATR)	May let train operate closer together	Long	✓	✓
ERTMS	May increase capacity	Long	✓	✓
Moving block	May increase capacity	Long	✓	✓

- 2.26 A potential challenge for any structure of industry incentives is to achieve mechanisms and incentives not only for unilateral action but also for multilateral action.
- 2.27 On a lightly-used network, none of these mechanisms may be required, and it may be possible for all applications for capacity to be accepted with little or no coordination or consultation of the operators.
- 2.28 As a network becomes more intensively used, there may be conflicts between the applications of different operators, but it may still be possible to reconcile them with relatively minor adjustments, requiring no capital investment, to the timing or stopping pattern of trains. Rerouting of services may be more difficult, as commercially viable alternatives may not be available, particularly for passenger services. CER (2013) noted “the general lack of redundancy in terms of railway routes which limits the possibilities for alternative routes that customers would accept, especially for passengers and (to a lesser extent) for freight”.
- 2.29 At higher levels of utilisation, investment to expand capacity may be needed, either by one or more operators, or by the infrastructure manager, or by both acting together, such as when a line is converted to electrification or resignalled.
- 2.30 Since restructuring following the Railways Act 1993, levels of demand on the British network have increased rapidly. As relatively simple and low cost measures to increase capacity become exhausted, successive Periodic Reviews have authorised Network Rail to invest in multiple schemes to increase capacity, often with associated investment coordinated by operators and their funders. However, this increasingly means that capacity constraints on the network are not identifiable to specific bottlenecks, as envisaged in Article 7.4, but are systemic. We show in Appendix D3 how a number of routes now operate with multiple

Final Report

constraints. Measures to remove any one constraint may allow only a small increase in capacity, performance or resilience before another is reached.

- 2.31 For the purposes of scarcity charging, this means that it may be increasingly difficult to identify, in the terminology of Article 7.4, an “identifiable segment of the infrastructure” which is the location of, or responsible for, “periods of congestion”. We illustrate this issue with a discussion of the East Coast Main Line (ECML) in Appendix D4.

Summary

- 2.32 There is no unit of railway capacity, whether as produced by the system operator, or as packaged and consumed by the system users.
- 2.33 Removal of one train service will prima facie release capacity, but this will not necessarily allow the provision of any other service being proposed by an applicant operator.
- 2.34 Additional railway capacity can be created by a number of means, only some of which can be achieved by either an operator or Network Rail acting alone.
- 2.35 Constraints to providing additional services may increasingly be systemic, rather than related to identifiable and localised “bottlenecks” or “pinch points”.

3 Economic approaches to scarcity charging

Introduction

3.1 In this Section we review, from an economic perspective, possible approaches to scarcity charging:

- The economic concept of opportunity cost
- Bilateral trading or barter, which needs no centralised pricing policy
- Market-clearing prices by which, in the absence of bilateral trading, a centralised pricing policy can improve allocative efficiency in some markets
- Lowest value in use, a possible alternative measure of value, in the absence of either bilateral trading or a means of identifying a market-clearing price
- Priority systems, an alternative to explicit pricing, as used on some networks

Opportunity cost

3.2 ORR’s “Invitation to tender & statement of requirement” set out the background to this study, as shown in Figure 3.1, highlighting the importance of the concept of opportunity cost.

FIGURE 3.1 BACKGROUND TO THIS STUDY

Where capacity is scarce, when an additional train is granted access to the network, other users, who would otherwise have used the train path, are denied this use, i.e. there is an opportunity cost. It is theoretically efficient for access charges to reflect this opportunity cost, i.e. the cost of scarcity since:

- Operators are made aware of the costs which their use causes and can decide whether they value their use enough to meet those costs.
- It provides a mechanism for deciding between competing demands for network capacity - so that decisions don’t have to rely on expert judgement (as is currently the case).
- It should facilitate more informed decision making about whether investment to increase capacity and alleviate or eliminate this scarcity is required.

3.3 It would therefore be desirable for a scarcity charge for the use of capacity to be set at the level of the opportunity cost, namely the highest value which could be achieved by another user of that capacity were it not already occupied.

3.4 However, application of the concept of opportunity cost depends on there being:

- Demand to operate additional services
- Means to operate additional services with capacity released by an existing user
- Means of estimating the value of these additional services

3.5 Setting aside the issues of the definition of units of consumption and creation of capacity discussed in Section 2, we discuss below a number of different means by which opportunity cost may either be revealed, through market mechanisms, or estimated, through market analysis.

Bilateral trading or barter

- 3.6 The simplest approach to setting a price is through bilateral trading or barter. A buyer and seller settle on a price, which neither will accept unless they consider themselves to be better off as a result of the trade. Barter requires no regulation or central administration.
- 3.7 We set out in Appendix D how increasing scarcity of airline slots at a limited number of airports, coupled with mechanisms permitting the exchange of slots, has led to bilateral trading between airlines to mutual advantage. We note, however, that such trading is not yet sanctioned by EU law and that airlines do not have formal contractual ownership of airport slots, which continues to be based on “grandfather rights” and subject to “use it or lose it” provisions.
- 3.8 However, as Figure 2.1 shows, bilateral trades in capacity would probably not be possible on a railway unless one operator withdrew in favour of another wishing to provide a service using the same train diagram or diagrams, and hence with the same stopping patterns and timings. This lack of direct substitutability, or fungibility, means that there is no unambiguous test of whether capacity is available at any given location. All that can be said with confidence is that:
- Each rail service added to the network is prima facie likely to make it more difficult to run additional services
 - Each rail service removed from the network is prima facie likely to make it easier to run additional services

Market-clearing prices

- 3.9 Opportunity cost may also be revealed through market-clearing prices, a theoretical approach which is well-developed in liquid markets with a smooth demand curve.
- 3.10 The principle of market-clearing prices is that a commodity is allocated to users in descending order of willingness-to-pay until it is exhausted. One approach is for the vendor to invite prospective buyers to submit sealed bids stating the volume required and the price offered, and then allocating the commodity to bidders, starting with the highest bid, until it has all been allocated.
- 3.11 In the absence of bids by which applicants inform the vendor what they are willing to pay, the vendor may estimate the likely market-clearing price, as when setting the face price for tickets for a public event. If the face price is set too high, not all the tickets will be sold. If the face price is set too low, a secondary market may emerge in which tickets are resold, at a profit to the initial purchasers, to those who value them more. For the vendor, the ideal outcome is that the price is set so that everyone willing to pay it obtains a ticket and there are no unsold tickets. At this market-clearing price the outcome is also a Pareto efficient equilibrium, and no further trading in tickets takes place because no one without a ticket values it more than anyone with one.
- 3.12 Market-clearing prices cannot, however, be applied directly to railway infrastructure, for at least three reasons related to the preconditions which we referred to in 1.9:

- First, because there is no market or auction in which capacity is sold.
 - Second, because even if such a market existed, there might also be practical difficulties in establishing a stable market-clearing price. Markets may be thin or illiquid, making it hard to find operators wishing to enter and exit the market on the same part of the network at the same time.
 - Third, because even if a market existed and a market-clearing price emerged, the nature of the consumption of infrastructure capacity means that a market-clearing approach, literally applied, may be inefficient, as we demonstrate next.
- 3.13 To illustrate this last point, assume, for example, that:
- Operator A wishes to operate 3 express trains, with incremental values of £60, then £30, then £20.
 - Operator B wishes to operate 1 stopping train, with incremental value of £40.
 - Timetabling exercises had shown that, if there is a stopping train, then there can be only one express train.
- 3.14 As Table 3.1 shows, in these circumstances a market-clearing approach to pricing would result in accepting Operator A's first train and Operator B's train and then declaring the infrastructure full, with a total value of £100. However, the optimum outcome would be to accept all three of Operator A's trains, with a total value of £110.

TABLE 3.1 MARKET-CLEARING MAY NOT MAXIMISE VALUE

	Incremental value	Market-clearing	Optimised
Express train 1	£60	£60	£60
Stopping train 1	£40	£40	
Express train 2 (given 1)	£30		£30
Express train 3 (given 1 and 2)	£20		£20
Total		£100	£110

Note: illustrative assumptions, see text

- 3.15 In practice, Table 3.1 is a simplification as both Operator A and Operator B would also require capacity at other times and locations to enable their services to be incorporated into workable train diagrams.
- 3.16 Nonetheless, this suggests that, irrespective of any other issues, there are theoretical reasons why reviewing a list of applications for capacity in descending order of value to their operator, and accepting any which do not conflict with those already accepted, may not deliver the most efficient outcome. The most efficient outcome might, in contrast, involve accepting a larger number of lower value but mutually compatible trains.
- 3.17 However, the number of possible combinations selected from n applications rises as $2^n - 1$, so that even four independent requests for trains could involve preparing, and then comparing, the total value to operators of up to 15 possible timetables.

With as many as 20 or 30 trains per hour on some routes, and interactions between routes, this approach would be unworkable in practice.

- 3.18 As we set out in Appendix B6, incorporating passenger operators' signals of value into an automated timetable optimisation process was examined in Sweden in the early 1990s and rejected. This was primarily because of the issue of contingency, that the value of a train path to one operator was dependent on all the other paths awarded to it and to its competitors. Passenger operators could not identify the value to them of a particular train path without first knowing all the other train paths in the timetable.
- 3.19 We also note an inherent feature of setting a scarcity charge based on market-clearing price, which is analogous to setting a face price for tickets to a public event or a reserve price at an auction. The introduction of any charge which deters applications for access also removes information on the shape of the demand curve, making it harder for Network Rail, ORR or other stakeholders to assess the potential demand for new paths without carrying out parallel market research or consultations. We return to this issue in Sections 6 and 9.

Lowest value in use

- 3.20 In the absence of a market in which prices can be observed, we also examined an approach to estimating opportunity cost based on the lowest value in use of capacity by the services in an existing timetable.
- 3.21 We note, however, that (even if a market-clearing price were efficient) if capacity had not already been efficiently allocated to the highest value users, then the lowest value in use might be below the market-clearing price, and would be insufficient to price off operators who could not afford the market-clearing price.
- 3.22 The lowest value in use approach would rely on the fact that, at least for passenger services, it is in principle possible to use standard industry demand and revenue models and other tools to estimate the gross or net value contributed by any element or subset of the current services. Initial work by ORR suggests that it is possible to estimate the value of any element of a rail service on an objective and consistent basis, which might be limited to the net financial value to the operator, but could include the valuation of other effects and externalities.
- 3.23 Any such analysis would require access to detailed information on the actual or modelled number of passengers using the service. We note that this is not public information and would not necessarily be available to Network Rail.
- 3.24 Assuming suitable access to the actual or modelled numbers of passenger using each service, the principle behind the lowest value in use approach would be to:
- Examine the value of all train services in the current timetable at a particular location and period
 - Identify the lowest value in use of these services
 - Use this value as a scarcity charge payable by all operators

- 3.25 Referring to Table 2.2, a lowest value in use approach, in theory at least:
- Would be sufficient to deter applications for capacity which would create less value than any of the existing users of the capacity
 - Might be sufficient to incentivise an existing operator to retime a service away from the peak, or reroute if commercially viable alternatives were available
 - Would, by definition, not be sufficient to incentivise any existing operator to withdraw a service, unless and until the value of the service subsequently fell
- 3.26 Provided that existing operators had scope to reroute or retime their services, or the value of the services fell, a lowest value in use approach could in principle influence their behaviour. If this were not the case, however, it would not be sufficient to incentivise any existing operator to withdraw, no matter what latent demand for high value services existed. A lowest value in use approach based on existing operators cannot, by definition, signal opportunity cost resulting from the existence of potentially higher value and replacement users of the same capacity.

Summary

- 3.27 Each existing rail service will have an opportunity cost, but it will depend not only on the demand by potential operators to provide additional services but also on whether removing the existing service would allow any of the proposed additional service to be operated.
- 3.28 In the absence of a measure of opportunity cost, market-clearing prices can be determined by auctions or estimated. Even if correctly estimated, however, a market-clearing price for rail capacity might not result in an optimal allocation of capacity. Any given price might both exclude services which could be part of an optimal timetable and include services which would not be part of an optimal timetable.
- 3.29 In the absence of an estimate of market-clearing prices, a body with access to actual or modelled passenger numbers could estimate the value in use of each current train service. Scarcity charges could then be set at, or with reference to, the lowest value in use, although if the market is not already efficient this is likely to be below the theoretical market-clearing price.
- 3.30 Given the successive divergence of each of these approaches from the underlying aspiration to identify opportunity cost, a simpler alternative would be either to set scarcity charges on an administered basis with no explicit attempt to estimate value.

4 International experience of scarcity-related charges

Introduction

4.1 ORR asked us to examine international experience of rail scarcity charging.

We would like the consultant to carry out a review of where rail scarcity charging has been adopted elsewhere. We are aware of its use for a limited period in Germany but understand that it was discontinued. Furthermore, we understand that Sweden has a reasonably progressive approach to allocating network capacity on the railway which may involve the use of prices. We would specifically like a comprehensive description of where it has been used elsewhere, the rationale for its introduction i.e. what problem or problems it was trying to solve, the approach to its calculation and design, the extent to which it was considered successful in meeting its objectives and any key lessons which might be applicable in a UK context.

4.2 We agreed with ORR to study the rail networks of Austria, Belgium, Germany, Sweden and Switzerland. We looked for international precedents for either:

- Scarcity charging, as defined in Directive 2001/14 Article 7.4
- Scarcity-related charging which, while not identified as a scarcity charge, had the effect of charging more for infrastructure which had high levels of utilisation or acted as a bottleneck

4.3 Our research is set out in Appendix C and our principal findings are repeated here.

Comparison of charging structures

4.4 Table 4.1 below compares the current charging structures of the five networks examined with those of Network Rail’s Network Statement 2015.

TABLE 4.1 CURRENT CHARGING STRUCTURES COMPARED

Article in Directive 2001/14		Network Rail	Austria (ÖBB)	Belgium (Infrabel)	Germany (DB Netz)	Sweden (Trafikverket)	Switzerland (SBB Infrastruktur)
7.4	Scarcity charge	x	x	x	x	x	x
	Charges relating to “periods of congestion”	x	✓	✓	x	✓	x
12	Reservation or cancellation charge	x	✓	✓	✓	x	✓
22	Congested infrastructure	✓	✓	x	✓	x	x
	Charges on “congested infrastructure”	x	✓	x	x	x	x

- 4.5 For illustrative purposes we have included Network Rail’s existing capacity charge, which is neither limited to infrastructure which has been declared congested nor specifically related to “periods” of congestion.

Scarcity charges as defined in EU law

- 4.6 None of the other networks describe any of their charges as a scarcity charge, reflecting the wording of Article 7.4.
- 4.7 In Germany, DB Netz’s former “Auslastungsfaktor” was claimed to be “an incentive aiming at increasing the performance of the network” but was rejected by the regulator for a number of reasons (see Appendix C5.3). In Sweden, Trafikverket applies a “passage charge” in a number of locations but these have not been declared to be congested infrastructure (see Appendix C6).

“Scarcity-related” charges

- 4.8 However, we identified four charges which in practice reflect “the scarcity of capacity of the identifiable segment of the infrastructure during periods of congestion” as set out in Article 7.4:
- ÖBB’s “supplement for congested infrastructure” and “incentive for capacity optimisation”
 - Infrabel’s combination of “H” and “T” parameters in track access charges
 - Trafikverket’s “passage charge”
- 4.9 We summarise these “scarcity-related” charges briefly below.

ÖBB’s “supplement for congested infrastructure”

- 4.10 ÖBB has declared the mainly 2-track 12 kilometre Mödling to Wien Meidling section of the Südbahn to be congested infrastructure, as provided for under Directive 2001/14 Article 22, and applies a fixed supplement of around €15 for every train using the section during morning and evening peaks. The supplement can, in principle, incentivise operators to reroute, retime or withdraw low value services.

ÖBB’s “incentive for capacity optimisation”

- 4.11 ÖBB also describes a supplementary charge of €0.5638 per train-kilometre applied at a large number of locations, for periods specific to each location, where the capacity utilisation index (CUI) exceeds a particular level. The incentive can, in principle, incentivise operators to reroute, retime or withdraw low value services.

Infrabel’s combination of “H” and “T” parameters in track access charges

- 4.12 Infrabel’s structure of track access charges for the whole network includes an “H” parameter which raises charges in semi-peak and peak hours and a “T” parameter which raises them for trains travelling materially slower (or faster) than a standard path (see 2.5). The incentive can, in principle, incentivise operators to reroute, retime or withdraw or to change the speed, performance or traction type of low value services. Infrabel informed us that the charge had reduced the demand for freight paths in peak periods but does not exclude freight services for which the operator is willing to pay the charge. We note, however, that Infrabel applies the charge network-wide, and hence that it may not be consistent with the Article 7.4 wording “identifiable segments of the infrastructure”.

Trafikverket’s “passage charge”

- 4.13 Trafikverket applies a fixed SEK 260 “passage charge” to any train using five defined sections of infrastructure in Stockholm, Göteborg and Malmö during peak periods. The charge can, in principle, incentivise operators to reroute, retime or withdraw low value services but Trafikverket informed us that no such effect had been observed, probably because the charge is so low. Trafikverket made clear that the charge is not a “scarcity charge” but a “mark-up”.
- 4.14 Of these charges only ÖBB’s “supplement for congested infrastructure” is contingent on the Article 22 test that “is not possible to satisfy requests for infrastructure capacity adequately” and hence the infrastructure has been congested. The other charges have been imposed before this point is reached, and Infrabel informed us that they explicitly expected to have at least some spare capacity at locations and times where the charges are applied.

The incentives to operators provided by scarcity-related charges

- 4.15 Table 4.2 below repeats from Table 2.2 the means of increasing capacity open to operators acting unilaterally, and identified which of the scarcity-related charges have, in theory at least, the potential to incentivise each of them.

TABLE 4.2 MEANS OF INCREASING CAPACITY BY OPERATORS UNILATERALLY

Means (For further detail, see Table 2.2)	Example						Delivery timescale	Investment	
	Network Rail Capacity charge	DB Netz “Auslastung”	ÖBB “supplement”	ÖBB “incentive”	Trafikverket “passage”	Infrabel “H” and “T”		Operator	Infrastructure
Reroute train	✓	✓	✓	✓	✓	✓	Short	x	x
Withdraw train	✓	✓	✓	✓	✓	✓	Short	x	x
Retime train	x	x	✓	✓	✓	✓	Short	x	x
Slow trains down	x	x	x	x	x	✓	Short	x	x
Add stops	x	x	x	x	x	✓	Short	x	x
High density trains	✓	✓	✓	✓	✓	✓	Short	✓	x
More/spare trains	x	x	x	x	x	✓	Medium	✓	x
Longer trains	✓	✓	✓	✓	✓	✓	Medium	✓	x
High performance	x	x	x	x	x	✓	Medium	✓	x
Electric trains	x	x	x	x	x	✓	Medium	✓	x
Fixed formations	x	x	x	x	x	x	Medium	✓	x
Shorter trains	x	x	x	x	x	x	Medium	✓	x

Note: reservation or cancellation charges will also slightly incentivise withdrawal

- 4.16 Network Rail's capacity charge varies by location, and hence may incentivise rerouting, but not by time of day, and therefore cannot incentivise retiming.
- 4.17 DB Netz's "Auslastungsfaktor" was not clearly related to "scarcity" but could in principle incentivise operators to reroute or withdraw trains but not to retime them to avoid "periods of congestion".
- 4.18 ÖBB's and Trafikverket's charges can also incentivise retiming.
- 4.19 Only Infrabel's charges have the potential to incentivise measures to consume capacity en route efficiently: slowing trains down, adding stops, having spare trains or buying high performance or electric trains.
- 4.20 None of the scarcity-related charges can, on its own, incentivise a complete recast of the timetable coordinated across all operators. In the case of Infrabel, we note that it appears that SNCB could in practice recast the entire passenger timetable and present it as an internally consistent and workable request for capacity.

The ability of charges to allocate capacity in periods of congestion

- 4.21 We noted in paragraph 3.19 that scarcity charges set in advance can only act as a reserve price, and cannot ensure that all available capacity is allocated. This is clearly both the case and the intention with Infrabel, which explicitly expects that some capacity will remain unallocated. If the objective of scarcity charges is to ensure that all capacity is allocated, then setting a scarcity charge which acts as a reserve price is potentially counterproductive.
- 4.22 Auction theory suggests that open ascending price auctions, with no reserve price, can ensure that any remaining packages of capacity which can be made available can be allocated to the user that values them the most. SBB Infrastruktur has made provision for auctions where infrastructure has been declared congested, but no SBB infrastructure has been declared congested. All the other approaches we have identified are analogous to reserve prices, which may both deter some bidders and also result in some capacity remaining unallocated.

The effect of charges on operator decisions

- 4.23 ÖBB's Network Statement suggest that scarcity-related charges have an effect on operator applications, but we identified no examples of where this is the case.
- 4.24 Infrabel informed us that the scarcity-related "H" and "T" charges had influenced freight operators to avoid requesting paths at peak periods, and that a market mechanism might be more effective than a priority system because in some cases it was commercially necessary for freight trains to operate at congested locations and times. However, Infrabel also pointed out that it was not considered practicable to charge freight operators on the same basis as passenger operators, even if they consumed the same capacity on the same track.
- 4.25 Infrabel noted that SNCB was replacing much of its fleet with higher performance trains, which might result in a reduction in charges due to the "T" parameter, although is not possible unambiguously to attribute the change to the charging structure.

Final Report

4.26 Trafikverket informed us that their assessment was that charges had so far had no impact on operator applications, but that this was unsurprising while charges remain relatively low. We agree that the limited extent to which charges influence behaviour may reflect the relatively low absolute levels of charges.

4.27 In summary, of the five networks studied, scarcity charges may have influenced:

- In Belgium, freight operators to avoid requesting paths at peak periods. Priority systems could have achieved the same effect but would have prevented some high value freight services. Specifically, priority systems can reflect ex ante estimates of the typical value of a type of train, but not the specific value of an any individual train, particularly for freight.
- In Belgium, potentially, passenger operators to buy higher performance trains, although this might have happened anyway in the absence of charges.

The effect of charges on infrastructure manager decisions

4.28 Infrabel informed us that the charging and investment planning functions were separate and that they were not aware of any case of charges directly influencing investment decisions.

4.29 We have not identified any evidence in other networks of any linkage between scarcity-related charges and investment decisions.

Summary

4.30 There are no scarcity charges, as defined in EU law, in any of examples studied.

4.31 All the “scarcity-related” charges in the international examples studied are set by administrative means, with no attempt to measure value, although SBB Infrastruktur makes provision for auctions.

4.32 Other infrastructure managers have told us that scarcity-related charges can influence operator decisions.

4.33 Other infrastructure managers do not appear to make any linkage between scarcity-related charges and the planning of investment and capacity.

5 Design of scarcity charging

Introduction

- 5.1 We assumed that any scarcity charge would probably need to have at least four elements to reflect some measures of location, time, scarcity and value:
- The location of congestion
 - The period(s) of congestion
 - The measure of scarcity
 - The measure of value
- 5.2 We began our search for options for the structure for a scarcity charge with a long list of options shown in Table 5.1, which also includes the existing capacity charge.

TABLE 5.1 LONG LIST OF OPTIONS FOR SCARCITY CHARGING

	Vary by location and “period” (a scarcity charge)				Vary by location
	All route sections, short time periods	Nominated locations, short time periods	Nominated locations, banded	Nominated locations, peak and off-peak	Whole route, all-day
Long Run Incremental Cost (LRIC)	?	?	?	?	?
Value of each service and CUI	?	?	?	?	?
Value of each diagram and CUI	?	?	?	?	?
Passenger-km on each diagram and CUI	?	?	?	?	?
CUI only	?	?	?	?	Capacity charge
Passengers through nominated location	?	?	?	?	?
Train count	?	?	?	?	?

See text for detailed definitions of options

- 5.3 Columns in the table represent different degrees of disaggregation of location and period including, for reference, disaggregation by location but not by period, the basis of the existing capacity charge.
- 5.4 Rows in the table represent different methods of estimating scarcity and value at each time and location, including measures making use of Network Rail’s Capacity Utilisation Index (CUI).

- 5.5 The CUI is calculated for all 6,688 “constant traffic route sections” (CTSs) of the network and is used to calculate the current capacity charge. Network Rail informed us, however, that while prima facie the ability to timetable additional trains falls as CUI rises, CUI alone is not a reliable guide to whether additional services can be operated, which requires a detailed timetabling exercise. This is consistent with the findings of previous studies summarised in Appendix B7.10.
- 5.6 It would in theory be possible to perform calculations of scarcity charges, by CTS, for all 25,000 train services on the network on a typical day, but might be desirable or necessary to limit the calculation to a smaller number of diagrams or a sample of services.
- 5.7 Taking into account the wording of the relevant EU Directives and international experience (see Appendix C), we assumed that scarcity charges should incentivise as many as possible of the behaviours achievable unilaterally by operators including, as a minimum:
- Withdraw an existing train service or application for a future service
 - Reroute an existing train service from a location which is congested
 - Retime an existing train service at a time which is congested
- 5.8 In each case, the objective would be that the withdrawal, rerouting or retiming of one train might, under some circumstances, result in capacity being available which would enable the operation of another service, of higher value, proposed by a different operator or funder.
- 5.9 We discussed the contents of the long list with ORR before identifying two options for further consideration, one with a specific measure of value and one without. We rejected options which are:
- Invariant with time. This would not be consistent with Article 7.4 or reflect “periods of congestion” and could not incentivise retiming of services.
 - Based on Long Run Incremental Cost (LRIC). Previous studies (most recently CEPA (2010)) have found this not to be a workable approach.
 - Based on passenger-kilometres or passenger numbers as a measure of value. Measures of passenger-kilometres or passenger numbers can reflect only the “gross” usage of the train concerned, rather the net increase in usage caused by its presence in the timetable.
 - Based on train counts. These provide no information of value to customers, although they are the basis of Infrabel’s “H” parameter, see Appendix C4.4).

Comparison of airline slot trading and scarcity charging

- 5.10 Table 5.2 compares:
- The actions required by airlines wishing to trade slots at a congested airport (which we discuss and compare with railway capacity allocation in Appendix D). Airlines do not require any processes to identify either which location (airport) or what time period they are discussing.
 - The actions which would be required for Network Rail, as infrastructure manager, to introduce a charge using either of the options considered.

TABLE 5.2 COMPARISON OF AIRLINE SLOT TRADING AND SCARCITY CHARGING

Feature of charge	Airlines	Railway infrastructure manager	
		Value included	Value excluded
Location of congestion		Infrastructure declared congested CUI above a threshold Industry consultation or consensus	
Period of congestion		CUI above a threshold Industry consultation or consensus	
Measure of scarcity	Lack of free slots when required	Function of CUI: <ul style="list-style-type: none"> ■ 1 if CUI is 100% ■ Declines with CUI ■ 0 if CUI is below a threshold 	Function of CUI: <ul style="list-style-type: none"> ■ 1 if CUI is 100% ■ Declines with CUI ■ 0 if CUI is below a threshold
Means of valuing	Mutual agreement	Industry models	To be determined
Measure of value		Lowest value in use during period of a: <ul style="list-style-type: none"> ■ Single service ■ Mean diagram 	An administered financial value to proxy the cost of “scarcity”
Units of scarcity charge	£ per slot owned £ per slot leased	£ per path in period	£ per path in period

Note: scarcity charge per path calculated as “measure of scarcity” x “measure of value”

5.11 We describe next the rationale for each element of the options.

The location of congestion

5.12 In the airline industry, airlines will already know the location of the airport at which they wish to buy and sell capacity.

5.13 For the rail industry, we considered that the location of congestion might be based on one or more of three measures:

- Infrastructure declared to be congested under Directive 2001/14 Article 22
- Infrastructure with a high capacity utilisation, as measured by Network Rail’s CUI or some similar measure
- Infrastructure which industry participants agreed was congested and should be subject to scarcity charging

Infrastructure declared to be congested

5.14 The capacity allocation procedures in Great Britain, a two-stage approach of awarding of rights followed by detailed timetabling of those rights which we describe in Section 8, are designed so that the declaration of infrastructure to be congested under Directive 2001/14 Article 22 should not normally occur.

- 5.15 Nonetheless, Network Rail’s 2015 Network Statement identifies (4.4.3) two areas of congested infrastructure, and NERA (2007) identified (2.5.1) a third, as summarised in Table 5.3.

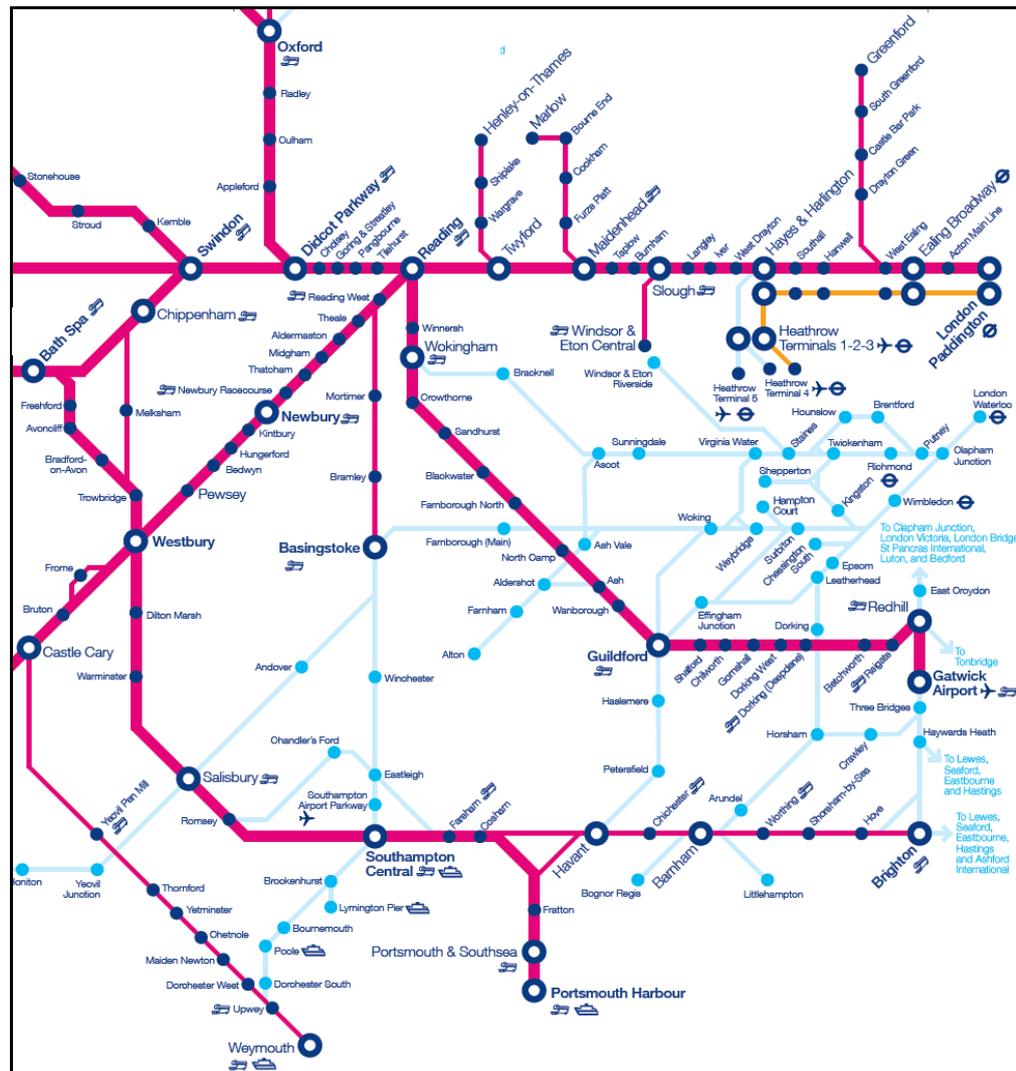
TABLE 5.3 NETWORK RAIL INFRASTRUCTURE DECLARED TO BE CONGESTED

Location	Comments	NS 2008	NS 2015
Gospel Oak to Barking	Approximately 22 kilometres of two-track unelectrified line connecting radial lines in inner north and east London. Declared congested in 2006.	✓	✗
Reading to Gatwick Airport	Approximately 85 kilometres of two track and four track line, all electrified at 750V DC, connecting the nodes of Reading, Wokingham, Guildford, Redhill and Gatwick Airport.	✓	✓
Glasgow and Southwestern Route	Approximately 160 kilometres of single and two track unelectrified lines connecting Troon and Ayr with the West Coast Main Line at Gretna Junction.	✓	✓

Note: these routes are illustrative: we have not examined all Network Statements

- 5.16 These sections of congested infrastructure were defined in the past, and their selection and boundaries might have been different if scarcity charging were introduced, but their characteristics illustrate a number of potential issues.
- 5.17 Nonetheless, Reading to Gatwick Airport, or on the Glasgow and Southwestern Route, may have no localised “bottleneck”, merely an inability to provide additional contiguous end to end train paths.
- Infrastructure declared to be congested: Reading to Gatwick Airport***
- 5.18 To illustrate this point, Figure 5.1 shows the location of the infrastructure between Reading and Gatwick Airport.
- 5.19 One approach to scarcity charging might be to declare the entire route from “Reading to Gatwick Airport” to be “the element of infrastructure” referred to in Article 7.4. However, if the entire section were declared to be congested, it is not clear whether all trains entering it would or should be charged a scarcity charge.

FIGURE 5.1 READING TO GATWICK AIRPORT “CONGESTED INFRASTRUCTURE”



- 5.20 At Reading, the lines towards Wokingham and onwards towards Gatwick Airport are connected both to terminating platforms and to through platforms on the Great Western Main Line, used by services from London Paddington via Oxford, Swindon and Newbury to the Midlands, Wales and the South West. Which, if any, of services using the latter platforms should pay the scarcity charge?
- 5.21 Between Reading and Wokingham, services to Gatwick Airport share lines with services via Bracknell to London Waterloo. Which, if any, of the latter services should pay the scarcity charge?
- 5.22 Between Ash and Guildford, services to Gatwick Airport share lines with services from Guildford via Aldershot to Ascot. Which, if any, of the latter services should pay the scarcity charge?
- 5.23 At Guildford, services to Gatwick Airport share lines with all services between London Waterloo and Havant and Portsmouth. Which, if any, of the latter services should pay the scarcity charge?

Final Report

5.24 Between Redhill and Gatwick Airport, services to Gatwick Airport share the slow lines with services between Brighton and London Victoria, London Bridge and Bedford. Which, if any, of the latter services should pay the scarcity charge?

5.25 This suggests that defining the **orbital infrastructure** from “Reading to Gatwick Airport” as congested might mean applying scarcity charges to all **radial services** between London and:

- Reading, the West Midlands (via Oxford), South Wales and the South West
- Reading via Bracknell
- Portsmouth and Havant via Guildford
- Brighton via Gatwick Airport

5.26 We return to this issue in Section 6 (7.15) but note that there might be a need, if scarcity charging were introduced, to define the “congested infrastructure” extremely carefully.

5.27 Our principal conclusion is that the examples of infrastructure which has been declared congested in the past provide no evidence that “congested infrastructure” will always relate to a more clearly-defined bottleneck such as the Strömbroarna in Stockholm (see Appendix C6.12).

Infrastructure with a high capacity utilisation

5.28 Scarcity charging might also be introduced on infrastructure with a high capacity utilisation as measured by Network Rail’s CUI or some other measure. We have not carried out any examination of CUI across the network but note that it might in principle be high even where there is little traffic and/or no credible scope for entry by another operator. CUI might, for example, be high:

- Approaching London Fenchurch Street, where there is only one operator
- On a single track line in a rural area, on which the spacing of loops had been set so that each section of track was nearly always occupied by a train

5.29 It might not be appropriate to apply scarcity charges purely on the basis of CUI.

Infrastructure which industry participants agreed was congested

5.30 We therefore suggest that the locations of scarcity charging might need to be subject to some form of industry review, and limited to locations where there was a consensus that, as a minimum, there was credible demand for further entry.

5.31 However this would mean that scarcity charges, unlike the capacity charge, would be based at least in part on a purely administrative processes rather than one determined solely by reference to an objective determination of the location of congested infrastructure.

The period(s) of congestion

5.32 In the airline industry, airlines will know the times at which no free slots are available.

5.33 For the rail industry, as with the definition of the location of congestion, the definition of the period of congestion might need to be both based on an objective measure such as CUI and subject to moderation, perhaps by industry consensus, to

take into account the existence of credible demand for entry. There might, for example, be little value in applying scarcity charging at any time at London Fenchurch Street.

- 5.34 We also envisage that the exact start and end times of charging periods could be seen as arbitrary and have disproportionate effect on operators timetabled to be just inside a charging period. As we discuss further in Section 6 (7.19), possible mitigating measures would be to set the charges very low, or to vary them smoothly over time, perhaps by using a moving hourly average CUI.

The measure of scarcity

- 5.35 In the airline industry, no measure of scarcity is necessary, because the market price for a slot will always fall to zero if effectively identical slots are freely available.
- 5.36 In the rail industry, however, there would be a need for a measure of scarcity independent of the value which operators were actually deriving from their use of capacity. We have assumed that this measure would need to be a function of a measure such as CUI which:
- Is 1 if CUI is 100%
 - Declines with CUI
 - Is 0 if CUI is below a threshold
- 5.37 It would not be practicable to say that the measure of scarcity was zero unless the CUI was 100%, both because few or no sections of track operate at such high CUI, and because it would mean that scarcity charging would not take effect until the infrastructure capacity was exhausted. We note that:
- Too rapid a fall in the scarcity charge with CUI might not be consistent with ORR's vision of fairness, stability and predictability (see Appendix B10)
 - Too slow a fall in the charge with CUI would blunt any relationship between the scarcity charge and scarcity, making the charge increasingly proportional to CUI

The measure of value

- 5.38 In the airline industry, no measure of value or cost is needed, as the price for the sale or lease of a slot is negotiated bilaterally and settled by mutual agreement, taking into account whatever conditions and forecasts the parties consider relevant.
- 5.39 For the rail industry, we have assumed that the measure of value for a scarcity charge should be based on the lowest value in use of the current operators within the charging period. This could be calculated either:
- On a gross basis, examining the revenue, costs and external effects directly attributable to the services concerned in the current timetable
 - On a net basis, by the more complex process of removing individual services from the current timetable and observing the overall effect

Final Report

- 5.40 A potentially more difficult problem, to which we have already referred, is identifying the “unit of use” of the infrastructure, which could in principle range from an arc between two adjacent stations to an all-day regular interval service.
- 5.41 We do not consider it realistic to consider an arc as a unit of consumption, for a number of reasons:
- It could result in anomalies where stations lie in the charging section. At Welwyn, for example, it could mean that trains calling at Welwyn North station would incur two charges whereas those passing non-stop through the constraint would incur only one. (It could be argued that this would reflect their relative costs at this particular location, but this might not be the case at others.)
 - It would rarely if ever be possible for an operator to remove a single arc from a service, which would typically require it to leave a train in one station and continue with another train waiting at the next one.
- 5.42 Equally, we do not consider it realistic to consider an all-day regular interval service as the unit of consumption, although exceptionally an open access operator might only be interested in operating on this basis.
- 5.43 A short-distance operator might reasonably argue that, if it removed one service in response to a scarcity charge it would remove an entire round trip, perhaps resulting in a train being unused for a period of two hours. However, it would not be possible to determine, ex ante, whether the round trip removed would include the preceding or the following service (see 2.14). To derive a value it might be necessary to perform the calculation for both services and take the lower value.
- 5.44 A long-distance operator, in contrast, might argue that removal of one service would result in the removal of another, which might not return to the same point, and effectively render an entire daily diagram or even a whole service uneconomic.
- 5.45 We provisionally conclude that an estimate of value in use of existing services would need to be based on either:
- A single service.
 - A whole diagram, if comprehensive information on operators’ diagrams were available to Network Rail. In the case of open diagrams, the value would perhaps need to be based on the mean daily value of the entire circuit. Rules might also be needed for dealing with split diagrams, where morning and afternoon workings are effectively independent and could be combined into different diagrams which, by implication, would have different values.
- 5.46 The measure of value would therefore be the lowest value in use of all the trains passing through the charging location, measured either over a single service or over the whole (mean) diagram.

The calculation process

- 5.47 The calculation of the scarcity charge for any given charging location and period of charging would be expressed in £ per train, calculated as the product of:
- The function of CUI, or a similar measure (5.36)

- The lowest value in use of all trains in the location and period (5.46)
- 5.48 We note that basing value on a single service would result in a relatively low charge, and that basing value on an entire diagram, if data were available, might result in a charge:
 - 2-3 times higher, if all the operators at the charging location were long-distance operators
 - 10-20 times higher, if all the operators at the charging location were short-distance operators
- 5.49 This suggests that the calculated level of lowest value in use, as a proxy for market-clearing price, as a proxy for opportunity cost at any given charging location, might vary over a wide range depending on the assumed minimum element of existing services which could be removed.
- 5.50 As an alternative to a detailed calculation of value in use, which seems likely to be subject to wide variation based on largely administrative decisions, we therefore also considered the option of excluding any direct measure of value and instead nominating a fixed price, which might be the same for each charging location. We summarise the precedents in Table 5.4.

TABLE 5.4 PRECEDENTS FOR ADMINISTRATIVELY SET CHARGES

Network	Charge	Appendix reference
ÖBB Incentive for capacity optimisation	€0.5638/train-kilometre	C3.2
ÖBB Supplement for congested infrastructure	€1.2173/train-kilometre	C3.10
Infrabel “H” and “T” parameters	Complex and multiplicative	C4.3
DB Netz former Auslastungsfaktor	20% supplement	C5.3
Trafikverket Passage Charge	SEK 260	C6.3

- 5.51 This suggests that one approach to introducing scarcity charges might be to introduce them at a notional level of, for example, £10 per train and monitor the market response and effect on behaviour. For a commuter operator sending 20 trains per hours through a charging section in one-hour weekday AM and PM peaks, a charge of £10 per train would result in an annual charge of around:
 - £50,000 per peak per direction
 - £100,000 if charged in one direction in both peaks
 - £200,000 if incurred in both directions in both peaks

Practical issues

- 5.52 In addition to the basic calculation of the scarcity charge it might also be necessary to address a number of practical issues. We discuss in turn below:
 - Trains passing through multiple charging locations
 - Recalculating charges

- Revising charges to operators
- Charges for freight services
- Revenue neutrality
- Transaction costs

Trains passing through multiple charging locations

- 5.53 A train passing through multiple charging locations might pay multiple charges.
- 5.54 Multiple charges apply in Sweden. It would in principle be possible for a train from Stockholm via Göteborg to Malmö and Copenhagen to pass through all five charging sections in one journey (see Appendix C6), although it would not be possible for it to be in all five during the charging periods 07:00-09:00 and 16:00-19:00.
- 5.55 We see no prima facie reason why a train passing through multiple charging locations should not pay multiple charges, particularly if removing it from all of them could, in principle, allow the operation of a number of services, with higher aggregate value, at some or all of them. However, we do not rule out the possibility of more detailed research revealing situations in which this might lead to anomalies.

Recalculating charges

- 5.56 We assume that operators applying for access would make their decisions on the basis of the scarcity charges in force at the time of their application. These would not necessarily be the same as the scarcity charges which would apply when their services began, because scarcity charges would need to be updated from time to time to reflect changes in the scarcity and value of capacity.
- 5.57 In principle, scarcity charges would need to be recalculated every time there was a change in either:
- Infrastructure capacity, and hence capacity utilisation
 - Timetable, and hence lowest value in use, if included in the measure
- 5.58 Without detailed examination, however, we cannot predict how volatile charges calculated on this basis would prove to be.
- 5.59 It would be possible to reduce any volatility, and to save calculation effort, by recalculating the charges less frequently, such as once for each Control Period. There would, however, be a risk that the relation between the charges and scarcity declined over time, particularly if:
- New infrastructure were opened during the Control Period, affecting and probably reducing the CUI even if there had been no change in services.
 - Services were added or removed during the control period, potentially raising or lowering the CUI respectively.
- 5.60 We provisionally suggest that the charges should be recalculated with each change in CUI, which would normally mean at each timetable change and whenever the infrastructure was modified.

Revising charges to operators

- 5.61 We assume that operators holding access rights would continue to pay scarcity charges at the current rate. If not, rising demand for capacity would not be signalled to operators through higher charges. Paying scarcity charges at the current rate would, in principle, provide operators an incentive for them to reroute, retime or surrender rights which rendered access unaffordable, although the extent to which franchised passenger operators could do so would depend on the flexibility in their franchise agreement.

Charges for freight services

- 5.62 Charges based on a measure of value in use for passenger services would not reflect the potential value of freight services, for which there are no equivalent models, or information on their value in final markets. The value in use of freight services might be lower and hence in principle imply that the lowest value in use, as a proxy for market-clearing price, as a proxy for opportunity cost, should also be lower. In principle, however, it would be possible to apply charges based on passenger value in use to all operators.
- 5.63 However, it is not clear how scarcity charges would be applied to paths which were deemed to be necessary for potential seasonal or intermittent freight services but in the event neither requested nor used by any operator. The principal choices appear to be:
- To continue to create paths for potential future freight traffic, irrespective of unsatisfied demand for passenger paths, but accept that, on the day, they may not be used or paid for.
 - To limit timetabling to paths which have been requested by operators willing to pay any necessary scarcity charge, and if necessary accept that freight services will not be possible at certain locations in certain periods.

Revenue neutrality

- 5.64 It would theoretically be possible for a scarcity charge not to be revenue-neutral, but this would potentially create incentives for Network Rail to allow scarcity charges to rise, even where additional services could be operated or capacity could cost-effectively be expanded. This fundamental issue, that a monopoly provider's rational response to a capacity shortage is to increase charges, and extract monopoly rents, rather than to increase capacity, was identified in early theoretical studies of a market-led railway. If the charge were not revenue-neutral, it might still be possible to ensure that the revenue from it was used in particular ways, such as:
- Ring-fenced as funds to be spent on capacity enhancement projects, reducing the net cost to be recovered from their users via the RAB.
 - Used to repay Network Rail's debt.
- 5.65 NERA (2007) assumed that it would be necessary to ensure that any reservation charge was revenue-neutral to operators as a whole, or even to freight operators as a group, but noted that revenue-neutral charges would still mean redistribution between operators, which we discuss in Section 6. CEPA (2010) referred to the "no

net loss no net gain” Clause 18.1 provision in franchise agreement, which ensures that changes in charges to franchisees are passed through to the funding authority.

5.66 ORR suggested that any scarcity charge would actually be revenue neutral because legislation requires Network Rail’s revenues to balance its costs. However this raises a number of issues, including whether scarcity charges are controlled to be revenue-neutral over all operations or over defined subsets of them. We discuss a number of these issues below, assuming illustratively that operators might be subdivided into charging groups or “pools” such as by:

- Franchised passenger, open access passenger or freight
- By Network Rail route

5.67 First, it would be necessary to decide how the net cost of scarcity charges would be returned to operators in any given charging group. Franchised passenger operators could be offered a percentage rebate on the Fixed Track Access Charge (FTAC), but this is not paid by freight or open access operators. One alternative would be to pay a rebate pro rata with some other measure, such as total train-miles operated during the relevant charging period.

5.68 Second, if a large proportion of all rail services remained funded by the Department for Transport, there is a risk that it would receive, in aggregate, a rebate almost equal to the scarcity charge for its services, blunting the incentives to it to avoid times of high opportunity cost. There are, however, at least two potential means by which the effect might decline:

- Increased funding of services by separate local authorities, each of which would pay only a small part of the total scarcity charges, and hence retain strong incentives to reduce their share of them. In this context, we note that it would probably be appropriate to make the scarcity charge revenue-neutral over a national “pool” of all operations, rather than within an individual Network Rail route, each of which might have a single or dominant operator or funder.
- Excluding scarcity charges from the “no net loss no net gain” Clause 18.1 provision in franchise agreement. However, bidders for franchisees might only accept this arrangement in franchises where they had considerable additional flexibility to modify services to minimise the incidence of the scarcity charge. A further possible subtlety would be for franchising authorities to indemnify franchisees against the level of scarcity charges for any given service (ΔP), but make them liable for changes in scarcity charges resulting from adding or removing services (ΔQ). This would indemnify franchisees against changes in their initial commitment, but ensure that they took scarcity charges into account in making any changes to their timetable within the terms of their franchise agreement.

5.69 However, both these approaches ultimately depend on the Department for Transport, either to ensure that more services are devolved or to ensure that franchisees have sufficient flexibility to avoid scarcity charges that they are willing to be responsible for paying them.

5.70 More widely, best practice would be for the scarcity charge to be borne by the party best able to manage it, but whether that was the Department for Transport

or the franchisee would depend on the franchise agreement, and might vary between franchises or even within a franchise.

- 5.71 Third, and whatever approach was adopted, high scarcity charges might lead to a situation in which some operators faced negative charges. If, for example, the average of all charges across 100 services in a “pool” were £20, and 50 of the services were required to pay a scarcity charge of £50, then redistributing the scarcity charge equally across all users would result in a charge of £45 for half the users and -£5 for the other half. (This issue cannot arise with airport slot trading because the premium for scarce slots is paid to the selling airline rather than into the regulated income of the airport.)
- 5.72 It is not clear whether this would be an efficient outcome or would result in perverse incentives. It might be possible to minimise or avoid this effect if scarcity charges were kept sufficiently small, but the materiality of the issue could only be determined once actual charges had been developed and their size relative to current charges was apparent.

Transaction costs

- 5.73 NERA (2007) noted, in the context of a reservation charge, that any new charge would result in transaction costs, such as for the development and operation of Network Rail’s billing systems. We have not examined the administrative implications of charging but note that it would be desirable to avoid recalculating the charges, for up to 25,000 trains per weekday plus weekend services, on up to 6,688 CTSs, every time there was a network or timetable change.
- 5.74 This suggests that the sequence for calculation charges would be:
- First, to examine sections or route, or nominated locations, by time period and identify as “charging locations” only those with a sufficiently high capacity utilisation to justify a scarcity charge.
 - Second, to calculate the value only of those trains using sections of route or nominated locations on which a scarcity charge would apply.
- 5.75 The aim would be to reduce the required number of calculations by several orders of magnitude to a manageable number.
- 5.76 Illustratively, we estimate that a full analysis of the net (rather than the gross) value of all weekday trains over the Welwyn viaduct on one route, the East Coast Main Line (ECML), would require around 500 model runs, or over 1000 including weekend services.

Summary

- 5.77 Scarcity charging could not directly reflect opportunity cost in the way that is achieved by bilateral trading for airport slots. It would, at best, be a proxy for market-clearing price, which in turn would be a proxy for opportunity cost.
- 5.78 A scarcity charge could be calculated which would, in principle, incentivise existing operators to reroute or retime their services where this was possible, but if correctly set at the lowest value in use might be insufficient to incentivise withdrawal. It would also deter applications with low value in use.

Final Report

- 5.79 Even a “robust” calculation based on lowest value in use might result in a range of around ten to one in the level of scarcity charges, depending on the definition of the minimum element of service which could be withdrawn.
- 5.80 It might be simpler to set a scarcity charge by administrative processes, as some other networks have done. This would be broadly similar to the approaches adopted by ÖBB and Trafikverket listed in Table 5.4.
- 5.81 A charge of £10 per train applied only in two weekday peak hours could result in charges of £200,000 per annum to large London commuter operators, who might not be permitted, or find it commercially viable, to reduce their peak services.
- 5.82 To send up-to-date signals, charges would need to be updated regularly and probably at least annually.
- 5.83 It might be desirable or necessary for scarcity charges to be revenue-neutral, but while the Department of Transport remains the dominant buyer of rail capacity this would blunt the incentives to both it and franchise operators, unless they were no longer able to pass through charges to their funder. They might only be willing to bear exposure to charges if they also had the flexibility to withdraw peak period services, particularly on London commuter routes.

6 Primary effects of scarcity charging

Introduction

- 6.1 In this Section we discuss the extent to which a scarcity charge would send signals and incentivise operators for the efficient use of capacity, as envisaged in economic theory. We focus on the primary or intended effects of influencing operator behaviour, including as far as possible the likely behavioural response of different types of operator. We end this discussion with the effects on industry workload and a summary of a possible illustrative outcome.
- 6.2 We note NERA’s view (see Appendix B9.1) that “It is not clear that there are widespread problems of poor capacity utilisation caused by train operators’ decisions.” Nonetheless, scarcity charging might act as an incentive:
- For existing operators to reroute, retime or withdraw trains (see Table 2.2)
 - For potential operators to withdraw their applications

The scale and timing of primary effects

- 6.3 The scale and timing of the primary effects would depend on the factors we summarise in Table 6.1.

TABLE 6.1 DETERMINANTS OF THE EFFECTS OF CHARGES ON OPERATORS

Factor	Determined by
The level and structure of charges International experience suggests that the low scarcity-related charges applied to date have no detectable effect, except possibly in applications for freight paths during passenger peaks	ORR
Funder or operator assessment of the demand, in final markets, for change In some networks, stability appears to be preferred to innovation and dynamism	Funder or operator assessment of passenger and freight market preferences
The flexibility given to operators by their funders, and the balance of risk between them	Funders, subject to operator willingness to accept a higher risk
The flexibility in the timetable to accommodate unilateral changes to timing and route	Historic capacity utilisation, subject to future ORR intervention
The timescales over which operators are able to react to the availability of released capacity	Operators ROSCOs and manufacturers

- 6.4 We assume in this Section that charges, whether set by reference to the value of a service or diagram (see 5.45) or by administrative means (see 5.50) were set sufficiently high to influence operator behaviour, would be sufficiently high to influence operator behaviour.

Final Report

- 6.5 We also include consideration of the sequence and timescales by which scarcity charging could influence operator behaviour and induce services changes. Where infrastructure is already constrained, it may be possible for further services to be accommodated, but it seems prima facie more likely that some services may need to be withdrawn before others are added. This raises the issue of how rapidly, if at all, services removed in response to the scarcity charge will be replaced.
- 6.6 In the airline industry, the “use it or lose it” provisions mean that unused slots are returned to the pool and hence the airline holding them loses the rights associated with holding them. An airline is unlikely to withdraw a service until it has found another airline willing to buy, or lease, the slot.
- 6.7 For the rail industry, in contrast, a mechanism based on scarcity charges rather than bilateral trades means that removing one service may not lead to replacement by another:
- There may be no demand to operate a higher value service at the same time
 - There may be demand, but insufficient contiguous capacity to enable a viable service to be introduced
 - There may be demand, and contiguous capacity, but the operator may need time to lease or buy rolling stock
- 6.8 Where infrastructure is congested, the removal of one service might lead to the immediate introduction of another. There is, however, the risk that a new service does not appear until rolling stock can be procured (which can take 2-3 years) or even at all, if the capacity released is of no commercial value to other operators.
- 6.9 ORR asked us not to assume that tight specification of franchise agreements, or the current mix of franchise and other operators, were a permanent feature of the railway, and hence to consider how scarcity charges might incentivise efficiency with a different mix of operators. More flexible franchises might increase the “churn” in the timetable if both:
- Patterns of demand become more volatile and franchisees conclude that there is a market requirement for more extensive or frequent change.
 - The timetable becomes sufficiently liquid that frequent change is possible.
- 6.10 Illustratively, we noted elsewhere in this study that:
- At Heathrow airport, around 5% of take-off and landing slots are transferred between airlines each year (see Appendix D2.26)
 - In Belgium, infrastructure manager Infrabel informed us that timetable of the sole domestic passenger operator, SNCB, has remained stable over long periods
- 6.11 Taking into account the likely need to services to be removed before new ones are added, and the potential for delay while this takes place, we discuss in turn the potential impacts of scarcity charges on:
- Non-timetabled services
 - Freight operators
 - Charter operators
 - Open access passenger operators

- Franchised passenger operators, which form the majority of current services

The effect on non-timetabled services

- 6.12 Network Rail is subject to a Network Code objective (see Table 8.1) “To share capacity on the Network for the safe carriage of passengers and goods in the most efficient and economical manner in the overall interest of current and prospective users and providers of railway services”.
- 6.13 While some prospective users may make formal applications for timetabled capacity, Network Rail must also keep capacity free for a number of non-timetabled services such as:
- Moving empty stock to and from the points where it provides a service
 - Taking stock to and from repair and overhaul
 - Delivery of new and removal of old stock
 - “Departmental” or engineering trains operated for Network Rail or its contractors.
- 6.14 While provision for all these operations is likely to be needed, it cannot necessarily be requested or paid for in advance by an identifiable user. In Portugal, for example, infrastructure manager REFER (see Appendix C8.4):
- Is not obliged to allocate all the capacity available
 - Reserves the right to hold back capacity for foreseeable ad hoc requests
- 6.15 We assume that, with a system of scarcity charging, Network Rail would continue to reserve capacity for non-timetabled services, but operators making use of this capacity would be required to pay any appropriate scarcity charge. If, over time, this depressed the usage of peak capacity reserved for non-timetabled services, then Network Rail could reduce the overall scale of provision made for them.

The effect on freight operators

- 6.16 Many freight trains operate with wagons dedicated to a particular commodity, such as coal, iron ore or petroleum, and run loaded in one direction and empty in the return direction. An exception is multimodal trains which offer a scheduled and “bookable” service of transporting containers from point to point, such as between a port and an inland terminal. The trains may run loaded on all scheduled journeys, although the containers they carry may be either loaded or empty.
- 6.17 The nature of the rail freight market means that Network Rail, in common with other rail infrastructure managers, needs to reserve paths for potential operators in advance of any firm commitment that any operator will either request the path in any particular future period or use it on any particular day.
- 6.18 Freight operators catering for seasonal, intermittent or occasional demand by their final customers may not make full use of paths allocated to them. NERA (2007) reported data from the March 2007 Freight Route Utilisation Strategy (RUS) showing that the take-up of domestic paths in 2004-5 varied between 95% for intermodal services and 37% for construction.

6.19 We assume that freight operators' response to scarcity charging would be:

- To take up fewer of any peak freight paths already timetabled
- To request fewer freight paths at times at which scarcity charges apply

6.20 We envisage that this would result in a net reduction in freight operations at locations and times where scarcity charging is applied.

The effect on charter passenger operators

6.21 Charter operators operate only a relatively small number of services. They typically wish to provide a tour, or out-and-back service, using atypical rolling stock, but rarely specifically require capacity at peak periods, or particular station calls, and may not impose pressure to minimise journey times. In this respect their requirements and flexibility may be similar to those of freight operators.

6.22 While charter operators do not typically require capacity at peak periods, scarcity charging might result in them avoiding locations and times where scarcity charging applied.

The effect on open access passenger operators

6.23 Open access passenger operators in Great Britain have to date focused on long-distance services serving particular destinations or markets not well served by franchised services. We note that different business models may emerge over time, particularly if scarcity charging makes it easier to obtain the capacity required. However, while open access services operate on a relatively small scale, having a train set idle for even a few hours may represent a material inefficiency, and their commercial viability may be sensitive to their ability to obtain diagrams allowing intensive use of a small fleet.

6.24 On balance, the net effect of scarcity charging on open access operators is uncertain and would depend on the relative importance of a number of factors, including their ability to pay, or avoid, scarcity charges, the extent of capacity becoming available, and the extent to which it allows them to operate efficiently.

The effect on franchised passenger operators

6.25 Franchised passenger operators are currently the dominant or only operator on large parts of the network, including the locations listed in Table 6.2. The degree of flexibility granted to franchisees to modify their timetables is in the gift of the specifying authority. We assume that, where they were permitted to do so, franchise operators would respond to sufficiently high charges by removing low value services and, potentially by introducing higher value ones.

6.26 In considering the potential effect of scarcity charging on franchise operators, we have taken into account analysis which suggests that scarcity charges might have unexpected effects at charging locations where there is a dominant or sole operator. We set out our thinking below.

Dominant or sole operators

6.27 If a consumer buys quantity Q of a product at price P , then the total amount is paid is PQ . With smooth demand and supply curves, and liquidity in the market

(see 1.9) and market-clearing prices (see 3.9), if the consumer buys one fewer unit of output, reducing quantity Q by 1, the price P will remain almost unchanged and the marginal change in PQ will be P , the unit price. One example is in foreign exchange markets, where an individual purchase of a small amount of currency will have no detectable effect on the market rate.

- 6.28 However, these idealised conditions may not apply with railway infrastructure, because each element of infrastructure is used by few operators, or often by only one:
- No single section of infrastructure handles more than 30 trains per hour per direction. A more common level of use at peak times may be in the range 10-20 trains per hour per direction (for an example, see Appendix D4), but most of these may be exercising contractual rights. This means that the scope for entry or exit at any one location in the short to medium term may be limited to a small number of paths.
 - On some sections of infrastructure, all, or almost all, operations are carried out by a single operator.
- 6.29 If only one operator could credibly use the infrastructure then, given that P would also be a function of capacity utilisation, (see 5.36), it would find that if rerouted, retimed or withdrew one service then both Q and P would fall, with the potential effect that the marginal saving would be many times the original price P . The effective marginal price, as seen by the operator, would be the ratio $\Delta QP/\Delta Q$.
- 6.30 Assuming, as an illustrative example, that:
- A section of infrastructure can handle 20 trains per hour at peak times
 - A single operator is currently operating 19 trains per hour
 - There appears to be no prospect of new entry
 - The scarcity charge has been set at £5 on the basis of the current CUI but would rise to £10 on the basis of the CUI at 20 trains per hour
- 6.31 In a liquid market, an operator would know that, if it added a train, it would cost an additional £5, and it would base its decision on the knowledge that this would be the marginal charge. A dominant operator, in contrast, would be able to predict that if it added one train its total charge would rise from £95 to £200, an increase of £105, or 21 times the intended marginal charge. The effective marginal charge, as seen by the dominant operator, would not be P but $\Delta QP/\Delta Q$.
- 6.32 Examination of the national rail network suggests that a large number of potentially-constrained areas are currently used by only one operator, as we set out in Table 6.2. If scarcity charges applied at any of these locations, and these operators both were liable for scarcity charges and had flexibility to reroute, retime or withdraw their services, then they might be incentivised to manage down their calculated CUI, or any other measure used as a basis for charging, with the aim of reducing their scarcity charges.

TABLE 6.2 INFRASTRUCTURE WITH ONLY ONE CREDIBLE OPERATOR

Potentially-constrained area	Operator
Glasgow Queen Street and approaches	ScotRail
Liverpool low level stations and approaches	Merseyrail
Waterloo and approaches	South West Trains
Future Crossrail core	Crossrail operator
Marylebone and approaches	Chiltern
Future Thameslink core	Thameslink operator
St Pancras high level and approaches	East Midlands Trains
Liverpool Street and approaches	Anglia
Fenchurch Street and approaches	C2C

Source: Steer Davies Gleave research, conclusions illustrative

- 6.33 This effect would also apply at any other point on the network where a dominant operator could be reasonably confident that a reduction in its services would not result in the introduction of replacement services by other operators.
- 6.34 All the operators listed in Table 6.2 are or will be franchise operators, and they may be prevented from their franchise agreements from reducing services at peak periods. However, if and where they had such flexibility, they might reduce their scarcity charges considerably by cutting services. This would be the case even if there were both no saving to them in rolling stock and considerable disbenefit to passengers through extended service intervals or greater crowding. This would probably not be efficient.
- 6.35 A further consequence of this effect would be the effect, on any dominant operator, of even limited new entry. Turning to the example described in paragraph 6.30, a future franchisee might contract to provide a 19-train service on the basis that the scarcity charges would be £95. Unexpected entry by another operator providing one more train, at a cost to the new entrant of £10, would raise the franchisee’s charge by £95 to £190. Conversely, withdrawal of a minority operator could result in a windfall benefit to the majority operator. This effect might not, prima facie, be consistent with ORR’s vision that charges should provide fairness, stability and predictability.
- 6.36 It could be argued that there would be no rationale for scarcity charges where there was a single operator, and we agree that this is the case, but the underlying principle is that the network is open to both entry and exit, and a sole operator might at any time be joined by others, who need not be competing in the same final markets to use the same infrastructure.
- 6.37 In summary, for a dominant or sole operator, a change in levels service level would change not only the number of charges paid (ΔQ) but also the unit charge (ΔP).

- 6.38 One approach to mitigating this effect would be to ensure that scarcity charges did not fall rapidly with CUI (or some other measure). If this were the case, however, they would become increasingly close to proportional to CUI, reducing the intended linkage to the incidence of scarcity.
- 6.39 A second approach to mitigating the effect would be for funders to indemnify franchisees against changes in scarcity charges, or to hold the rates of scarcity charge for each operator at the level in force when they were awarded access rights. This would, however, have a number of effects:
- Operators might face charges set up to 15 years earlier, bearing no relation to current levels of scarcity, as we noted in paragraph 5.61.
 - Operators applying for equivalent scarce capacity at different times might face difference scarcity charges.
 - Operators with access rights subject to an historic and high scarcity charge might have an incentive to cancel them with the aim of reapplying for them with a lower, or no, charge.
- 6.40 The most effective means of mitigating the effect would be to avoid situations with a dominant or sole operator. This could in principle be achieved if existing franchises were subdivided into smaller packages, but would require action by the funding authorities to reverse the trend, since initial franchising, to less direct competition between franchisees (see Appendix B7).

Summary of effect on franchised passenger operators

- 6.41 If franchised passenger operators were given greater flexibility to do so by their funders, we would expect them to respond to scarcity charging by rerouting, retiming or withdrawing services. However, whether they were the dominant or sole operator in relatively illiquid markets, the effective marginal scarcity charge might be many times that seen by other operators. A possible consequence is that franchised operators would be over-incentivised to remove services where scarcity charging applied, even if there were disbenefits to their passengers and/or no realistic prospect of new entry to make use of any capacity released.
- 6.42 A potential offsetting effect is that franchised operators might introduce new services of higher value.

The effect on workload

- 6.43 Scarcity charging would, in principle, have two effects on the workload of timetabling and capacity allocation:
- An increase in the workload associated with dealing with rerouting, retiming or withdrawing existing services.
 - A decrease in the workload associated with dealing with new applicants deterred by prior knowledge of the scarcity charge.
- 6.44 The overall effect on workload is unclear, but prima facie a more dynamic market, if that were the outcome, would result in a higher workload.

Summary

6.45 Table 6.3, which we stress is purely illustrative, summarises the possible effects of scarcity charging.

TABLE 6.3 POSSIBLE OPERATOR RESPONSES TO SCARCITY CHARGING

Market	Possible withdrawal or retiming	Possible new services
Non-timetabled	Take-up of reserved capacity declines, so less capacity needs to be reserved	Replacement or retimed services outside charging periods
Freight	Take-up of reserved capacity declines, so less capacity needs to be reserved Less demand during charging periods	Replacement or retimed services outside charging periods
Charter	Likely to avoid scarcity charging, given the nature of charter services	
Open access	Uncertain effect, depending on business models of open access operators	
Franchised	Removal or retiming of lower value services during charging periods	Higher value services, or no replacement services

Note: assumptions illustrative, see text for details

6.46 We provisionally conclude that the primary effects of scarcity charging would be:

- Rebalancing services from peak to off-peak
- Rebalancing the mix of services offered

6.47 However, dominant or sole operators might have unintentionally high effective marginal scarcity charges, which may over-incentivise franchised passenger operators to withdraw peak services, depending on the terms of their franchise agreements.

7 Secondary effects on operators of scarcity charging

Introduction

- 7.1 We set out in Section 6 our assessment of the primary effects of scarcity charging on operator behaviour, including the potential distortion of incentives where there is a dominant or sole operator, which is prima facie likely to be a franchised operator.
- 7.2 In this Section we discuss a number of other potential secondary and unintended effects, and the scope for mitigation measures, dealing in turn with:
- The mismatch between market-clearing prices and optimal use
 - Distribution between operators
 - The reserve price effect
 - Behaviour at the boundary of charging locations and periods
 - Inefficient allocation of risk between operators and funders
 - Distribution between customers
 - Peak services withdrawn and not replaced
 - Delayed service expansion

The mismatch between market-clearing prices and optimal use

- 7.3 We showed in Table 3.1 how scarcity charges set at the lowest value in use might still exclude services which would form part of the optimal use of capacity and could be included with the use of a relatively simple priority system. If the existing timetable included Express train 1 and Stopping train 1, then:
- The lowest value in use would be £40
 - The highest scarcity charge compatible with the optimum outcome, which would need to include Express train 3, would be £20
- 7.4 This highlights the risk that a scarcity charge set at the lowest value in use might also deter operators from applying to operate lower value services which would or could form part of the optimum package. In summary, a scarcity charge based on an estimate of the lowest value in use, even if calculated on an objective, robust and up-to-date basis:
- May not be high enough to reflect opportunity cost
 - May not be low enough to incentivise applications which might form part of the optimum package
- 7.5 The materiality of the mismatch would ultimately depend on the extent to which scarcity charges diverge from opportunity costs, which our analysis in Section 3 suggest could be large. We note, however, that:
- The root cause of the mismatch is that any predetermined scarcity charge is likely to be a poor estimate of opportunity cost
 - There are no obvious mitigation measures

Distribution between operators

- 7.6 ORR's own vision for the charging structure is that it should provide fairness, stability and predictability (see Appendix B10). Any change to charges is likely to result in a redistribution of charges between operators which may, prima facie, be perceived as unfair or unpredictable.
- 7.7 The materiality of distribution effects cannot be estimated without detailed proposals for the structure and level of scarcity charges, but we note that:
- Freight and passenger open access operators must collectively pay at least as much in charges, because they need to pay the charge but they cannot benefit from a reduction in FTAC, which they do not pay, as a mechanism to ensure revenue neutrality.
 - If freight and passenger open access operators must collectively pay at least as much, and scarcity charges are revenue neutral, then passenger franchises and their funders must collectively pay no more in charges.
 - The distribution of the benefits of scarcity charges between franchised operators and funders, in the long term, implicitly depend on the competitive process of franchise bids. This may mean that funders capture most, if not all, of the benefits.
 - Among franchised services, however, the net effect may be an increase in charges for those users of congested infrastructure and a reduction in charges for those avoiding it.
- 7.8 In summary, the likelihood is that revenue-neutral scarcity charges should reduce charges to franchisees (and/or their funders) not using congested infrastructure, might increase charges to freight and passenger open access operators, and would leave the balance of reductions or increases being borne by franchisees (and/or their funders) using congested infrastructure.
- 7.9 The outcome would ultimately depend not only on the level and structure of charges and the charging locations and periods but also on the ability of operators to modify services in response to the charges. We note, however, that:
- The root cause of the issue of distribution between operators is inherent in the industry structure, in which the obligation or need to use congested infrastructure varies between operators
 - There are no obvious mitigation measures

The reserve price effect

- 7.10 As we note in paragraph 3.19, any charge set in advance is analogous to a reserve price. If the objective of scarcity charges were to ensure that all capacity was allocated, then setting a reserve price would be potentially counterproductive.
- 7.11 At first sight this effect might be addressed through a two-stage process in which applications for capacity were first invited without posting a charge, to gauge unconstrained demand, and a scarcity charge introduced at the second stage in order to moderate demand. However this would require modifications to the current access arrangements to include:

- Fixed periods in which applications for future capacity could be submitted
 - Delays while a timetabling exercise was carried out to see if all demand could be accommodated
 - If needed, publication of a scarcity charge calculated on some basis (which, if not the current timetable, could not be a timetable including all applications)
 - Delays while applicants decided whether and how to respond to the proposed charge
- 7.12 In practice, and even if the inflexibility, delays and costs of such a process were acceptable, operators able to observe the process over a number of years might still be deterred from applying if their experience was that it normally resulted in a scarcity charge which they could not afford.
- 7.13 The materiality of this effect would ultimately depend on the size of the level of scarcity charges and the extent of any latent demand for capacity. We note, however, that:
- The root cause of the issue is that any predefined charge may deter applications which could be accepted
 - Possible mitigation measures include a two-stage application process, if this proved to be workable and beneficial, which would need to be imposed by ORR

Behaviour at the boundary of charging locations and periods

- 7.14 Scarcity charges applied to particular locations and periods could have unintended effects at the boundaries of those locations and periods.
- 7.15 We set out in paragraph 5.25 how declaration of the orbital Reading to Gatwick Airport as congested infrastructure could, if it resulted in scarcity charges, affect large numbers of radial services on Great Western, South Western and Southern services. Depending on the exact definition of the “congested infrastructure”, even minor changes to the Working Timetable, such as shifting the same train to a different line, might result in a service moving in and out of charges.
- 7.16 If an operator held rights outside the charging location, would they still avoid the charge if their service was transferred to a line inside it? Equally, if an operator held rights inside the charging location, would they still pay the charge if their service was transferred to a line outside it? The decision to transfer a train to a different lines might be made either when planning the Working Timetable for a particular period or tactically in response to disruption.
- 7.17 An operator that had paid a charge to use a charging location might expect to be excused the charge if they did not, although an alternative approach would be for the charge to be payable if the effective level of service provided was the same as if the charging location had been used. This would, however, mean decoupling the application of the charge from the use of particular infrastructure, and we have not reviewed the legal practicalities of such an approach.
- 7.18 Similarly, if an operator held rights outside the charging period, would they avoid the charge if their service was retimed (which might be from 07:59½ to 08:00½)? Equally, if an operator held rights inside the charging period, would they still pay the charge if their service was retimed (which might be from 08:00½ to 07:59½)?

- 7.19 The materiality of the outcome would ultimately depend on when and where scarcity charges would apply and at what level. We note, however, that:
- The root cause of the boundary issue is the potentially arbitrary selection of boundaries in space and time
 - Possible mitigation measures would be to vary scarcity charges smoothly over time, perhaps by using a moving hourly average CUI, and lie mainly with ORR

Changing the balance of risk between operators and funders

- 7.20 As we noted in 5.70, best practice would be for the scarcity charge to be borne by the party best able to manage it, but whether that was the funder or the franchisee would depend on the franchise agreement, and might vary between franchises or even within a franchise.
- 7.21 To date, the practice of the Department for Transport and other franchise funders has normally been to bear the risk of any change in the level of charges faced by operators as a result of Periodic Reviews. The rationale is that franchisees have little power to manage their exposure to charges while subject to prescriptive franchise agreements.
- 7.22 If scarcity charges were introduced, a range of outcomes are possible, of which we focus on two.

Continuation of current arrangements

- 7.23 In this scenario, franchisees would continue to be given only limited flexibility to change timetables to manage their exposure to scarcity charges. In these circumstances franchisees might be unwilling to bear the risk of scarcity charges, which instead would be passed through to the funding authority. Changes to timetables in response to scarcity charges might therefore be a function of central consultation, planning and funding processes. We also note that, if the Department for Transport remained the dominant funder of passenger rail services, a policy that charges should be made revenue-neutral through reductions in FTAC would much dilute the effect of charges on its financial position, as the majority of scarcity charges, or any change in them, would be returned to it, at least in the longer term.

More flexible franchises

- 7.24 Alternatively, franchisees might be given greater flexibility to change timetables, and in particular reroute, retime or withdraw peak period services, and as a result be willing to accept exposure to scarcity charges. In these circumstances funders might partly or wholly exclude scarcity charges from pass-through. A possible outcome would be more dynamic, or volatile, timetables, with franchisees restrained from making changes only by the likely response of their customers.
- 7.25 The outcome would ultimately depend on the policies of funding authorities and the willingness of franchisees to bear risks. We note, however, that:
- The root cause of the issue of aligning risk with ability to manage it is inherent in the repeated decisions, since 1993, that most services are socially necessary and must be protected in franchise agreements
 - Possible mitigation measures lie mainly with the funding authorities

Distribution between customers

- 7.26 Any change to the pattern of services resulting from scarcity charges would be likely to reduce the range or frequency of services available to some passenger and potentially allow subsequent increases to the range of frequency of services available to others. Nonetheless, if scarcity charging proved to be effective as a means of improving allocative efficiency, the benefits to those who gained would exceed the disbenefits to those who lose.
- 7.27 As we note in Section 6 (6.46), however, we provisionally conclude that the primary effects of scarcity charging would be:
- Rebalancing services from peak to off-peak
 - Rebalancing the mix of services offered
- 7.28 The materiality of the outcome would ultimately depend on the extent to which franchisees were free to respond to scarcity charges and the balance of incentives on them to do so. We note, however, that:
- The root cause of the distribution effect is that, if railway capacity is not currently allocated efficiently, then any move to more efficient allocation is likely to redistribute services and create losers.
 - Possible mitigation measures would be either to have tighter franchise specifications, weakening or removing the rationale for scarcity charges, or to restructure franchise payments to reward equity as well as efficiency. We have not attempted to assess how this could be done in practice, but note that any possible mitigation measures would lie mainly with the funding authorities.

Peak services withdrawn and not replaced

- 7.29 A related, but potentially more serious outcome, would be if the effect of scarcity charges was that franchised operators withdrew peak services but that any replacement services were either delayed, perhaps until suitable capacity and rolling stock were available, or never emerged, because no outstanding proposal for new services on the route could be accommodated with the capacity released. Taking the example of the East Coast Main Line (Appendix D4), for example, removal of services over the Welwyn viaduct to Letchworth Garden City, or to Cambridge and Peterborough, would not necessarily mean that commercial proposals for additional longer distance services could be accommodated.
- 7.30 A potential outcome, therefore, is that scarcity charges at some locations result in the removal of existing services without facilitating the addition of new ones. The effect might also be exacerbated if, as we argue in Section 6, franchisees which were dominant or sole operators were over-incentivised to withdraw peak services.
- 7.31 The materiality of the outcome would ultimately depend on the extent to which franchisees were free to respond to scarcity charges and the balance of incentives on them to do so. We note, however, that:
- The root cause of the issue is that railway capacity is not interchangeable between operators, and removal of one service would not necessarily enable the introduction of another.

- Possible mitigation measures would be to have tighter franchise specifications, weakening or removing the rationale for scarcity charges. Any such mitigation measures would lie mainly with the funding authorities.

Distorted investment incentives

- 7.32 The distortion arising where there is a dominant or sole operator could also affect the incentives to invest in additional capacity. A dominant or sole operator might have a strong incentive to hold back from consuming additional capacity (ΔQ) where doing so would also result in a higher unit scarcity charge (ΔP). If demand were rising, the operator would have an incentive to defer any increase in services for as long as possible. This could result not only in sub-optimal provision of services, and use of infrastructure and rolling stock resources (as we set out in 6.34), but also distort any signals to Network Rail to expand capacity which might normally be sent through rising capacity utilisation.
- 7.33 Unless effectively mitigated, therefore, the response of a dominant or sole operators to scarcity charges might distort any signals to Network Rail to accommodate additional capacity or to expand the network (see 1.4), which we discuss in Section 8.
- 7.34 In summary, scarcity charging might be counterproductive as a means of signalling the need for investment in additional infrastructure capacity, as it could delay, rather than advance, the use of capacity. This might also, potentially, result in free-riding, whereby potential entrants would defer formal application for capacity until additional capacity had been brought into use and they were able to avoid any payment of scarcity charges.
- 7.35 The outcome would ultimately depend on the structure, level and location of scarcity charges and the credible patterns of operator exit and entry at the charging locations. We note, however, that:
- The root causes are that markets may be illiquid, and that dominant or sole operators are common and are likely to remain so unless existing franchises were subdivided into smaller packages
 - Possible mitigation measures lie mainly with Network Rail, through continuing to base investment appraisal on robust projections of future capacity and demand

Summary of mitigation measures

- 7.36 We stress that the above analysis is not exhaustive and that it is possible that detailed study would reveal other secondary effects of scarcity charges which would need to be anticipated and mitigated. Nonetheless, we summarise in Table 7.1 the unintended effects and possible mitigation measures.

TABLE 7.1 UNINTENDED EFFECTS AND MITIGATIONS

Effect	Root cause(s)	Mitigation	Action(s)
Mismatch between market-clearing and optimal use	Scarcity charges cannot reflect variations in opportunity cost	None: scarcity charges may be both too high and too low	
Distribution between operators	Inherent in a mixed industry	None: operators have different requirements	
Behaviour at boundaries	Design of scarcity charging system	Smooth boundaries and transitions	
Reserve price effect	Scarcity charges deter applications which might be optimal	Fixed two-stage application timetable	ORR
Inefficient allocation of risk	Socially necessary services	Revisions to franchise agreements	Funding authorities
Distribution between customers	Inherent in a mixed industry	Franchise has tighter specification or payment linked to equity	
Peak services are withdrawn and not replaced	Operator flexibility to profit maximise	Franchise has tighter specification	
Delayed service expansion	Inherent in scarcity charging	Do not use charges to plan investment	Network Rail

Note: list may not be exhaustive

7.37 Taken together, this analysis suggests that the principal mitigation measures would be:

- For ORR to ensure that scarcity charges varied smoothly, and potentially to introduce a two-stage process of fixed annual bidding rounds at which scarcity charges would be related to the initial demand.
- For funding authorities either to tighten franchise specifications, which could negate the rationale for charges, or to adopt measures including greater flexibility, or payment structured to reflect the value of each service and/or to address equity issues.
- For Network Rail to ensure that investment decisions were not based on apparent demand for train paths, which might not reflect actual demand in final markets.

Summary

- 7.38 We noted in Section 6 how the primary effect of scarcity charging might be:
- Rebalancing services from peak to off-peak
 - Rebalancing the mix of services offered
- 7.39 There might, however, be a number of secondary or unintended effects, some of which are inherent in the nature of the industry and could not readily be mitigated. We focus here on five effects in Table 7.1 which we consider likely to be most important and difficult to mitigate.
- 7.40 First, a posted price scarcity charge based on current value in use, or any other measure, may be a poor estimate of the opportunity cost of the capacity consumed by any individual existing service.
- 7.41 Second, careful design would be required to ensure that scarcity charging did not distort incentives near the boundaries of the locations and periods of charging.
- 7.42 Third, dominant or sole operators, which will often be franchised operators, may have unintentionally high effective marginal scarcity charges, which may over-incentivise the withdrawal of peak services and distort their behaviour.
- 7.43 Fourth, optimisation of the allocation of capacity will probably involve redistribution of the coverage of passenger services between geographical markets. Some locations and communities will gain but others will lose.
- 7.44 Fifth, a potentially inefficient outcome is that scarcity charging, coupled with more flexible franchise specifications, would result in the withdrawal of peak services without any immediate or subsequent introduction of services to other destinations being either proposed or possible.

8 Effects on capacity allocation by Network Rail

Introduction

- 8.1 ORR asked us to examine the effect of a scarcity charge on capacity use and allocation, as set out below.

We would like to understand better the extent to which if we were to introduce robustly calculated charges which reflect the cost of scarcity these would still need to co-exist with an administrative system for allocating capacity. In order to obtain this better understanding we would like the consultant to provide a reasoned opinion on the extent to which a scarcity charge would bring about an improved understanding of capacity and inform decision making around the provision and use of the network. Evidence to support the consultant's views might be in the form of a case study of where the absence of a price which includes the cost of scarcity has meant that a decision has been made using expert judgement. The consultant should consider how its view would change if there was introduction of a shadow charge only i.e. there was no customer behavioural response. We would like the consultant to give specific consideration to how a scarcity charge might inform decisions in respect of all traffic on the network including franchised and open access passenger and freight flows.

- 8.2 We discussed in Sections 6 and 7 the extent to which a scarcity charge would send signals and incentivise operators for the efficient use of capacity.
- 8.3 We discuss in this Section whether and how a scarcity charge, or shadow charges, could result in more efficient capacity allocation by Network Rail. In the remainder of this Section we discuss in turn:
- The administered process of railway timetabling
 - The administered process of allocating infrastructure capacity
 - The implications for scarcity charging
 - The alternatives of shadow scarcity charges or a priority system

Railway timetabling

- 8.4 We summarise in Appendix B2 how the definitive allocation of capacity on a railway system is the timetable, and how timetables have been prepared since the early nineteenth century using administrative procedures.
- 8.5 Timetabling is an art and not a science and cannot be reduced to a series of procedural rules which automatically lead to an optimum result. Instead it generally relies on a combination of objectives and criteria, against which successive iterations of a potential timetable are tested until a satisfactory result is reached. When devising a wholly new timetable, rather than incremental changes to an existing one, it typically proceeds top down “What approach to this timetable is likely to be appropriate?” rather than bottom up “When should this train run?”.

The Network Code

8.6 The Network Code, developed collaboratively by the industry since the Railways Act 1993, recognises these practicalities in the form of the objectives and criteria set out in Table 8.1. The Network Code does not, and cannot, specify a procedural approach to developing a timetable.

TABLE 8.1 OBJECTIVES AND CRITERIA IN THE NETWORK CODE

Content		Text
Objective	4.6.1	To share capacity on the Network for the safe carriage of passengers and goods in the most efficient and economical manner in the overall interest of current and prospective users and providers of railway services
Criteria	4.6.2	Maintaining, developing and improving the capability of the Network
		The spread of services reflects demand
		Maintaining and improving train service performance
		Journey times are as short as reasonably possible
		Maintaining and improving an integrated system of transport for passengers and goods
		The commercial interests of Network Rail (apart from the terms of any maintenance contract entered into or proposed by Network Rail) or any Timetable Participant of which Network Rail is aware
		Seeking consistency with any relevant Route Utilisation Strategy
		As far as possible, International Paths included in the New Working Timetable at D-48 are not subsequently changed
		Mitigating the effect on the environment
		Enabling operators of trains to utilise their assets efficiently
Avoiding changes, as far as possible, to a Strategic Train Slot other than changes which are consistent with the intended purpose of the Strategic Path to which the Strategic Train Slot relates		

Source: Network Code - Part D

8.7 It would, in principle, be possible to add to the Network Code an criterion related to scarcity charging, such as:

- To minimise the total scarcity charges paid by operators, while still complying with their access rights
- To distribute the scarcity charges fairly among operators

8.8 However, the Network Code also specifies that “Where, in light of the particular circumstances, Network Rail considers that application of two or more of the relevant Considerations will lead to a conflicting result then it must decide which of them is or are the most important in the circumstances and when applying it or them, do so with appropriate weight.”

- 8.9 Thus, while new criteria could in principle be introduced into the Network Code, their effect, if any, would depend on what importance Network Rail chose to attach to them in the circumstances, and cannot readily be predicted.

Incorporating price signals

- 8.10 Timetabling had never been expected or designed to deal with price signals because it normally took place within an integrated railway. It was only after railway restructuring in Sweden in 1988 and Great Britain after 1991 that consideration was given to linking capacity allocation to value, as we summarise in Appendix B.
- 8.11 For freight railways, where individual trains serve different markets and can be considered independently, the iteration process can use optimisation software to hunt for the “best” timetable against criteria such as minimising total journey time or deviations, which might in principle be explicitly valued, from ideal departure and arrival times. This approach was examined in Sweden’s newly-liberalised railways, as we describe in Appendix B2, but was not adopted.
- 8.12 For passenger railways, in contrast, each passenger train, or connection, between two stations contributes to the overall service to the final customer. This means that the value of each train will depend on all the other trains in the timetable (we discuss this issue of contingency in Appendix B6) and that timetable development cannot be based on prior signals of value. In effect:
- Ex post, once a timetable has been prepared, it is possible to estimate its total value to passengers, or the effect of changes to it, with industry tools such as MOIRA (Sweden has an equivalent tool).
 - Ex ante, before a timetable has been prepared, an operator cannot signal the value of a requested train path. This is because its value to both operator and passengers will depend on the other paths or diagrams offered to the operator, to provide regular services or connections, and to the operator’s competitors.
- 8.13 For example, if two operators wish to provide an hourly service from A to B, the optimum for the passenger is for them to run half an hour apart, but the optimum for each operator is to run immediately before their competitor. (Attempts for operators to run immediately before competitors in the deregulated bus industry, where timetabling is not based on social objectives, are well documented.) The value of each path to its operator will be highly dependent on the existence and timing of the other operator’s path. An operator requesting a path departing every hour, on the hour, cannot signal its value to Network Rail without first knowing whether and when there will be a competing service.

Allocation of capacity

The two-stage process

- 8.14 Allocation of capacity to rail operators in Great Britain takes place in two stages:
- Operators are awarded and exercise firm contractual rights to use Network Rail’s infrastructure, often through a framework agreement lasting more than one timetable period. This may require sufficient timetabling work to confirm that the rights can be accommodated.

- Only operators with contractual rights are entitled to have their requirements considered in the timetabling process. This results in the development of a full Working Timetable, consistent with all contractual rights, against the objectives and criteria set out in Table 8.1.
- 8.15 All operators are expected to provide diagrams to support their applications and are unlikely to accept offers of capacity which do not constitute efficient and robust all-day diagrams.
- 8.16 EU law limits the duration of framework agreements, and we understand from ORR that there are European Commission proposals to limit the proportion (by some measure) of capacity which can be allocated for longer than one year or timetable period. At present, however, much of the capacity of the network, and almost all of the capacity in some locations, has already been awarded for a number of timetable periods and is held as firm contractual rights. This means that the typical recurring workload of Network Rail is not to reallocate all the capacity from a zero base, but to examine a limited number of requests for additional and changed capacity and to see whether they can be accommodated.
- The model contract*
- 8.17 Access rights are all approved by ORR, after consultation and on the basis of sufficient work to be sure that Network Rail can offer the rights without conflicting with others. ORR may take into account performance, as with the refusal to allow Southern a fourth fast path on the Brighton Main Line.
- 8.18 If it appears that the addition or altered rights can be accommodated, this will be done using a model contract, approved by ORR, setting out the operator's rights. ORR advised us that the model contract specifies:
- As a minimum, the "quantum" (number) of rights to operate between an origin and destination with a particular formation of train
 - Additionally, where requested and practicable, addition details such as maximum journey times, regularity of services, connections, or "clock face" departure times
- 8.19 Departmental traffic's requirements vary: the High Speed Track recorder may need specific paths, but movement of materials to work sites cannot be predicted. Network Rail now pays for all possessions, and ORR does not consider the current capacity retained for departmental traffic to be excessive.
- 8.20 Freight access rights may be firm or requested within a timetable period for up to one year. Firm freight rights may be:
- Level 1, with timings specified to within an hour: this may be necessary for services such as scheduled and reservable container trains
 - Level 2, with rights to a quantum of paths but no specific timings
- 8.21 Passenger access rights may be either firm or contingent, which will only apply if they can be fitted.
- 8.22 Once rights are granted ORR does not normally become involved in exactly how Network Rail creates a timetable consistent with them.

The timetabling process

- 8.23 Only when an operator has rights does it become a participant in the timetabling process. The fact that rights should only be granted after it has been established that they are mutually compatible should avoid any requests for capacity which cannot be satisfied adequately and which would trigger a declaration of congested infrastructure. (Directive 2001/14 Article 16 appears to allow the infrastructure manager to set requirements of applicants to ensure that they are credible.)
- 8.24 ORR informed us that there have been a small number of cases in which it has not been possible to satisfy all rights, and infrastructure has been declared congested, as we illustratively identified in Table 5.3. This can occur in circumstances such as where:
- Franchise agreements unintentionally specify services for which rights cannot be awarded: in these circumstances the franchising authority, not the operator, is at risk if it specifies services which ORR cannot or will not approve.
 - Decisions regarding late applications for rights have to be made on the basis of limited information, and in extreme circumstances rights may be awarded which cannot subsequently be timetabled.

Implications for scarcity charging

- 8.25 We return now to the brief set by ORR, and set out our provisional response.

We would like to understand better the extent to which if we were to introduce robustly calculated charges which reflect the cost of scarcity these would still need to co-exist with an administrative system for allocating capacity. In order to obtain this better understanding we would like the consultant to provide a reasoned opinion on the extent to which a scarcity charge would bring about an improved understanding of capacity and inform decision-making around the provision and use of the network. Evidence to support the consultant's views might be in the form of a case study of where the absence of a price which includes the cost of scarcity has meant that a decision has been made using expert judgement. The consultant should consider how its view would change if there was introduction of a shadow charge only i.e. there was no customer behavioural response. We would like the consultant to give specific consideration to how a scarcity charge might inform decisions in respect of all traffic on the network including franchised and open access passenger and freight flows.

- 8.26 We deal below in turn with:
- Robustly calculated charges which reflect the cost of scarcity
 - The need to co-exist with an administrative system for allocating capacity
 - Understanding of capacity and decision-making around provision and use
 - A case study of where price could remove the need for expert judgement
 - The effects on franchised and open access passenger and freight flows

Robustly calculated charges which reflect the cost of scarcity

- 8.27 Applying the economic theory set out in Section 3 on the practical basis set out in Section 5 suggests that it would be possible to make a calculation of the value of an existing train service or diagram on a number of bases, but that this might vary

over a wide range according to the administrative rules used and might only poorly reflect opportunity cost. Additionally, the analysis in Section 6 suggests that a scarcity charge correctly calculated for a minor operator might still over-incentivise a dominant operator if marginal service changes affected the scarcity charge paid for all services.

8.28 In addition, any scarcity charge calculated in this way would provide:

- No information on whether there were other requests for capacity
- Little information on whether the presence of an existing service was a barrier to granting other requests for capacity
- Little information on whether removal of an existing service would permit the granting of other requests for capacity

The need to co-exist with an administrative system for allocating capacity

8.29 In addition to the systems and processes required to calculate a scarcity charge, an administrative system would still be required:

- To determine whether additional access rights could be awarded
- To prepare a Working Timetable

Understanding of capacity and decision-making around provision and use

8.30 The analysis in Section 5 shows that a scarcity charge must either be set without reference to value, as has taken place in the international examples studied, or based on detailed analysis of the value in use of each service in the current timetable.

8.31 In principle operators and funders could already carry out such detailed valuation of existing services, and we would expect that open access and freight operators do so, but understand that this rarely takes place for franchised services. If service-by-service valuation were carried out as part of the process of setting a scarcity charge, it would seem logical to provide it to operators and their funders, if necessary on a confidential basis.

8.32 It would appear, prima facie, that Network Rail and operators would be better able to make informed decisions on capacity provision and use with this detailed analysis, rather than on the more limited information signalled by a single scarcity charge derived from it. A scarcity charge alone would necessarily provide less information than the more detailed analysis needed to set it.

A case study of where price could remove the need for expert judgement

8.33 Given the nature of the timetabling process, the only circumstances in which expert judgement could be avoided would be where two or more applicants required effectively identical capacity.

8.34 In these circumstances it would in principle be possible to have an auction of the capacity to the highest bidder, as is provided for in the Czech Republic or Switzerland (see Appendix C8.2). There are no provisions for auctions in Great Britain, and an auction in which bids are made by operators is a different process from a predetermined charge set by the infrastructure manager. However an auction could not necessarily take into account “prospective users and providers of

railway services” or, more widely, forecasts of future changes in capacity and demand.

The effects on franchised and open access passenger and freight flows

- 8.35 Scarcity charges could, in principle, be applied to all types of operator, although in the case of paths reserved for intermittent traffic, it is not clear how they would be paid when the paths were not used.
- 8.36 Scarcity charges would almost certainly result in a rebalancing of charges among operators, as discussed in Section 6. Without a detailed proposal, however, we cannot predict what effect this would have on the mix of operators or whether it would be consistent with ORR’s objectives or government policy.

Shadow scarcity charging

- 8.37 As we set out in 8.7, it would, in principle, be possible to add to the Network Code a criterion related to scarcity charging. This could also apply if scarcity charging were used as a shadow charge, such as:
- To minimise the total scarcity charges paid by operators, while still complying with their access rights
 - To distribute the scarcity charges fairly among operators
- 8.38 As we set out in 8.9, however, scarcity charging would not necessarily have any weight unless the Network Code was modified to elevate it above other criteria, and this would also be the case for shadow scarcity charging.

Priority systems

- 8.39 One alternative to allocation by price is allocation by priority, a process based at least implicitly on the assumption that priority services have the highest opportunity cost (that is, the reduction in value resulting from their withdrawal would be greater than for other services). Directive 2001/14 specifically provides, in Article 22, that where “the infrastructure has been declared to be congested, the infrastructure manager may in addition employ priority criteria to allocate infrastructure capacity”.
- 8.40 We identified a number of variations of priority systems in Appendix C8:
- In Denmark, Banedanmark requires applicants to prioritise their own paths, which provides information on the relative importance they attach to each.
 - In Portugal, REFER’s priority system reflects time of day and week, frequency, regularity, distance, type of service and whether in service type of service, and could in principle also reflect any or all of speed, stopping pattern and direction of travel.
- 8.41 These priority rules can reflect a number of factors which influence consumption of capacity, need not deter apparently low-value applicants for whom capacity might be found, and do not depend on the calculation of a charge which may be too low to have any material effect.
- 8.42 In practice, priorities could be informed by calculating the typical values generated by different kinds of services using an approach similar to that already

examined by the ORR. One possible outcome is that, once developed, scarcity charging would have little or no effect that could not be achieved more cheaply and transparently through priority rules. In the hypothetical example shown in Table 3.1, the efficient outcome could equally be achieved using publishing priorities, such as that:

- Express trains have priority over stopping trains
- Trains consuming a “standard path”, which could be set at the speed of an express train, have priority over slower or faster trains

8.43 In addition, a system based on published priorities need not deter any application, because it introduces the possibility or probability of not being awarded capacity, rather than the certainty of an additional payment.

8.44 Without detailed analysis, we cannot anticipate the relative values attached to different train service or diagrams, but there is a possibility that, for example, the effect of scarcity charging is practically always equal to a priority rule.

8.45 We stress, however, the counterargument that high value services (or diagrams) could be inefficiently excluded if they were assigned a low priority through an administered process. Priority rules cannot identify when the relative value of a train service is widely different from its relative priority. In particular:

- Infrabel in Belgium noted that a system which always prioritised passenger services over freight would not permit even high value freight trains.
- Trafikverket in Sweden specifically rejects predetermined rules in favour of greatest economic value (see Appendix C6.31), as does the objective in the Network Code (discussed in Table 8.1).

Summary

8.46 The allocation of railway capacity is through timetabling, which is not a fixed procedure. It is instead governed by a broad objective and a series of criteria, set out in the Network Code, against which possible timetables can be evaluation.

8.47 We have not identified either how timetabling could be made a fixed procedure or how that procedure could explicitly incorporate pricing signals.

8.48 It would in principle be possible to make reference to scarcity charges in one or more additional criteria in the Network Code. Given that no criterion takes precedence over any other, however, this would not necessarily lead to the additional criteria making any difference to timetabling and capacity allocation decisions.

8.49 Scarcity charging based on value would require a valuation of each existing service. These valuations could inform capacity allocation decisions better than the resulting scarcity charge.

8.50 Scarcity charging on a shadow basis, while it would avoid any of the secondary effects to operators described in Section 7, would not address any of these issues.

8.51 Instead of a system of scarcity charging based on values or administered prices, it would be possible to use a priority system, but even complex priority rules would not necessarily reflect the actual value of any given train service.

9 Role in investment decision-making

Introduction

- 9.1 ORR asked us to examine the role of a scarcity charge in investment decision-making, as set out below.

A key rationale for introducing a scarcity charge is that it could provide information on what investment is required on the network and so facilitate more informed decision making. At present, decisions on whether to undertake investments are subject to standard cost benefit analysis (CBA) on a project by project basis. We would like the consultant to contrast these two approaches and to provide a reasoned view on whether, if we could calculate a scarcity charge robustly, the extent to which it will result in better decision-making relative to the existing CBA analysis and the reasons for this. We would like the consultant to provide a view on whether these benefits can be fully realised even if scarcity charges are only introduced on a shadow basis i.e. there is no customer behavioural response.

- 9.2 We discuss in turn below:

- The investment planning process
- International precedents
- A comparison of existing processes and the use of a scarcity charge

The investment planning process

- 9.3 In a typical consumer industry, investment in expansion of capacity may be based largely or wholly on information on sales and revenues. This is particularly likely to be the case for products bought only infrequently, or in emergencies, or by only a small proportion of the population, for which potential customers cannot readily be identified or studied in advance.
- 9.4 Nonetheless, when planning a major investment, many large companies will carry out market research and prepare forecasts of demand taking into account factors such as projected economic activity, population and demographics. Investments such as new manufacturing plants or logistics centres are unlikely to be committed without detailed research and planning.
- 9.5 In the railway, operators could in principle make investment decisions purely on the basis of analysis of current sales and revenue. In practice they are likely to prepared forecasts of demand and develop a business case for investment. This is particularly likely to be the case if they will require authorisation, funding or finance, in areas such as leasing or upgrading rolling stock or expanding services.
- 9.6 We set out in Appendix B5.7 how it was originally assumed that Railtrack would not need, or even have access to, information on final passenger and freight markets, and would identify the need for investment from the pricing signals sent by operators requesting capacity. This proved unworkable, and Network Rail now has access to industry models of demand, and assesses proposed investments in the

context of forecasts of capacity and demand set out in Route Utilisation Strategies (RUSes).

- 9.7 We discussed the investment decision-making process with Network Rail, who pointed out that rail capacity is not provided direct to individual consumers but to a limited number of informed buyers, funders and suppliers. With some exceptions, such as the confidentiality of franchise bids and operator finances, the industry is collegiate and open, and mechanisms such as alliances and the Rail Delivery Group (RDG) are expected to increase this openness.
- 9.8 The industry invests in a number of areas, not all of which relate directly to capacity, including accessibility (Access for All), stations, level crossing removal, electrification on environmental or energy security grounds, information systems and ticketing systems.
- 9.9 We noted in Table 2.2 how railway capacity may be created by operators, by Network Rail, or by a combination of them. While Network Rail investment in capacity must ultimately be underwritten either by operators or by Ministers in a High Level Output Specification (HLOS), the origins of proposals can be diverse including:
- Rolling stock leasing companies (ROSCOs) with proposals for modifications to rolling stock, such as refurbishment, higher performance or additional vehicles
 - Transport authorities with proposals for additional services, stations or facilities
 - User groups with proposals for additional services
- 9.10 The identification, optimisation, planning and implementation of investments is a diffuse process including lobbying, policy-making, formal consultation, preparation of the RUSes and development of business cases typically based on cost benefit analysis (CBA) on the basis of valuation parameters typically specified by funders (such as through WebTAG). It is not possible to impose a wholly standardised approach, particular because the need for commercial negotiation means that many parties do not have access to information such as Network Rail's revenue from charges.

International precedents

- 9.11 Our review of international experience revealed no evidence of any linkages between scarcity-related charges and investment and capacity planning.

Comparison of existing processes and use of a scarcity charge

- 9.12 There are a number of reasons to suppose that a scarcity charge is likely to perform poorly compared with project-by-project cost benefit analysis (CBA) for a number of reasons. We discuss in turn below how:
- Scarcity charges would at best be estimates of opportunity cost
 - Scarcity charges cannot indicate which means of expansion is optimal
 - Scarcity charges at one location cannot indicate where investment is optimal
 - Scarcity charges cannot indicate when investment is optimal
 - Scarcity charges cannot take into account future supply and demand
 - Scarcity charges are not appropriate for planning major investments

Scarcity charges would at best be estimates of opportunity cost

- 9.13 As we set out in Section 5, there is little evidence that an actual or shadow scarcity charge could be calculated “robustly”:
- The element related to capacity utilisation could only be as good as the measures of capacity utilisation available. Additionally, the likely need to base scarcity charging in part on an administrative assessment of where capacity expansion might be needed, as we propose, would preclude basing an assessment of where capacity expansion was needed on scarcity charges.
 - The element related to value could, at best, be a spot estimate of the value of services in the current timetable, calculated using a number of administered rules. We estimate that different rules on how to estimate this value would mean that the estimates might vary over a wide range (see 5.49).
 - The definition of location would be a spot value reflecting current conditions.
 - The definition of “periods of congestion” would be a spot value and potentially arbitrary, although this could be smoothed or based on a moving measure of utilisation centred on the train being charged.

Scarcity charges cannot indicate which means of expansion is optimal

- 9.14 A scarcity charge alone could not indicate which measure, or combination of measures, listed in Table 2.2 was the most effective means of creating additional capacity, which might be possible in the short term without investment. Any efficient process of capacity expansion might need to consider all these measures and to focus on the ones which were most cost-effective in the short, medium or long term, depending inter alia on other planned changes in the network.

Scarcity charges at one location cannot indicate where investment is optimal

- 9.15 As we set out in 2.30, capacity constraints on the network are increasingly not identifiable to specific bottlenecks, but are systemic. We note in Appendix D4 that, while the notional “bottleneck” on the East Coast Main Line is the Welwyn viaduct, the “solutions” implemented involve a platform at Kings Cross, additional tracks between Finsbury Park and Alexandra Palace, a flyover at Hitchin and new platforms at Peterborough, over 120 kilometres from Kings Cross. Attributing a scarcity charge, or a stream of income from it, to Welwyn viaduct would send no clear signals on where and what enhancements were needed.

Scarcity charges cannot indicate when investment is optimal

- 9.16 Equally importantly, a scarcity charge would provide no information on the optimum timing of investment, and could in practice distort incentives both:
- To operators, to defer the introduction of services (see 7.32)
 - To Network Rail, to defer investment so as to benefit from scarcity charges (see Appendix B5.5). Revenue-neutral charges would reduce this effect but could also reduce the incentive to major funders to consume capacity efficiently (see 5.68).

Scarcity charges cannot take into account future supply and demand

- 9.17 A scarcity charge, even if calculated robustly, could not be both a short term value designed to match supply and demand (analogous to a share price) and a

considered view of long term value (analogous to a valuation derived from thorough due diligence). We note the need, in the Network Code, to consider “prospective users and providers of railway services” which we understand to mean a longer term view, rather than just any currently outstanding or unsatisfied applications. Specifically, a current value could not take into account:

- Forecast growth, or fluctuations in underlying passenger and freight demand
- Planned changes to the infrastructure
- Planned changes to passenger and freight services using the infrastructure

9.18 Note also that growth, infrastructure and service changes on competing or new routes may also be material: appraisals of projects on the West Coast Main Line and other routes need to take into account proposals for HS2.

9.19 In contrast, cost-benefit analysis is capable of taking the longer view, reflecting planned and possible changes, examining and comparing a wide range of options and, under certain circumstances, informing decisions on the timing of investment.

Scarcity charges are not appropriate for planning major investments

9.20 Network Rail made the point to us that predetermined scarcity charges are analogous to liquidated damages as a “rule of thumb” for addressing minor and recurring issues. This simplified approach seems prima facie unlikely to be appropriately robust for the appraisal of large, infrequent and irreversible investments in infrastructure capacity.

Scarcity charges on a shadow basis

9.21 Finally, we note that scarcity charges could be applied on a shadow basis and would not distort the incentives to Network Rail to defer investment. They would, however, remove all incentives to operators and would retain all the other weaknesses, relative to cost-benefit analysis, listed above.

Summary

9.22 We have not identified any evidence in other networks of any linkage between scarcity-related charges and investment decisions.

9.23 Network Rail’s customers for infrastructure capacity, and their funders and suppliers, are informed buyers and generate a wide range of proposals for improvements including investment in infrastructure capacity.

9.24 Information on the current revenue from scarcity charges notionally associated with a particular location, even if made available to all interested parties, would be no substitute for detailed analysis to identify what, where and when expansion of capacity would meet the cost-benefit criteria set by the ultimate funders or any expansion.

10 Conclusions and possible next steps

Introduction

10.1 We summarised in Section 3 the economic rationale for scarcity charging and possible approaches to applying it to the provision of rail infrastructure. In this section we summarise our findings from:

- Section 4, international precedents for scarcity charging
- Section 5, whether scarcity charging could reflect opportunity cost
- Sections 6 and 7, whether scarcity charging could improve operator use of capacity
- Section 8, whether scarcity charging could improve allocation of capacity
- Section 9, whether scarcity charging could improve investment decision-making

International precedents for scarcity charging

- 10.2 There are no scarcity charges, as defined in EU law, in any of examples studied.
- 10.3 All the “scarcity-related” charges in the international examples studied are set by administrative means, with no attempt to measure value, although SBB Infrastruktur makes provision for auctions.
- 10.4 Other infrastructure managers told us that scarcity-related charges can influence operator decisions. We have not identified any evidence in other networks of any linkage between scarcity-related charges and investment decisions.

Could scarcity charging reflect opportunity cost?

- 10.5 Scarcity charges would, at best, be a proxy for market-clearing price, which in turn would be a proxy for opportunity cost. Scarcity charging could not directly reflect opportunity cost in the way that is achieved by bilateral trading for airport slots.
- 10.6 A scarcity charge could be calculated which would, in principle, incentivise existing operators to reroute or retime their services where this was possible, but if correctly set at the lowest value in use might be insufficient to incentivise withdrawal. It would also deter applications with low value in use.
- 10.7 Even a “robust” calculation might result in wide range in the level of scarcity charges, depending on the definition of the minimum element of service which could be withdrawn in response.
- 10.8 It might be desirable or necessary for scarcity charges to be revenue-neutral, but while the Department of Transport remains the dominant buyer of rail capacity this would blunt the incentives to both it and franchise operators, unless they were no longer able to pass through charges to their funder. They might only be willing to bear exposure to charges if they also had the flexibility to withdraw peak period services, particularly on London commuter routes.

Could scarcity charging improve operator use of capacity?

10.9 We noted in Section 6 NERA’s view (see Appendix B9.1) that “It is not clear that there are widespread problems of poor capacity utilisation caused by train operators’ decisions.”

10.10 Table 10.1 below repeats Table 6.3, which illustrates the possible effects of scarcity charging at sufficiently high levels to influence operator behaviour.

TABLE 10.1 POSSIBLE OPERATOR RESPONSES TO SCARCITY CHARGING

Market	Possible withdrawal or retiming	Possible new services
Non-timetabled	Take-up of reserved capacity declines, so less capacity needs to be reserved	Replacement or retimed services outside charging times
Freight	Take-up of reserved capacity declines, so less capacity needs to be reserved Demand for peak period paths declines	Replacement or retimed services outside charging times
Charter	Likely to avoid scarcity charging, given the nature of charter services	
Open access	Uncertain effect, depending on business models of open access operators	
Franchised	Fewer peak services Risk that scarcity charging over-incentivises peak service withdrawal	All-day long-distance services

Note: assumptions illustrative, see text after Table 6.3 for details

10.11 The effects are uncertain, but sufficiently high scarcity charging might have the effect of:

- Rebalancing services from peak to off-peak
- Rebalancing the mix of services offered

10.12 There might, however, be a number of unintended effects, some of which are inherent in the nature of the industry and could not readily be mitigated. We focus here on five effects (Table 7.1 provides a longer list) which we consider likely to be most important and difficult to mitigate.

10.13 First, a posted price scarcity charge based on current value in use, or any other measure, may be a poor estimate of the opportunity cost of the capacity consumed by any individual existing service.

10.14 Second, careful design would be required to ensure that scarcity charging did not distort incentives near the boundaries of the locations and periods of charging.

10.15 Third, dominant or sole operators, which will often be franchised operators, may have unintentionally high effective marginal scarcity charges, which may over-incentivise the withdrawal of peak services and distort their behaviour.

10.16 Fourth, optimisation of the allocation of capacity will probably involve redistribution of the coverage of passenger services between geographical markets. Some locations and communities will gain but others will lose.

- 10.17 Fifth, a potentially inefficient outcome is that scarcity charging, coupled with more flexible franchise specifications, would result in the withdrawal of peak services without any immediate or subsequent introduction of services to other destinations being either proposed or possible.

Could scarcity charging improve allocation of capacity?

- 10.18 The allocation of railway capacity is through timetabling, which is not a fixed procedure. It is instead governed by a broad objective and a series of criteria, set out in the Network Code, against which possible timetables can be evaluation.
- 10.19 We have not identified either how timetabling could be made a fixed procedure or how that procedure could explicitly incorporate pricing signals.
- 10.20 It would in principle be possible to make reference to scarcity charges in one or more additional criteria set out in the Network Code. Given that no criterion takes precedence over any other, however, this would not necessarily lead to the additional criteria making any difference to timetabling and capacity allocation decisions.
- 10.21 Scarcity charges based on value would require a valuation of each existing service. These valuations could inform capacity allocation decisions better than the resulting scarcity charge.
- 10.22 Scarcity charging on a shadow basis, while it would avoid any of the secondary effects to operators described in Section 7, would not address any of these issues.
- 10.23 Instead of a system of scarcity charging based on values or administered prices, it would be possible to use a priority system, but even complex priority rules would not necessarily reflect the actual value of any given train service.

Could scarcity charging improve investment decision-making?

- 10.24 We have not identified any evidence in other networks of any linkage between scarcity-related charges and investment decisions.
- 10.25 Network Rail's customers for infrastructure capacity, and their funders and suppliers, are informed buyers and generate a wide range of proposals for improvements including investment in infrastructure capacity.
- 10.26 Information on current revenue from scarcity charges notionally associated with a particular location, even if made available to all interested parties, would be no substitute for detailed analysis of what, where and when expansion of capacity would meet the cost-benefit criteria set by the ultimate funders or any expansion.

Possible next steps

- 10.27 ORR invited us to set out suggestions for possible next steps if it wished to investigate the concept of scarcity charging further.
- 10.28 We suggest that the most appropriate next steps would be to carry out a case study of a route on which one or more applications for capacity have not been satisfied, with the aim of identifying whether scarcity charging could result in a low value service being replaced with a higher value one.

Final Report

- 10.29 The process would be to:
- Calculate measures of value in use of each current service, on one or more bases, with the aim of finding the absolute and relative values associated with different individual train paths and diagrams.
 - Consider whether and how, in the absence of any contractual obligations, each service might be rerouted, retimed or withdrawn, and the approximate level of scarcity charge at which the operator might choose to do so.
 - Assuming the introduction of a scarcity charge sufficiently high to result in changes to one service, carry out a timetabling exercise to identify whether it would be possible to readjust the timetable, consistent with the rights of the remaining operators, to include one or more of the unsatisfied applications.
 - Estimate whether the unsatisfied application would be commercially viable with the paths made available and at the calculated scarcity charge.
 - Use MOIRA and other tools to compare the overall value of the original and revised timetables.
- 10.30 We note that this would be a potentially complex exercise, but suggest that at least an indicative understanding of all these effects would be desirable before proceeding further with scarcity charging.

APPENDIX

A

BIBLIOGRAPHY

A1 ORR REPORTS AND STUDIES

ORR (2005)

“Structure of costs and charges review, Conclusions”

ITS (2006)

“Scoping study for scarcity charges, Final report for the Office of Rail Regulation”

NERA (2007)

“The Impact of a Reservation Charge, Final Report for the Office of Rail Regulation”

CEPA (2010)

“High level review of track access charges and options for CP5”

ORR PR13 First Consultation (2011)

“Periodic review 2013, First Consultation”

NERA (2011)

“Using incentives to Improve Capacity Utilisation, Report for ORR”

ORR (2012)

“Track access contract (Passenger services)”

ORR PR13 Financial and Incentive Framework Consultation (2012)

“Setting the financial and incentive framework for Network Rail in CP5”

ORR PR13 Volume Incentive Consultation (2012)

“Periodic Review 2013, Volume Incentive Consultation”

ORR Long term regulatory statement (2013)

“Opportunities & challenges for the railway, ORR’s long-term regulatory statement”

ORR PR13 Final determination (2013)

“Periodic Review 2013: Final determination of Network Rail’s outputs and funding for 2014-19”

A2 OTHER REPORTS AND STUDIES

Nilsson (2002)

“PRICING THE USE OF SWEDEN’S RAILWAYS; Are Charges in line with Marginal Costs?”

ITS (2002)

“Implementing rail infrastructure charging reform - barriers and possible means of overcoming them”

IEA (2003)

“A Market in Airport Slots” by Keith Boyfield, David Starkie, Tom Bass & Barry Humphreys

Final Report

EU Task Force on Rail Infrastructure Charging (2004)

“Summary Findings on Best Practice in Marginal Cost Pricing”

CATRIN (2009)

CATRIN, Cost Allocation of TRansport INfrastructure cost “D8 - Rail Cost Allocation for Europe, Annex 3A - Estimating the opportunity cost of slots - A case study in Great Britain”, by Daniel Johnson, Richard Connors and Chris Nash (ITS)

Heathrow Airport (2012)

“An introduction to Secondary Slot Trading”

House of Commons Library (2012)

“Airports: aviation slots”

CER (2013)

“Scarcity charging in rail”

A3 OTHER DOCUMENTS

European Commission

Directive 91/440/EEC (English version)

Directive 2001/14/EC (English, German, Italian and Swedish versions)

Great Britain

Network Rail “Freight Route Utilisation Strategy, March 2007”

Network Rail “East Coast Main Line Route Utilisation Strategy, February 2008”

Network Rail “The 2015 Network Statement”

Austria

ÖBB Infra “Network Statement 2014 of ÖBB-Infrastruktur AG” (English version)

ÖBB Infra, “Network Access Product Catalogue Train Path, Train Run and other Services 2014 of ÖBB-Infrastruktur AG” (English and German versions)

Belgium

Infrabel “Network Statement Valid from 09/12/2012 to 14/12/2013, Version of 21/06/2012” (English version)

Infrabel “Network Statement Annex B.6: Priorities for the allocation of capacities in a congested situation : Congested lines” (English version)

Infrabel “Network Statement Annex B.7: Priorities for the allocation of capacities in a congested situation: The types of lines and the types of trains” (English version)

Infrabel “Network Statement Annex B.8: Priorities for the allocation of capacities in a congested situation: List of lines by type” (English version)

Czech Republic

SZDC “The Network Statement on nationwide and regional rail systems, Valid for preparation of Timetable 2014 and for Timetable 2014, as amended by Amendment No 1/2013 effective since 1 March 2013” (English version)

Denmark

Banedanmark “Network Statement 2015” (English version)

Banedanmark “Bilag 4.4, Kapacitetsudnyttelse 2014” (in Danish)

Germany

Bundesnetzagentur “Überprüfung des Trassenpreissystems (TPS) gemäß § 14f Abs. 1 AEG hier: Auslastungsfaktor”, letter 1 July 2011 (in German)

DB Netz “DB Netz AG Network Statement 2014” (English and German versions)

DB Netz “Anlage 4.3.D zu den Schienennetz-Benutzungsbedingungen der DB Netz AG SNB 2014, Nutzungsvorgaben für den als voraussichtlich in naher Zukunft für überlastet erklärten Schienenweg Hamburg Hbf” (in German)

DB Netz “Trassenpreissystem (TPS) Liste der Entgelte der DB Netz AG 2014 für Zugtrassen, Zusatz- und Nebenleistungen” (in German)

DB Netz “Liste der Entgelte für Trassen gültig ab 15.12.2013” (in German)

Norway

Jernbaneverket “Network Statement 2015” (English version)

Portugal

REFER “Network Statement 2014” (English version)

Romania

CFR “Network Statement” (English version, March 01, 2013)

CFR “Annex 17, Infrastructure Sections with Congested Capacity” (English version)

CFR “Annex 18a, DECISION OF THE DIRECTOR GENERAL No. 62 of June 19, 2012 on the declaration of the Sighisoara - Atel Infrastructure Section as infrastructure with congested capacity” (English version)

CFR “Annex 18b, ORDER No. 1165 of July 10, 2012 on approving the priority criteria with regard to the railway infrastructure capacity allocation on the congested capacity sections, as well as the and supply of transparency” (English version)

Sweden

Trafikverket “Network Statement 2014, Edition 09/12/2012” (English version)

Trafikverket “Underlagsrapport, PM Effektbeskrivning av förslag till nya avgifter för T15” (in Swedish)

Trafikverket “Järnvägsnätsbeskrivning 2015, Samrådsutgåva, 2013-09-27” (in Swedish)

Trafikverket “Järnvägsnätsbeskrivning 2014, bilaga 4.2 - Prioriteringskriterier” (In Swedish)

Final Report

Trafikverket “Järnvägsnätsbeskrivning 2015, bilaga 6.1 - Tåglägesavgift, Passageavgift och emissionsavgift” (In Swedish)

VTI “Differentierade marginalkostnader inom järnvägen” (in Swedish, 2013)

Switzerland

SBB Infrastruktur “Network Statement 2014. For ordering and execution of timetabled transport operations from 15 December 2013 to 13 December 2014, November 2013” (English and German versions)

International

Eurotunnel “Eurotunnel Network Statement, 2015 Working Timetable”

APPENDIX

B

SCARCITY CHARGING AND CAPACITY ALLOCATION AND INVESTMENT

B1 INTRODUCTION

B1.1 This Appendix lists and reviews a number of developments relevant to scarcity charging on railways, summarised in the table below.

DEVELOPMENTS IN SCARCITY CHARGING

Date	Event
1988	Sweden separates rail infrastructure (Banverket) and operations (SJ) and sets access charges on a marginal cost basis.
1991	Europe, Directive 91/440/EEC sets out initial requirements for separation of infrastructure and operations and for separate accounting.
1991-2	Studies of rail access and charging, including how pricing could incentivise efficient use of existing capacity and provision of additional capacity.
1992-3	Sweden studies whether capacity allocation can be improved by algorithms based on maximising objective functions of value specified by operators.
1993	Railways Act 1993 provides for the restructuring of British Rail.
1994	Railtrack created with “vesting charges” approved by the Secretary of State, subsequently reviewed and approved by the Rail Regulator.
1996-7	First round of passenger franchises let. Railtrack subsequently studies charges related to the consumption of capacity.
2001	Europe, Directive 2001/14/EC makes specific provision for scarcity pricing (Article 7.4) and declaration of infrastructure as congested (Article 22).
2006	ORR “Scoping Study for Scarcity Charges” by ITS Leeds.
2008	ORR “The Impact of a Reservation Charge” by NERA. ORR keeps the case for a reservation charge under review.
2007	ORR considers scarcity and reservation charges as part of Periodic Review 2008.
2010	ORR “High level review of track access charges and options for CP5” by CEPA.
2011	ORR “First Consultation” plans further research into inefficiencies in capacity utilisation, incentives to help allocate capacity, and a track access charge to incentivise better use of capacity.
2011	ORR “Using Incentives to Improve Capacity Utilisation” by NERA concludes that “it is not clear that there are widespread problems of poor capacity utilisation caused by train operators’ decisions”.
2012	ORR “Financial and Incentive Framework Consultation” and “Volume Incentive Consultation” confirm research plans set out in First Consultation.
2013	ORR “Long term regulatory statement” sets out a vision of a charging structure with incentives for efficient provision and use of the infrastructure.
2013	ORR “Final Determination” for CP5.

B1.2 We discuss in further detail below a number of key developments.

B2 RAILWAYS' ADMINISTRATIVE PROCEDURES

- B2.1 Partly as a result of the limitations of the available technology, pricing has not generally been used widely within railways as a means of allocating infrastructure capacity to operators, the subject of this study, or operator capacity to customers.
- B2.2 Administered systems originating in the nineteenth century were used to allocate railway capacity, through the preparation of a timetable. Timetabling can be relatively simple on a lightly-used railway, but becomes increasingly complex as constraints emerge. Timetabling staff have developed a number of “rules of thumb” to achieve workable timetables, including particularly identifying:
- Dominant services, around which others need to be timetabled, such as first timetabling long-distance passenger services, then regional and local services which connect with them
 - Dominant constraints, which may be a particular section of track (Welwyn viaduct), junction (Woking) or station (Birmingham New Street)
- B2.3 Many of the problems of timetabling are analogous to packing objects of various shapes into a limited space. Packing is intuitive and approaches include analogous rules of thumb, none of which relate to the value of the objects, such as:
- Pack large, solid objects first and then smaller, flexible ones around them
 - Pack tight corners as well as possible before moving on to larger spaces
- B2.4 In this context, Section 4.6 of the Network Code, which determines the Decision Criteria to be used by Network Rail in allocating capacity through devising a timetable, sets out an **objective** (4.6.1) and a number of **criteria** (4.6.2) but does not specify what **procedures** (or rules of thumb) Network Rail may or must use.
- B2.5 At the time that rail restructuring was first being considered, in the late 1980s, computer technology was limited and yield management had not yet been established even in the airlines. Rail fares in Great Britain were generally limited to singles, returns and seasons, with the main discounted tickets being the Cheap Day Return, at shorter distances, and the Weekend Return, followed by the Saver, at longer distances. With no mechanism to manage demand by linking fares to the emerging loadings of specific trains, it was often necessary to add trains, or calling points, specifically to deal with known locations and periods of crowding or overloading. Even on long-distance routes, it was necessary to vary capacity throughout the day to deal with relatively peaky demand.

B3 1988 SWEDEN SEPARATES INFRASTRUCTURE AND OPERATIONS

- B3.1 The first market-opening initiative in Europe was taken by the Swedish government, which decided to separate railway infrastructure and operations. The approach was to charge for access to infrastructure on a marginal cost basis, analogous to that for road, with a fixed annual charge per vehicle (analogous to the road vehicle tax) and a low distance-based variable charge (based on estimates of the direct cost of wear and tear on the network). For a period, however, incumbent operator Statens Järnvägar (SJ) was the only operator on the network, but County authorities were given the power to procure local rail services from other operators.

B4 1991 DIRECTIVE 91/440/EEC

B4.1 Directive 91/440/EEC set out initial requirements for separation of infrastructure and operations and for separate accounting.

B5 1991-2 RAIL ACCESS, ALLOCATION, CHARGING AND INCENTIVES

B5.1 Following Directive 91/440/EEC the Department for Transport commissioned extensive studies of how access to the rail network of Great Britain could be opened to competition. Even in advance of the chosen approach, franchising a wide range of passenger services, there was an intention that open access should be permitted and recognition that it could lead to conflicts for limited capacity. There was extensive consideration of how market-based mechanisms could be used to incentivise both:

- The efficient allocation and use of existing capacity
- The efficient provision of additional capacity justified by market requirements

B5.2 At one point it was proposed that the railway timetable would be a dynamic “open system” with potential operators able to identify and reserve “spare” capacity, with the aim that the timetable could be revised every eight weeks in response to emerging market demand. There was, however, no direct evidence that such frequent timetable changes were either practicable or desirable, either to operators or particularly by passengers.

The vision for capacity allocation

B5.3 Given the then preference for market solutions, it remained the assumption that Railtrack, the infrastructure manager, would negotiate access charges bilaterally with potential customers. Under this approach, both operators and infrastructure manager would signal value during a negotiating process:

- Operators would signal to Railtrack the value they placed of train paths, singly or in combination
- Railtrack, as infrastructure manager, would signal to operators the direct and indirect costs, including opportunity costs, of providing each train path

B5.4 It was assumed that the excess of “what the market would bear” over marginal cost would incentivise Railtrack, through bilateral negotiation:

- To accept as many non-conflicting customers as possible
- When conflicts arose, where possible to negotiate with customers to resolve the conflict, such as by charging less for a slightly modified train path, or alternatively to expand capacity for which customers were willing to pay
- More widely, to adapt or expand the infrastructure to customer requirements where the cost of doing so was less than the customer’s willingness to pay

B5.5 However, as a monopoly provider of infrastructure, Railtrack’s commercial response to a capacity shortage was most likely to be to increase charges, and to extract monopoly rents, rather than to increase capacity. Analysis also showed that the capital costs of new infrastructure were likely to be many times greater than the costs of operating and maintaining existing infrastructure.

Final Report

B5.6 These issues were taken into account in the design of regulatory arrangements for the industry. In particular ORR was empowered not only to approve all access agreements but also to direct Railtrack to provide access rights, and the capacity necessary to deliver them, in exchange for a reasonable return on the efficient cost of providing the capacity required.

The vision for investment planning

B5.7 The original assumption was that Railtrack would not need, or even have access to, information on final passenger and freight markets, which it was assumed would remain confidential to the operators. It was instead expected to identify the need for investment from the pricing signals sent by operators requesting capacity. Preparation was made to deal with frequent requests from operators for changes to their access rights, and to respond rapidly with revised timetable proposals and prices.

B5.8 In practice, however, the information provided by pricing signals was of little or no use for network planning, and Railtrack almost immediately attempted to build, using the limiting information inherited from British Rail, its own models of demand in final passenger markets.

B5.9 Only after a number of years, in which it had become clear that Railtrack could not forecast demand for capacity without first forecasting passenger demand, was it given access to the industry standard forecasting models, albeit with restrictions on seeing detailed information on each operator.

B5.10 The Strategic Rail Authority began a process of Route Utilisation Strategies (RUSes), examining markets, demand, capacity and investment needs in the longer term. With its abolition following the Railways Act 2005, responsibility for the RUSes was transferred to Network Rail. RUSes have been produced since 2006 and have been absorbed into Network Rail's Long Term Planning Process (LTPP).

B6 1992-3 SWEDEN STUDIES CAPACITY ALLOCATION BY ALGORITHM

B6.1 Following what was perceived as limited success in introducing new services into the railway, Sweden established a "Committee for increased competition in the railways" (Kommittén för ökad konkurrens inom järnvägen) which carried out a range of studies of the efficient allocation and use of capacity. Thinking in Sweden focused on the requirements of a largely rural and single track network, constraining trains travelling in opposing directions to meet and pass in infrequent loops, so that it would rarely or never be possible to offer operators the exact timings requested for their services. While charges would be based only on marginal costs, one approach considered was for applicants to specify their ideal departure time, stopping patterns and overall journey time, and the disbenefit of any divergence from it. Banverket, the infrastructure manager, would then use optimisation software to create the timetable with the highest value to operators based on their declared (or assumed) objective functions.

B6.2 Under this approach, operators would signal value to the infrastructure manager in the form of the disbenefits of perturbed timings, and the level of perturbation at which a path would no longer be acceptable. The infrastructure manager would then allocate capacity in a way intended to maximise the collective benefits to operators.

- B6.3 Note that this approach did not attempt to identify an “opportunity cost” associated with each path. It would, in principle, have been possible to rerun the optimisation with each service removed, and then calculate the net effect on all other services in the timetable. However, the value measured would depend as much on the origin, destination and stopping pattern of the train as on the time at which it passed any given point or constraint en route. In addition, this would not be a measure of either the value of a hypothetical additional train or whether it could be accommodated.
- B6.4 The optimisation approach could in theory also devise timetables for multiple track lines, or in urban areas, or modify timetables to include additional trains with minimum perturbation to existing services. However, while broadly suitable for freight services carrying different commodities on the same line, it was unworkable for passenger services for a number of reasons, including the commercial need for operators to:
- Make efficient use of rolling stock over the day, week and year as a whole
 - Provide services at regular intervals
 - Provide connections among their own services and with others
- B6.5 The key issue was one of “contingency”, whereby neither the value of a train path to one passenger operator, nor its cost to other passenger operators, can be defined in isolation. The value of each path depends on all the other train paths allocated to the operator, its competitors, and its collaborators in providing connections.
- B6.6 Achieving regular intervals and connections was more easily achieved by the “Taktfahrplan” approach used in Switzerland and Germany, in which a carefully-designed pattern of services and connections was repeated at even intervals throughout the day, sometimes with services added to the basic pattern during peak periods.
- B6.7 Studies in Sweden also examined an approach of permitting open access operators, subject to them compensating existing operators for any loss of revenue. The essence of the approach was:
- An open access operator would propose a new service and the timetable would be reworked to see if it could be included.
 - If it could, revenue models (analogous to MOIRA) would be run to predict the loss of revenue to existing operators.
 - The open access operator would then be quoted a supplementary access charge equal to the compensation to existing operators for their losses.
 - The open access operator would only proceed if the net revenue from the new service exceeded the change in revenue on existing ones.
- B6.8 In this approach:
- The infrastructure manager would signal the opportunity cost (in terms of foregone revenue) of introducing the new service.
 - The potential new operator would signal its willingness (or not) to pay at least this amount in compensation.

Final Report

- B6.9 This approach was not adopted in practice, but we note that in principle it provides a precedent for signalling opportunity cost to operators through charges. However:
- I It was intended to support decisions on the addition of an open access service (analogous to ORR’s application of the “Five stage test”), rather than to prioritise additional services or resolve conflicts between actual or potential services.
 - I It required a bespoke calculation for each service: there could be no “look up table” of the costs of adding services at particular locations or times. This was because the impact of each service on other services would depend on both its calling pattern and the timing of each station call relative to the other services.

B7 1996-7 THE FIRST ROUND OF PASSENGER FRANCHISES

- B7.1 By the time the first round of passenger franchises were let, over the period 1996 to 1997, the role of market forces in access and charging had been limited in a number of ways:
- I ORR’s power to approve all access agreements, rather than accept agreements negotiated bilaterally between Railtrack and the relevant operator
 - I ORR’s power to direct Railtrack to enter into access agreements and to provide capacity
 - I ORR’s policy of Moderation of Competition, originating in its duty “to have regard to the financial position of the Franchising Director”, to limit the scope for poaching or “cherry picking” profitable business from franchises
- B7.2 Nonetheless, some scope for competition remained where separate long distance and commuter franchises both offered services between at least some stations on a number of radial routes from London:
- I Anglia Main Line: Anglia Railways and Great Eastern
 - I East Coast Main Line: GNER and West Anglia Great Northern
 - I Midland Main Line: Midland Mainline and Thameslink
 - I West Coast Main Line: Virgin West Coast and North London Railways
 - I Great Western Main Line: Great Western and Thames Trains
- B7.3 In addition, the Department for Transport initially kept franchise specifications as flexible as possible, with the aim of allowing franchisees to optimise their services above a certain minimum level. Winning franchise bids often took advantage of this flexibility to increase service frequencies, which on many long distance routes rose from two hourly to hourly or hourly to half-hourly. Parallel changes in the fares structure, allowing some peak fares to rise and off-peak fares to fall, resulted in fares varying by as much as a factor of ten by time of day. Instead of managing capacity to match peak demand, it became more possible to manage demand to match an increasingly “flat” profile of capacity through the day.
- B7.4 As a result, the rail timetable in Great Britain has become increasingly like the “Taktfahrplan” timetables of Switzerland and Germany, with regular interval services throughout the operating day, supplemented by additional services at peak periods.

- B7.5 However the combination of this higher service frequency, rising passenger demand, and the policy of Moderation of Competition meant that it became increasingly difficult to optimise or increase services further, because changes to services of one operator would conflict with the access rights or competitive protection of the other.
- B7.6 In the absence of identifiable market mechanisms to resolve this difficulty, one approach adopted by the Department for Transport was to have a single franchise on each route, which could internalise the problem of optimising the timetable and balancing passenger demand across different services. Both the Anglia Main Line and the Great Western Main Line now have a single integrated franchise operator at their London terminals of Liverpool Street and Paddington, although the suburban services of these operators will eventually be replaced by Crossrail.
- B7.7 Thus the overall trend, since the Railways Act 1993, has been:
- Away from the original vision of loosely specified franchises with plentiful spare capacity used by competing open access operators
 - Towards a single tightly-specified franchise exhausting most the available capacity, particularly at times when demand is highest
- B7.8 Our research for the European Commission on the potential impacts of the Fourth Railway Package liberalising domestic rail services highlighted the issue that a combination of flexible pricing and smaller and more frequent trains may improve railway finances in the short term but exhaust available capacity in the medium term.
- B7.9 In the years after franchising, Railtrack commissioned at least two studies of the scope for varying charges to reflect the extent to which trains consumed capacity, with the aim of incentivising them to minimise their impact on other services. The approach was analogous to estimating how much each train contributed to the Capacity Utilisation Index (CUI), which can be calculated on a number of bases, including the UIC 406 standard.
- B7.10 However, the studies identified a number of difficulties with such an approach:
- The results were highly sensitive to the definition of the sections of route over which capacity consumption was calculated. The smaller the section, the lower the CUI and the lower the apparent consumption of capacity.
 - The availability of capacity in any two adjacent sections of route was no guide to whether it was possible to run a train successively through them, which required that capacity in each section was not only available but also contiguous.
 - A calculation based on this method might produce perverse incentives. For example, the capacity consumption measure would be lowest if all trains with the same pattern ran together, whereas the passenger benefit would be greatest if they were evenly spaced throughout the hour and day.
- B7.11 We have not established whether Railtrack discussed these studies with ORR.

B8 2006 ORR'S SCOPING STUDY FOR SCARCITY CHARGES

- B8.1 In 2005-6 ORR commissioned ITS Leeds to carry out a scoping study of scarcity charges. The study introduced distinct definitions of scarcity (conflicts between operator requests for paths) and congestion (delays caused by trains to each other during operation). It highlighted the principle that the most efficient allocation of resources results from pricing at marginal social cost. It also examined approaches to identifying value including auctions, demand and revenue models and calculation of long run average incremental costs.
- B8.2 The study concluded that “the most practical approach to the estimation of scarcity charges is the calculation of the opportunity cost of slots using models to estimate the net social benefit provided by alternative uses of these slots”.
- B8.3 However it provided no detailed proposals of how such calculations could be carried out in practice.

B9 2011 ORR'S USING INCENTIVES TO IMPROVE CAPACITY UTILISATION

- B9.1 In 2011 NERA examined the scope of charges to improve infrastructure capacity utilisation, and consulted industry stakeholders on scarcity charges, but concluded that:
- It is not clear that there are widespread problems of poor capacity utilisation caused by train operators' decisions.
 - A number of regulatory, contractual and planning mechanisms are suitable for addressing specific problems.
 - Charges based on opportunity costs would be likely to vary significantly between different types of service, different parts of the network, and different parts of the day.
 - Much will depend on how much freedom franchisees will have to make significant changes to service patterns under the Department for Transport's revised franchise policy.
- B9.2 NERA's report included no detailed analysis of how charges based on opportunity costs could be calculated or would be applied. They also implied that any such approach would need to offer cost-effective improvements relative to existing regulatory, contractual and planning mechanisms.

B10 2013 ORR'S LONG-TERM REGULATORY STATEMENT

- B10.1 In July 2013 ORR published its long-term regulatory statement, which set out ORR's vision of a simpler, more efficient charging structure, reproduced below.

ORR's vision is of a simpler, more efficient charging structure. It is possible to conceive of a charging structure which reflects the full cost of providing capacity including both short-run wear and tear costs of running trains on the network and the costs of scarcity and congestion. These costs may vary by geography and time of day but we understand the need to balance the pursuit of cost reflectivity with simplicity and practicability.

ORR considers that an efficient charging structure should provide:

- Cost recovery: A mechanism for Network Rail to recover the efficient costs it incurs in providing track and station infrastructure used by train operators.
- Incentives for efficiency of use: Users make better use of products, including capacity, by responding to signals sent through prices based on cost. Charges provide signals to train operators, their suppliers and funders for the efficient use and development of vehicles and the infrastructure.
- Incentives for cost efficiency and allocation: Charges allow costs to be allocated. Where charges allocate costs to those who have caused them to be incurred they provide an incentive to reduce those costs.
- Incentives for efficient provision of goods and services: Charges send signals to providers as to the goods and services they should provide. In this case, charges could provide an incentive to Network Rail to respond to signals sent by users through prices and their consumption decisions about what they are willing to pay for and what Network Rail should therefore provide, as long as those charges cover the cost of provision.
- Fairness, stability and predictability: To support long-term investment in the railway, the impact on different customer groups will be important to understand. The distributional and commercial effect of any revised charging structure would need to be considered carefully.

- B10.2 ORR's Periodic Review 2013 Final Determinations also include, as paragraph 16.30, a statement that "charges provide" the effects listed in the first four bullet points.
- B10.3 A number of points are particularly relevant to scarcity charging, as we summarise in the table overleaf.
- B10.4 Nothing in ORR's current vision is inconsistent with the original Railways Act 1993 vision of a market-led railway in which the greatest possible use is made of pricing signals to incentivise efficient behaviour by all parties.
- B10.5 However, that vision assumed a mechanism based on negotiation, with simultaneous and iterative signalling of cost by Railtrack and value by operators. In contrast, ORR now specifically envisages a mechanism based on charges, with Network Rail signalling cost and the operators responding on the basis of willingness to pay.

ORR’S VISION: RELEVANCE TO SCARCITY CHARGING

Wording	Relevance
<p>It is possible to conceive of a charging structure which reflects the full cost of providing capacity including both short-run wear and tear costs of running trains on the network and the costs of scarcity and congestion.</p>	<p>Makes specific reference to the costs of scarcity and congestion, defined (3.17) as:</p> <p>The cost of scarcity is either the cost of depriving another user of access to the network or the cost of expanding capacity to meet the excess demand.</p> <p>The cost of congestion is a cost associated with one operator’s use of the network on another user.</p>
<p>These costs may vary by geography and time of day.</p>	<p>Both scarcity and congestion vary by geography and time of day.</p> <p>Directive 2001/14 (see Appendix C) refers to the scarcity of capacity of the identifiable segment of the infrastructure during periods of congestion” implying that it is a function of both geography and time of day.</p>
<p>Users make better use of products, including capacity, by responding to signals sent through prices based on cost.</p>	<p>This suggests that charges based on costs will influence existing users to consume capacity in a more efficient way.</p>
<p>Charges provide signals to train operators, their suppliers and funders for the efficient use and development of vehicles and the infrastructure.</p>	<p>This suggests that charges based on cost will influence the use and development of rolling stock.</p>
<p>Charges could provide an incentive to Network Rail to respond to signals sent by users through prices and their consumption decisions about what they are willing to pay for and what Network Rail should therefore provide, as long as those charges cover the cost of provision.</p>	<p>This restates the early 1990s vision for charges as characterised in paragraph B5.2.</p>
<p>Fairness, stability and predictability</p>	<p>Fairness raises the issue of the distribution of charges among operators and the assumption that they will be equal where operators have the same requirement or impose the same cost.</p> <p>Stability and predictability highlight the need for decisions on long term investment in rolling stock or infrastructure to be based on consistent signals.</p>

APPENDIX

C

INTERNATIONAL EXPERIENCE

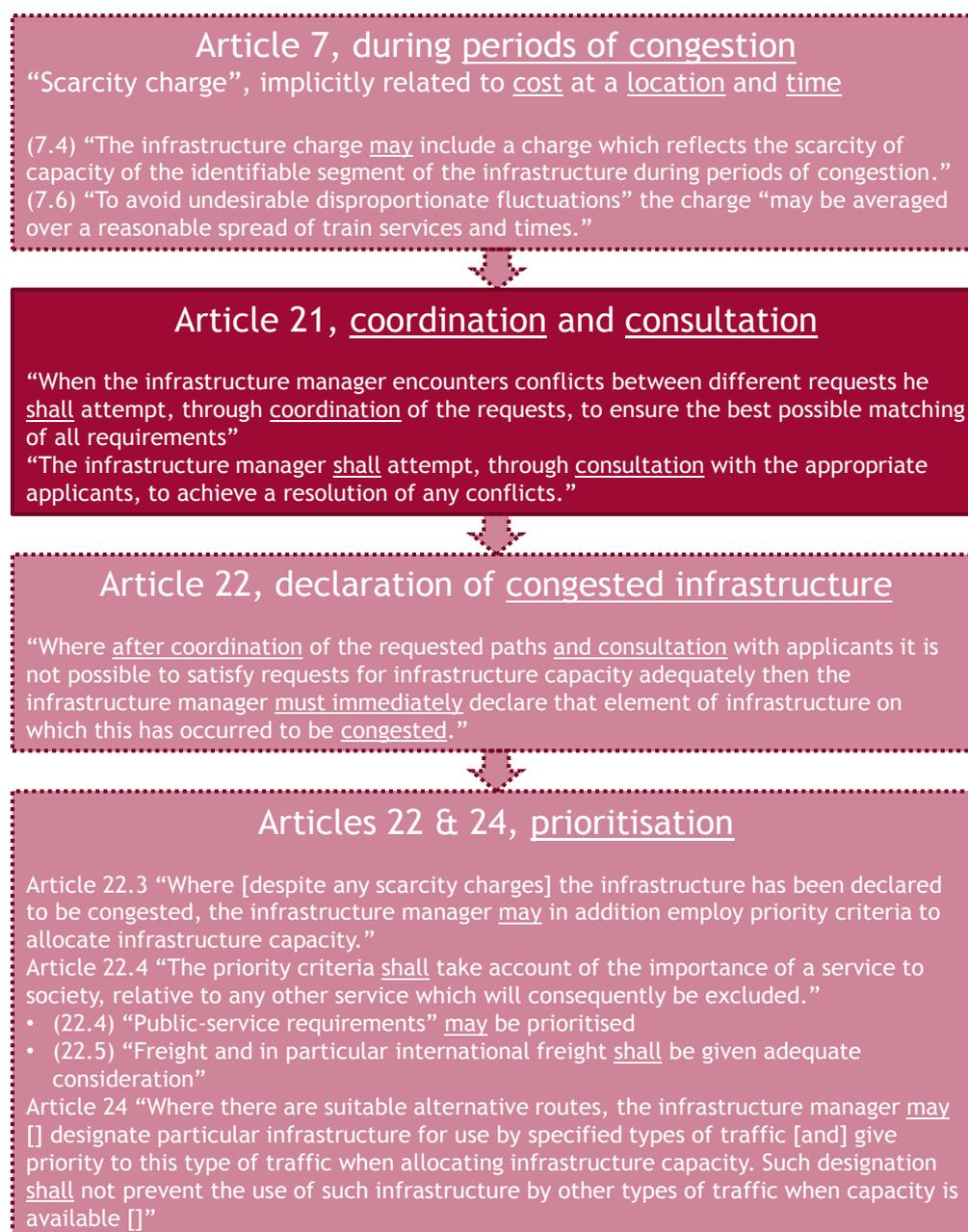
C1 THE LEGAL FRAMEWORK FOR SCARCITY CHARGING

Directive 2001/14/EC

C1.1 Directive 2001/14/EC sets out a framework for the allocation of capacity which we summarise in the figure below. The Articles highlighted deal in turn with:

- Article 7, applying an ex ante charge, which will deter requests for capacity
- Article 21, coordination and consultation with the aim of resolving conflicts
- Article 22, declaring infrastructure congested, and plans to relieve congestion
- Articles 22 and 24, resolving remaining conflicts with priorities

SUMMARY OF DIRECTIVE 2001/14/EC



Source: Directive 2001/14, underlining of key words added by Steer Davies Gleave

Final Report

C1.2 We note that it is not clear whether it was intended that “periods of congestion” in Article 7.4 can only exist on infrastructure which is declared “congested” in Article 22. The English version of the Directive uses the words “congested” and “congestion”. However, the equivalent wordings in the Italian version of the Directive are “congestione” and “saturata”.

C1.3 One possible interpretation is that the distinction is as set out in the table below.

“CONGESTED INFRASTRUCTURE” AND “PERIODS OF CONGESTION”

	Article 22	Article 7.4
Wording	“Congested infrastructure”	“Periods of congestion”
Test specified	After coordination of the requested paths and consultation with applicants it is not possible to satisfy requests for infrastructure capacity adequately	Not specified, but implicitly related not only to locations but also to times
Interpretation	No discretion in definition Applies to the defined element of infrastructure at all times	Discretion in definition, but it: <ul style="list-style-type: none">• Must apply only at some times• May apply only in some places
Problematic issues	Directive 2001/14 is unclear on how the “congested infrastructure” is defined It could, in principle, include the entire route of any unsatisfied applications (As context, Network Rail has declared as congested a section over 160 kilometres long)	Given discretion, charge may be applied in a number of ways. Defining criteria might include: <ul style="list-style-type: none">• Capacity utilisation, based on a measure such as UIC 406• A measure of performance• A measure of the deviation of capacity awarded from operators’ aspirations• A measure of the number of operators withdrawing after coordination and consultation under Article 21

C1.4 However, Infrabel and Trafikverket drew to our attention the scope, under Directive 2001/14 Article 8, to “levy mark-ups on the basis of efficient, transparent and non-discriminatory principles” and suggested that it was possible to introduce scarcity-related charges as mark-ups, without formally referring to any of congested infrastructure, periods of congestion, or scarcity.

C1.5 Our research has also identified a number of such charges which have a “scarcity-related” effect but are not declared to be scarcity charges.

Article 7 “scarcity charges”

C1.6 Article 7.4 allows the infrastructure manager to include a scarcity charge “on the identifiable segment of the infrastructure during periods of congestion”. Article 7.6 allows that “To avoid undesirable disproportionate fluctuations, the charges [] may be averaged over a reasonable spread of train services and times.”

- C1.7 We conclude that, under Article 7.4, infrastructure managers are free to define, identify, and apply charges related to “periods of congestion”, subject to other general principles including non-discrimination.

Article 12 “reservation charges”

- C1.8 Article 12 makes provision for reservation charges “Infrastructure managers may levy an appropriate charge for capacity that is requested but not used. This charge shall provide incentives for efficient use of capacity.” We assume that the key feature of reservation charges is that they become payable once capacity is reserved, whether paid in advance (reservation) or arrears (cancellation).

Article 21 “coordination process”

- C1.9 Article 21 requires that, if any conflicts are encountered, the infrastructure manager coordinates and consults with operators to achieve a resolution of the conflict. In most railways this reflects the core timetabling process which would have operated within an integrated railway in the absence of the Directive.

Article 22 “congested infrastructure”

- C1.10 Article 22 requires that, if not all requests for capacity can be satisfied, the infrastructure manager declares the capacity to be congested.
- C1.11 Article 22 allows that, **only** if the capacity has been declared to be congested, the infrastructure manager may define priorities which:

- (22.4) Shall take into account the importance of a service to society
- (22.4) May prioritise public service requirements
- (22.5) Shall give adequate consideration to (but not necessarily prioritise) freight

Article 24 “specialised infrastructure”

- C1.12 Article 24 also provides for “specialised infrastructure” designated for the use of specified types of traffic, to which the infrastructure manager may give priority when allocating infrastructure capacity, provided that such designation does not prevent the use of such infrastructure by other types of traffic when capacity is available. Article 24 therefore permits the setting of priorities on infrastructure which has not been declared congested, provided that it has been designated “specialised infrastructure” after consultation with interested parties.
- C1.13 The overall structure is capable of interpretation or operation in a number of ways which we characterise below as:
- “Bottom up” with market mechanisms such as those in Article 7.4
 - “Cooperative” with coordination and consultation as set out in Article 22
 - “Top down” with conflict resolution using priority rules under Article 22 or 24

“Bottom up” market-led solution

- C1.14 In this vision, Article 7.4 dominates, and it might be assumed that the application of (published) scarcity prices alone, based on some measure of cost, will ensure

Final Report

both that the infrastructure capacity is allocated efficiently and that there are no remaining conflicts.

- C1.15 If this were the case, none of the provisions in the other Articles would need to be triggered. The need for priorities, for example, would only occur in the extremely unlikely event that two services were willing to pay exactly the same scarcity charge (as can happen in a sealed bid auction, but not an open ascending price auction) and a measure of social value was used as a “tie break” to decide which should be accepted.
- C1.16 Increasing charges on “the identifiable segment of the infrastructure during periods of congestion” would be used to influence operators to reroute, retime or withdraw their services until they could all be accommodated.
- C1.17 However, as we set out in Table 2.2 of the main report, some means of making more effective use of capacity, such as a recast of the timetable, could not be achieved unilaterally by operators and would require coordinated action by a number of parties. While increasing charges until all remaining applications could be accommodated would “clear the market”, it would not necessarily optimise the use of existing capacity, because it could only exploit mechanisms which could be used unilaterally by operators. Table 3.1 illustrates how market-clearing approaches may not lead to optimal allocation of capacity.

“Cooperative” negotiation-led solution

- C1.18 In this vision, no scarcity charges are applied and Article 21 alone suffices, with coordination and consultation leading to all requests for capacity which cannot be met ultimately being withdrawn (or rerouted or retimed to avoid places and times with congestion), with the result that there is no need to declare any infrastructure congested or to apply priorities.
- C1.19 In practice, while priorities declared in Network Statements are only applied under Article 22, they will influence the coordination and consultation process in Article 21. This is because, during coordination and consultation, applicants will be aware which of them will be prioritised in the event of not accepting the outcome of the coordination and consultation process. This seems likely to explain why, in many countries, little or no infrastructure has been declared congested.

“Top down” conflict resolution solution

- C1.20 In this vision, at least some unmet requests for capacity are not withdrawn, triggering the Article 22 requirements that the infrastructure manager must draw up a capacity analysis and a capacity enhancement plan and then, once the infrastructure has been declared congested, apply the priorities set out in the Network Statement, **based on some measure of value**, at least until the capacity has been enhanced.

C2 INTERNATIONAL EXPERIENCE

- C2.1 We carried out a brief review, summarised in the table overleaf, of Network Statements and other related documents produced by European infrastructure managers.

INTERNATIONAL EXPERIENCE: SUMMARY OF FINDINGS

Comparator	Findings
Austria	Additional charge for one section of congested infrastructure, Mödling to Wien Meidling, a two-track section of the Südbahn with intermediate stations served by local trains.
Belgium	Network Statement states that there are no congested lines, but charges vary by time of day and standard train paths consumed.
Czech Republic	No reference to scarcity charges, but provisions for auctions as a last resort to resolve capacity allocation.
Denmark	Prioritisation, and applicants must prioritise their own paths to aid allocation, but no congestion or scarcity charging.
France	Prioritisation, but no congestion or scarcity charging.
Germany	Scarcity charges in the past but no longer.
Italy	Prioritisation, but no congestion or scarcity charging.
Norway	Identifies three sections of congested infrastructure, and provides for scarcity charges, but they are all zero.
Portugal	Sophisticated administrative rules - prioritises by traffic type and time of day, and provision to reserve capacity for foreseeable ad hoc requests - but no reference to congestion or scarcity charging.
Romania	Identifies sections of congested infrastructure, but this is only pending rehabilitation and the “capacity enhancement plan” is to finish the work. No evidence of scarcity charging.
Spain	Prioritisation, but no congestion or scarcity charging.
Sweden	<p>“Passage charges” of SEK 260 for five sections of congested infrastructure:</p> <ul style="list-style-type: none"> • Stockholm Central to Karlberg, 2km of multiple track. • Stockholm Central to Stockholm Södra, 2km of twin track. • Around 35km of the network around Göteborg, including twin and single track and intermediate stations. • Around 35km around Malmö, the historic twin track line including intermediate stations. • Around 17km in Malmö, the CityTunnel opened in 2010. <p>Both Stockholm sections will be relieved by the new Citybanan project: it will be interesting to see what will happen to charges.</p>
Switzerland	Provision for some charges, including optional trains, “to be set as scarcity prices”.

C2.2 On the basis of these findings we agreed with ORR that we would carry out more detailed research in five countries: Austria (ÖBB), Belgium (Infrabel), Germany (DB Netz), Sweden (Trafikverket) and Switzerland (SBB Infrastruktur). We describe our findings to date below and then briefly refer to relevant features of the regimes in other countries.

C3 AUSTRIA (ÖBB)

Current charging structure (Network Statement 2014)

C3.1 The table below summarises the charging structure set out in ÖBB’s Network Statement 2014.

CURRENT CHARGING STRUCTURE IN AUSTRIA

Article		Comments
7.4 Scarcity charge	✘	No reference to scarcity charge
Other charges which relate to “periods of congestion”	✓	“Incentive for capacity optimisation” €0.5638 per train-kilometre
12 Reservation charge or cancellation charge	✓	Detailed in Product Catalogue, tiered by proportion of unused paths, allowance for force majeure
22 Congested infrastructure	✓	1 section (Mödling to Wien Meidling)
Actual charges on “congested infrastructure”	✓	“Supplement for congested infrastructure” €1.2173 per train-kilometre

Article 7.4 scarcity charges

C3.2 ÖBB’s “Product Catalogue” also describes (1.1.4.4) an “incentive for capacity optimisation” (Anreiz zur Auslastungsoptimierung) set at €0.5638 per train-kilometre. The mechanism is intended to flatten peaks in demand and is considered to be an incentive for optimising infrastructure occupancy to improve operational quality on congested routes, particularly at peak times. The price differentiation does not apply to congested infrastructure such as Mödling-Wien Meidling, but is applied at particular times on particular sections of routes where certain capacity utilisation thresholds are exceeded. The timings at which the incentive applies can be highly specific, such as only 0600-0700 between Wien Hausfeldstraße and Siebenbrunn-Leopoldsdorf, or 0600-1000, 1200-1400, 1600-2100 and 2300-2400 between Klagenfurt and Villach.

C3.3 We have not identified how the charge was calculated, but note that:

- A national standard level of charge is unlikely to provide any consistent reduction in capacity utilisation.
- An operator of regular interval services in any area with charging would inevitably pay the charge for at least some trains.
- An operator of less frequent trains might be able to avoid the services in one area, but it might be difficult or impossible to retime a long-distance passenger or freight train to avoid the charge completely without rerouting and/or holding in loops, platforms or sidings to wait for charge-free windows.

C3.4 This suggests that charges like ÖBB’s “incentives for capacity optimisation” would be most likely to affect infrequent or irregular open access, interurban, international or freight operations, allowing regular interval urban, suburban and regional services to be preserved.

- C3.5 At first sight, however, the same effect could be achieved merely by prioritising regular interval services in the timetabling process, without the added complexity of a charge. Awareness of these priorities, defined in accordance with Article 22, could increase the willingness of operators to adjust or withdraw their applications during the coordination and consultation required in Article 21.

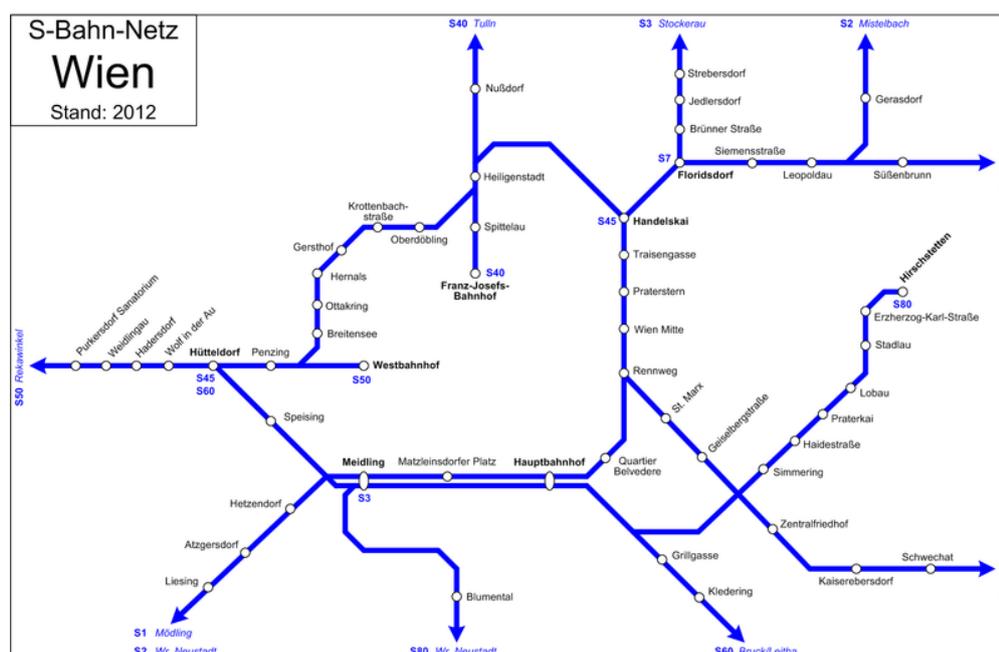
Article 12 reservation charges

- C3.6 ÖBB's sets out details of cancellation charges in the Product Catalogue. These are tiered by the proportion of unused paths and there is allowance for force majeure.

Article 22 congested infrastructure and prioritisation

- C3.7 ÖBB's Network Statement 2014 notes (4.4.3) that "The following route has been declared as congested in accordance with the Railways Act [], whereby [] supplementary fees will be charged during the following time periods: Mödling - Wien Meidling 05:00 - 09:00; 15:00 - 19:00".
- C3.8 Mödling to Wien Meidling is on the main Südbahn southwest approach to Vienna from Graz and Klagenfurt, and in the bottom left hand corner of the figure below.

"SUPPLEMENT FOR CONGESTED INFRASTRUCTURE" IN WIEN



Source: Wikipedia commons

- C3.9 The total distance is around 12 kilometres of mainly 2-track line shared with S-Bahn trains from Mödling stopping at Liesing, Atzgersdorf and Hetzendorf. At Meidling the new Westbahn joins the route at a grade-separated junctions, and trains then continue east towards the new through Hauptbahnhof.
- C3.10 The separate "Product Catalogue", which lists ÖBB's specific charges, identifies (1.1.4.1) the supplement for congested infrastructure (Zuschlag für überlastete Infrastruktur) as €1.2173 per train-kilometre or around €15 for the whole 12 kilometres. Previous work for ORR, ITS (2006), had noted (3.4) that "Austria has a surcharge of €0.50 per train-kilometre for certain bottleneck stretches in the

Final Report

vicinity of Vienna”. The charge is now higher than that reported in ITS (2006) but have not identified how the charge was calculated.

C3.11 The fixed charge is neutral between operators and provides no incentive either for the S-bahn trains to omit stops at Liesing, Atzgersdorf and Hetzendorf or for longer-distance trains to add stops or slow down.

C3.12 ÖBB’s Network Statement (4.4.1.2) also sets the priorities set out in the table below.

PRIORITIES FOR PATH REQUESTS IN AUSTRIA

Priority	Designated as congested	Not designated as congested
1	Paths based on a framework agreement, where this path is necessary to fulfil the commercial conditions laid out in the framework agreement (regularity/frequency, journey duration, connections)	
2	Train paths for “not-for-profit social services” during peak traffic times	Paths for fixed-cycle traffic and paths crossing a national border
	Other paths, prioritised according to: <ul style="list-style-type: none">• the social value of their services• freight traffic, in particular international freight traffic, prioritised over passenger traffic	Other paths

C3.13 On infrastructure designated as congested, currently only Mödling to Wien Meidling, the priority is framework agreements, then local and regional commuter services, then international freight, domestic freight and finally other passenger services. Again, the supplement for congested infrastructure may reinforce this prioritisation and may deter some requests for paths low in the priority order. However, of those willing to pay the charge, a domestic or international freight operator willing to pay little would be prioritised over a high value passenger operator willing to pay more.

C3.14 This raises the fundamental issue of whether the aim is a “bottom up” solution, in which prioritisation solely reflects the value declared by operators on the day, or a “top down” solution, in which prioritisation is based on published rules determined in advance and implicitly based on value estimated in advance.

C3.15 ÖBB also defines priorities for infrastructure not designated as congested, although these are not covered by Article 22 and would only be covered by Article 24 on infrastructure which had, after consultation, been designated as specialised infrastructure. The legal basis for applying these priorities is therefore unclear. The priority is framework agreements, then regular interval and international services. The “incentive for capacity optimisation” may reinforce this prioritisation and may also deter some requests for paths which are low in the priority order.

C4 BELGIUM (INFRABEL)

Current charging structure (Network Statement 2013)

C4.1 The table below summarises the charging structure set out in Infrabel’s Network Statement 2013.

CURRENT CHARGING STRUCTURE IN BELGIUM

Article		Comments
7.4 Scarcity charge	✗	No reference to scarcity charge
Other charges which relate to “periods of congestion”	✓	Access charges by section have factors relating to: <ul style="list-style-type: none"> • (“H”) whether peak, semi-peak or normal • (“T”) divergence of speed from standard path
12 Reservation charge or cancellation charge	✓	Administrative costs of reservations are charged Cancellation charge tiered by notice given
22 Congested infrastructure	✗	None currently declared (Annex B.6)
Actual charges on “congested infrastructure”	✗	“There are currently no plans for a specific charge for the utilisation of congested infrastructures.”

Article 7.4 scarcity charges

C4.2 Infrabel’s network is subdivided into a number of sections which are listed in Annexe C.5 of the Network Statement. The sections are defined by operational criteria over which the charging department has no influence.

C4.3 Infrabel applies a structure of track access charges inherited at its creation in 2005. Track access charges for each section are multiplied (or “marked up”) by two parameters which jointly have the effect of relating charges to the utilisation of each section of the infrastructure:

- The “H” parameter: 1 in normal hours, 3 in semi-peak hours and 4 in peak hours
- The “T” parameter, which rises with the discrepancy between journey time over the section and a standard journey time

C4.4 The “H”-parameter can be seen as related to the **scarcity of capacity**, as it is based on the number of trains per hour by direction in a reference year. Annex F.5 of the Network Statement lists the periods in which each section of infrastructure is peak (Category 1) or semi-peak (Category 2). Infrabel informed us even at peak times it is normally possible to operate additional trains: the capacity of the network is not exhausted and it has been possible to satisfy all applications for access. The classification remains stable from year to year as changes to use of the network have been gradual.

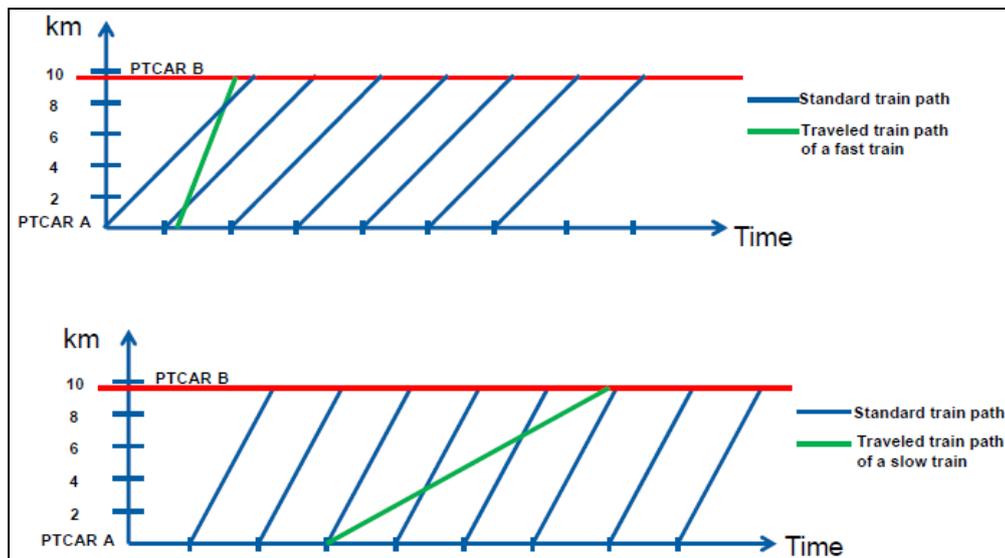
C4.5 Higher charges at peak and semi-peak times incentivise operators, and particularly freight operators, to avoid locations, periods and directions of peak demand.

C4.6 The “T”-parameter can be seen as at least partly related to the measures of **consumption of capacity** examined by Railtrack in the 1990s (see B7.9).

Final Report

- C4.7 It is derived from a standard train path based on the time required to transit each section at the line speed, less an allowance for normal deceleration and acceleration (for example, where the line speed is 220 km/h, the transit time might be calculated on the basis of an average speed of 180 km/h). Each train is therefore penalised for travelling either faster or slower than the standard path, as shown in the figure below.

“T-PARAMETER” IN CHARGES IN BELGIUM



Source: Infrabel

- C4.8 For passenger trains, the “T” parameter is 1 for transit times within 5% of the standard path, and then rises by 0.15 for each additional 5% deviation. This provides operators with at least some incentive to harmonise their timetabled speeds as closely as possible with the standard path.
- C4.9 For freight trains, different parameters are used: Infrabel advised us that on a mixed use railway it would not be practicable to increase charges to freight trains in the same way as passenger trains. The standard path can also be different on fast and slow lines where there are multiple tracks, encouraging segregation of services according to speed.
- C4.10 Infrabel retains the income from charges, so operators are incentivised to minimise the effects of both “H” and “T” parameters. Infrabel’s charges are only intended to cover operational costs: charges resulting from the “H” and “T” parameters are neither “ring-fenced” for investment in the location in which they arise nor used to inform the investment planning process.
- C4.11 Infrabel informed us that the overall structure of charges was under review, and might be modified in 2017, inter alia to reflect the possible liberalisation of domestic services. As yet no specific decisions had been taken on the future of the “H” and “T” parameters.

Article 12 reservation charges

- C4.12 Infrabel charges for the administrative costs of reserving paths but this is related to costs and is not a reservation charge. The charge is low and the workload is

broadly similar whether a path is required once, or every hour throughout the year, so the cost is most material to freight operators requesting individual (or small numbers of) paths at short notice.

- C4.13 Infrabel also imposes cancellation charges for unused paths. The charge is zero for paths cancelled 60 days in advance but rises to 100% for paths cancelled at less than 24 hours' notice, which cannot be reallocated.

Article 22 congested infrastructure and prioritisation

- C4.14 Infrabel defines priority criteria for congested infrastructure but has not declared any capacity to be congested. Even at peak times it could accept additional trains, and it has completed all its timetabling within the coordination and consultation mechanisms provided for in Article 21. The process is based on application by April and allocation by August in each year.
- C4.15 Potentially the most congested part of the network is the Brussels “junction”, the main north-south route through the centre of the capital. However:
- Freight services are not permitted to use this route on safety grounds
 - Eurostar services operate on separate high speed lines
 - Thalys and former Fyra services were managed by SNCB
 - All other services are operated by SNCB, the sole domestic operator
- C4.16 In effect, SNCB manages applications for capacity on behalf of all the operators needing to use the infrastructure, and it has to date not made any applications which cannot be satisfied. We note that this situation might change if the domestic passenger market were to be liberalised as proposed in the Fourth Railway Package.

C5 GERMANY (DB NETZ)

Current charging structure (Network Statement 2014)

- C5.1 The table below summarises the charging structure set out in DB Netz's Network Statement 2014.

CURRENT CHARGING STRUCTURE IN GERMANY

Article		Comments
7.4 Scarcity charge	×	No reference to scarcity charge
Other charges which relate to “periods of congestion”	×	Former “Auslastungsfaktor”, now withdrawn, had no linkage to “periods of congestion”
12 Reservation charge or cancellation charge	✓	Reservation charge for optional paths (6.2.5.2) Cancellation charge tiered by notice given
22 Congested infrastructure	✓	8 sections, plus 2 which may soon become congested
Actual charges on “congested infrastructure”	×	None at present

Final Report

Article 7 scarcity charges

- C5.2 DB Netz does not currently apply any scarcity charges.
- C5.3 In 2001 DB Netz introduced an “Auslastungsfaktor” or “utilisation factor” claimed to be an “incentive aiming at increasing the performance of the network” according to a section of the relevant law based on Article 7.4. This took the form of a 20% increase in access charges on certain sections of route. However, industry observers concluded that:
- There was no explanation of how the factor had been calculated, and its application provided potential for discrimination.
 - The factor, applied in 2001 to 71 sections of route or 2,377 kilometres (reduced in 2006 to 9 sections of route or 361 kilometres) included sections not declared as congested but excluded sections declared as congested.
 - The factor did not relate either to particular times or to measures such as utilisation which vary over time, and hence was not consistent with the Article 7.4 reference to “periods of congestion”.
 - The factor did not effectively incentivise rerouting, as alternatives to the routes on which it was applied were often poor substitutes with considerably longer journey times and hence higher operating costs.
 - The factor did not result in any reduction in traffic on the routes to which it was applied: in most cases there was significant traffic growth.
 - There was no evidence that DB Netz had used the additional charges revenue to invest in capacity enhancements.
- C5.4 In November 2006 newly-created network regulator, the Bundesnetzagentur (BNetzA) challenged elements of DB Netz’s Network Statement including the Auslastungsfaktor, but its findings were contested by DB Netz. In July 2011 the BNetzA published a final decision declaring the Auslastungsfaktor to be invalid, to which DB Netz initially objected. In November 2011 DB Netz agreed to withdraw the factor, which was removed from charges from the beginning of the December 2012 timetable period.
- C5.5 In the preamble to its decision, the BNetzA’s put forward four arguments against the Auslastungsfaktor:
- Lack of consistency with the legal requirement for “scarcity of capacity [] during periods of congestion” derived from Article 7.4.
 - Lack of objective justification for the charges, leading to discrimination.
 - Lack of incentives for efficiency. We note that this is not specifically referred to in Article 7.4, which requires only that capacity is scarce.
 - The view that scarcity charging could only be levied on sections of route where a capacity enhancement plan exists and is being implemented. This differs from both our interpretation of the legislation (see C1.3) and the way in which “scarcity-related” charges have been applied by ÖBB or Infrabel.
- C5.6 While details of DB Netz’s income from the Auslastungsfaktor have not been published, the BNetzA’s November 2011 press release indicated that the market in rail services would experience an overall cost reduction of €13 million after its withdrawal.

- C5.7 Further changes to charging are being considered in Germany but these are unlikely to include a scarcity charge as defined in Article 7.4.

Article 12 reservation charges

- C5.8 DB Netz sets a reservation charge for optional paths, and in the case of cancellation applies a charge tiered by the amount of notice given. If 60 days' notice is given, the cancellation charge is limited to the charge for developing the train path, but with less notice is given up to 40% of the base price for the train path is also charged.

Article 22 congested infrastructure and prioritisation

- C5.9 DB Netz's Network Statement 2014 declares a number of sections of infrastructure to be congested or probably congested in the near future, as shown in the table below.

CONGESTED INFRASTRUCTURE DECLARED IN GERMANY

Status	Infrastructure	Notes
Congested	5200 Gemünden (Main)-Würzburg	Two consecutive sections of route totalling over 120km
	5910 Würzburg-Fürth (Bayern)	
	5900 Fürth (Bayern)-Bamberg	
	4000 Offenburg-Gundelfinger	
	4000 Leutersberg-Weil am Rhein	
	3600 Hailer-Meerhoz-Fulda	
	1720 Uelzen-Stelle	
	Station Berlin-Spandau	Station
Probably congested in the near future	1210 Niebüll-Westerland (Sylt)	Two-track line on causeway
	Hamburg Hauptbahnhof	Station

- C5.10 We note that two elements of congested infrastructure are not sections of line but stations, at Berlin Spandau and the Hamburg Hauptbahnhof. We are aware from previous studies that Hamburg Hauptbahnhof was already a major operational constraint in 2006, when services through Schleswig-Holstein from the holiday island of Sylt on the Nord-Ostsee-Bahn could not be accommodated there.
- C5.11 Priority rules for each section of congested infrastructure are included in Annexes to the Network Statement. The proposed priority rules for Hamburg Hauptbahnhof, in the event that it is declared congested, are summarised in the table overleaf.

CONGESTED INFRASTRUCTURE: PRIORITIES IN HAMBURG HAUPTBAHNHOF

Higher priority	Lower priority
Local trains which fit a half platform	Local trains which do not fit a half platform
Trains requiring no change of locomotive	Trains requiring change of locomotive
Trains requiring no change of formation	Trains requiring a change of formation
Reversing trains continuing a service	Reversing trains forming separate services

C5.12 We noted, in the indicative list of means of increasing capacity set out in Table 2.2 of the main report, that using longer trains may not allow two trains to occupy the same platform, but not the other points. However:

- The second and third points could be consolidated into a single requirement, that the operator use fixed-formation trains
- The fourth point relates more to real time operation than to timetable planning

C5.13 The proposed priorities for the Hamburg Hauptbahnhof are not yet in use and have not been subject to industry or regulatory scrutiny or tested in practice. They nonetheless imply that, if stations constrain capacity, it may not be possible to define its use and consumption with measures used on sections of route.

C5.14 DB Netz also sets priority rules, which do not appear to be restricted to congested infrastructure: regular-interval or integrated network services, then cross-border train paths, then train paths for freight traffic.

C5.15 Where the priority rules do not resolve a conflict, capacity is awarded to the applicant generating the higher charge. For conflicting ad hoc applications, priority is given to the first applicant.

Article 24 special infrastructure

C5.16 DB Netz declares a number of elements of “special infrastructure”, including high speed lines, and in some cases prioritises their use for passenger trains by day and freight trains by night.

C6 SWEDEN (TRAFIKVERKET)

Current charging structure (Network Statement 2014)

C6.1 The table overleaf summarises the charging structure set out in Trafikverket’s Network Statement 2014.

Article 7.4 scarcity charges

C6.2 The Network Statement makes no reference to scarcity charges but states (6.1.1) that “A passage charge is paid during peak hours on normal working days for passages to and from Stockholm, Göteborg and Malmö. The charge is a special charge.” We note that “special charge” has no defined meaning in Directive 2001/14. Trafikverket informed us that, while the charge is “scarcity-related”, it is not referred to as a scarcity charge within the meaning of Directive 2001/14 Article 7.4 and is considered a “mark-up” under Article 8.

CURRENT CHARGING STRUCTURE IN SWEDEN

Article		Comments
7.4 Scarcity charge	✘	No reference to scarcity charge
Other charges which relate to “periods of congestion”	✓	“Passage charge” 5 sections (Stockholm, Göteborg, Malmö) SEK 260 per train path per section
12 Reservation charge or cancellation charge	✘	No reference to reservation charge Train paths and “passage charge” (see below) are based on allocated, non-cancelled capacity
22 Congested infrastructure	✘	None currently declared
Actual charges on “congested infrastructure”	✘	No reference to charges for congested infrastructure

- C6.3 The passage charge was introduced in 2011 at a level of SEK 150 per train path but has since been increased to SEK 260 per train path. While some other charges are based on measures of costs, the passage charge is not, and is intended to act as an incentive, and its effects are monitored as a basis for possible adjustment. Trafikverket stressed that it does not have to seek prior approval from its regulator before imposing a new charge, but the charges have not been challenged and continue to evolve. It considers that the charges remain below the costs permitted under Article 7.2.
- C6.4 A background report on charging proposals for 2015 estimated that the 2013 passage charges were incurred 200,000 times by passenger trains and 14,600 times by freight trains. This suggests a total income of around SEK 56 million (around £5 million) or only around 4% of total forecast track access charges in 2015. The background report also analysed the expected effect of the 2015 charges in detail, and in particular:
- Anticipated some traffic moving from high fee to low fee routes. (It is not clear how relevant this potential effect would be in Great Britain with, for example, freight diversionary routes.)
 - Claimed that “Similarly tolls in big cities mean that trains are moving their departure and/or arrival outside the hours charged”. (Given the number of regular interval local services specified and supported by the Swedish county authorities, we suspect that this relates primarily to long-distance SJ, open access passenger and freight operators.)
- C6.5 However, Trafikverket confirmed that, while these effects would occur in principle, none had been observed at the relatively low levels of charges to date.
- C6.6 The background report also showed that, in line with Sweden’s policy of marginal cost pricing of transport infrastructure, total infrastructure charges form only a small proportion of operating costs (4-11%) or fares (3-11%). From this it also derives an estimate of the effect of charges on demand and hence the socio-economic cost of raising charges. The passage charge only represents 4% of total

Final Report

track access charges, or much less than 1% of either operating costs or fares over the network as a whole.

- C6.7 This raises the issues of whether charges imposed on operators would be passed through to final customers. In Great Britain the scope to do this would be limited by fares regulation, although there might be a case for passing scarcity charges to passengers, and particularly peak period users including commuters, to signal the cost of their journeys.
- C6.8 We also reviewed a consultation document on the 2015 charges which provided more detail on the charging structure, summarised in the table below.

“PASSAGE CHARGE” IN SWEDEN

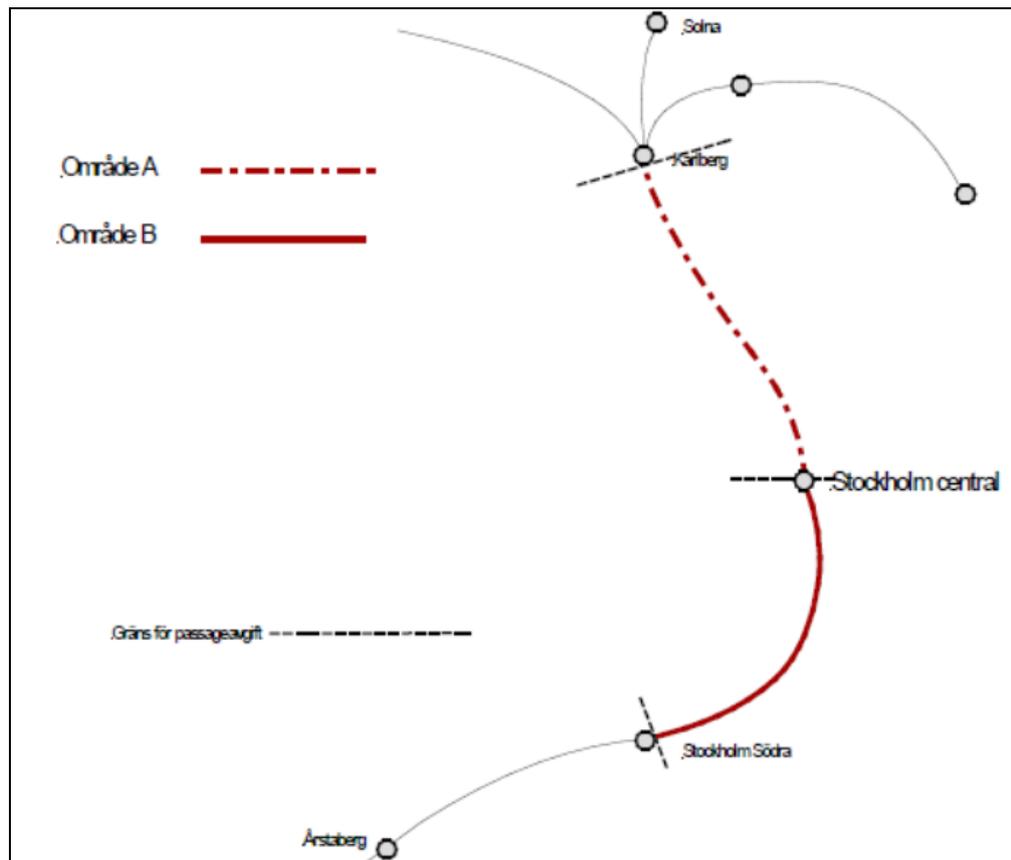
Feature	Details
Locations	Stockholm (2), Göteborg (1), Malmö (2), as described below
Times of application	Any train in the defined location 07:00-09:00 or 16:00-19:00 on weekdays. We have not identified any exception for periods such as public or school holidays).
Level of charge	SEK 260 per location. The charge is not-distance based and is the same irrespective of train type, speed or stopping pattern.

- C6.9 Trafikverket informed us that the three main city areas had been selected initially as most traffic originates and terminates there, and that this would remain the case for the foreseeable future, although the intention was to align the charges with capacity utilisation as measured by UIC standards.
- C6.10 The charging areas were defined so as to capture all traffic to and from the three cities, although it was also necessary to align their boundaries with points at which trains were “registered” (analogous to Network Rail’s recording points). The current definitions are not considered optimal and are likely to be reviewed, developed and altered over time.

Stockholm

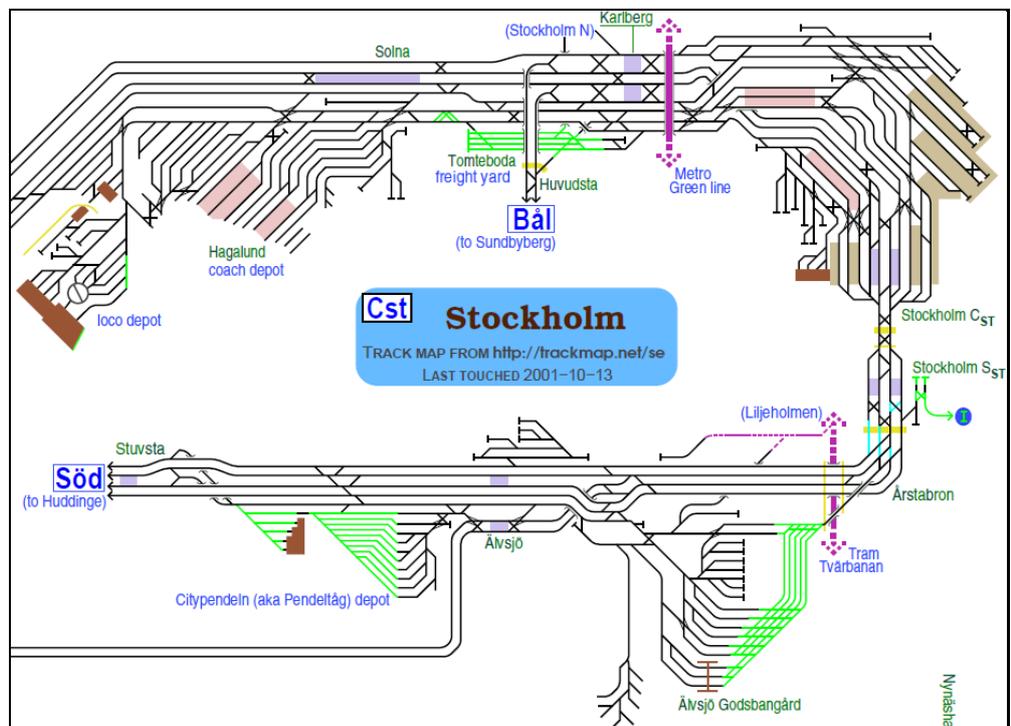
- C6.11 In Stockholm, the charge is levied for two sections of track, between Karlberg and Stockholm Central and Stockholm Central and Stockholm Södra (South) as shown in the first figure overleaf. Distances are around 2 kilometres from Karlberg to Stockholm central and a further 2 kilometres from Central to Södra, with no intermediate stations.
- C6.12 The second figure overleaf shows the most well-known constraint, the two track section between Central and Södra, an historic two track bottleneck over the “Strömbroarna” bridges across the outflow from Lake Mälaren.
- C6.13 The charge is levied independently on the two sections of track, presumably reflecting both the differences between their characteristics and the number of trains which reverse in Stockholm Central. A train passing through both sections during the charging period pays both charges (and we note that the limited number of platforms available in Stockholm Central may limit operators’ scope to hold trains there to avoid the charging period).

“PASSAGE CHARGE” IN STOCKHOLM



Source: Trafikverket

“PASSAGE CHARGE” IN STOCKHOLM: TRACK LAYOUT



Source: trackmap.net/se, 2001, note that the detailed track layout may have changed

Aspedalen, Aspen, Jonsered, Partille and Sävenäs). The line south is in tunnel, including an underground station at Liseberg (not shown), as far as the portal at Almedal. These lines are two track, but the line north appears to be single track over a short distance between Marieholm and Göteborg-Kville.

C6.20 The charge appears to be applied once for any train entering the highlighted network. There is no differentiation according to:

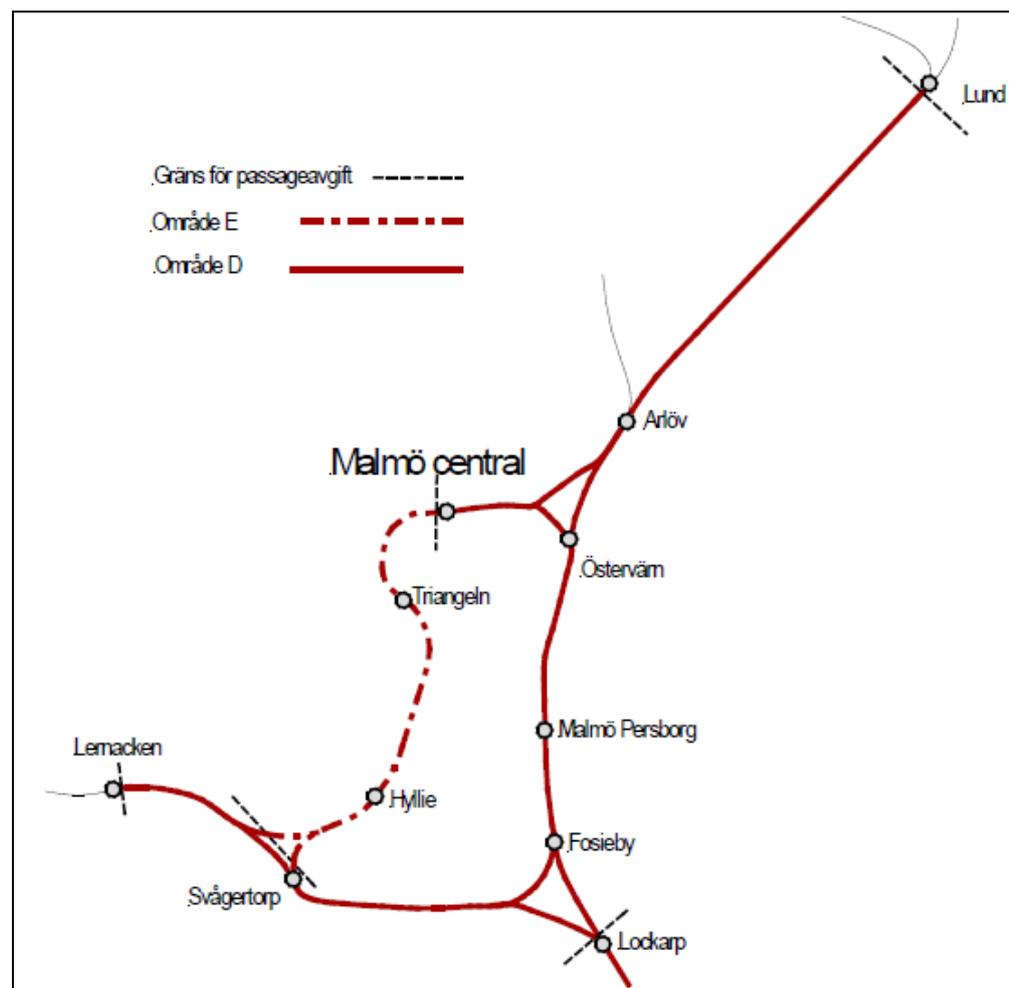
- Type or speed of train
- Stop (Stenkullen/Lerum/Aspedalen/Aspen/ Jonsered/Partille/Sävenäs, Liseberg)
- Manoeuvres, such as entering and leaving Göteborg Central on different routes

Malmö

C6.21 In Malmö, the charge is levied for all tracks in two areas shown in in the figure overleaf:

- Area E, between Malmö Central and Svågertorp (CityTunnel, opened in 2010)
- Area D, between Malmö Central and Lund, Lockarp, Svågertorp and Lernacken

“PASSAGE CHARGE” IN MALMÖ



Source: Trafikverket

Final Report

- C6.22 If both areas are used during the charging period, a maximum of two tolls are paid, as in Stockholm.
- C6.23 Area D comprises two-track line with stations at Lund (and then Hjärup, Åkarp and Burlöv), Malmö Central, Persborg and Svågertorp (Malmö Syd), with the next station beyond Lernacken being Kastrup in Denmark. Distances are around 16 kilometres from Lund to Malmö Central and a further 18 kilometres from Central via Fositeby to the shore at Lernacken.
- C6.24 Area E is 17 kilometres long, with four tracks through Hyllie and Malmö Central stations but two tracks in the central tunnelled section including the station at Triangeln. It appears that a train travelling from Lernacken, reversing at Hyllie and continuing to Svågertorp would pay the Area E charge even without entering the two track section.
- C6.25 Trafikverket informed us that, despite the CityTunnel (area E) opening in 2010, the historic line (area D) remains congested, with 480 trains per day between Malmö and Lund, although the need to include the southern branch at Lockarp, which leads to the port of Trelleborg, may be reviewed.

Article 12 reservation charges

- C6.26 Trafikverket does not apply reservation charges. Charges are “based on allocated, non-cancelled capacity”.

Article 22 congested infrastructure and prioritisation

- C6.27 Trafikverket informed us that they have not yet declared any infrastructure to be congested, although they have made provisions to do so as required under Article 22.
- C6.28 The Network Statement states (4.4.3) that “If a conflict of interests was not resolved during coordination or dispute resolution, the Swedish Transport Administration shall declare the relevant part of the infrastructure to be “congested”. The Swedish Transport Administration notifies the applicants of the decision and publishes it on the Administration’s website.”
- C6.29 The Network Statement also states (4.4.3) “If the infrastructure is declared congested, a capacity analysis and a capacity reinforcement plan shall be developed” and (4.4.6) “The capacity reinforcement plan also contains a cost and benefits analysis for possible measures, information about the measures which the Swedish Transport Administration intend to implement on the basis of this analysis, and a schedule for this work. The schedule encompasses a maximum of three years. The measures that are analysed and proposed may relate to the infrastructure, modifications to train paths or to the railway undertakings’ vehicles and carriages.” In practice it is not always possible to expand capacity within three years, as the longstanding bottleneck in central Stockholm shows.
- C6.30 The Network Statement (4.4.3) states that when infrastructure is declared to be congested, the timetable will be determined using priority criteria listed in an Appendix.
- C6.31 However, “the basic principle of [Trafikverket’s] prioritisation criteria is to choose the resolution of conflicts between applicants that provide the greatest economic

value.” This is consistent with the provisions of Article 22.4 “The priority criteria shall take account of the importance of a service to society, relative to any other service which will consequently be excluded”.

- C6.32 Trafikverket does not set priorities based on the type or classification of trains, but instead uses a model incorporating a number of simplifications and assumptions. We understand that this approach is broadly analogous to ORR’s use of MOIRA and other demand and revenue modelling tools.

C7 SWITZERLAND (SBB INFRASTRUKTUR)

Current charging structure (Network Statement 2014)

- C7.1 The table overleaf summarises the charging structure set out in SBB Infrastruktur’s Network Statement 2014.

CURRENT CHARGING STRUCTURE IN SWITZERLAND

Article		Comments
7.4 Scarcity charge	×	No reference to scarcity charge (except for services)
Other charges which relate to “periods of congestion”	×	Provision for scarcity prices for “keeping paths available for optional trains”, but this is a service
12 Reservation charge or cancellation charge	✓	No reference to reservation charge Cancellation charges apply 30 days after definitive allocation
22 Congested infrastructure	×	Provision, but no reference to congested routes
Actual charges on “congested infrastructure”	×	No reference to charges for congested infrastructure

Article 7 scarcity charges

- C7.2 SBB Infrastruktur’s Network Statement makes no reference to a scarcity charge for capacity allocation (Section 4) but does so for the provision of ancillary services (Section 5).
- C7.3 It states that “The infrastructure manager defines the prices of the following ancillary services without discrimination, insofar as these can be provided with the existing infrastructure and available personnel, and publishes these: a. Keeping paths available for optional trains; b. Track allocation in the event of a delay requested by the network user and not caused by scheduled traffic; c. Stabling of train compositions; [] f. Use of loading tracks and facilities; []”.
- C7.4 In addition “The prices covered by paragraphs 1a-c and f are to be set as scarcity prices as a function of demand and investment value on a location-by-location basis. All other prices are to be set analogously in accordance with the principles laid out in Article 19. In addition, a pro rata element can be added for capital and amortisation costs in respect of assets used mainly in the provision of ancillary services.”

Final Report

- C7.5 In summary, the Network Statement sets a precedent for scarcity charges in relation to:
- Keeping paths available for optional trains
 - Track allocation in the event of a delay requested by the network user and not caused by scheduled traffic
 - Stabling of train compositions
 - Use of loading tracks and facilities
- C7.6 We have not attempted to identify whether and how this principle has been applied in practice to these services.

Article 12 reservation charges

- C7.7 SBB Infrastruktur allows a grace period of 30 days after definitive train path allocation, after which the operator (not the applicant) will be liable to pay compensation. This differs from networks where cancellation is permitted during the timetable period, with charges tiered according to the notice given.

Article 22 congested infrastructure and prioritisation

- C7.8 SBB Infrastruktur prioritises regular timetable passenger services for track access, and these are always accommodated (we assume that any conflicts are internalised in the process of devising the public timetable specification), but sets the following priorities after “a declaration of congestion”:
- A Freight trains, to which, for technical reasons, in particular the critical clearance gauge, no alternatives can be offered
 - B Freight trains which, for geographical reasons, are less easy than others to divert onto alternative routes
 - C Trains which run more than once during the timetable year, depending on how frequently they run
 - D Other trains: a two-stage bidding process will be held (the winner will pay the second-highest bid plus a maximum of CHF 1,000.00)
- C7.9 We have not identified whether the bidding process has ever been used.

C8 OTHER EXAMPLES

- C8.1 In our review of Network Statement we identified a number of other features which may be relevant to this analysis.
- C8.2 In the Czech Republic (SZDC), as in Switzerland, there is provision for an auction of capacity if neither coordination or consultation resolves all disputes, but we have not identified whether a detailed auction process has been devised or used.
- C8.3 In Denmark (Banedanmark), once infrastructure has been declared congested, applicants must prioritise their own applications, thereby signalling to infrastructure manager Banedanmark additional information on the relative value of different requests by the same applicant. One effect may be to create a disincentive to making multiple or speculative applications.
- C8.4 In Portugal (REFER), the infrastructure manager reserves the right to hold back capacity for foreseeable ad hoc requests: it is not obliged to allocate all the

capacity available. It also sets priorities which vary by time of day, with domestic and international freight at night, suburban passenger services in the commuter peaks, and international or interurban passenger services at other times.

C9 SUMMARY

Comparison of charging structures

C9.1 The table below compares the current charging structures of the five networks examined with those of Network Rail's Network Statement 2015. For illustrative purposes we have interpreted Network Rail's capacity charge as neither limited to infrastructure which has been declared congested nor specifically related to "periods" of congestion.

CURRENT CHARGING STRUCTURES COMPARED

Article	Network Rail	Austria (ÖBB)	Belgium (Infrabel)	Germany (DB Netz)	Sweden (Trafikverket)	Switzerland (SBB Infrastruktur)
7.4 Scarcity charge	x	x	x	x	x	x
Other charges which relate to "periods of congestion"	x	✓	✓	x	✓	x
12 Reservation charge or cancellation charge	x	✓	✓	✓	x	✓
22 Congested infrastructure	✓	✓	x	✓	x	x
Actual charges on "congested infrastructure"	x	✓	x	x	x	x

Article 7 "scarcity charges"

C9.2 None of the networks describe any of their charges as a scarcity charge.

C9.3 However, we identified four charges which in practice reflect "the scarcity of capacity of the identifiable segment of the infrastructure during periods of congestion" as set out in Article 7.4:

- ÖBB's "supplement for congested infrastructure" and the "incentive for capacity optimisation"
- Infrabel's combination of "H" and "T" parameters in track access charges
- Trafikverket's "passage charge"

C9.4 Only ÖBB's "supplement for congested infrastructure" is contingent on the Article 22 test that "is not possible to satisfy requests for infrastructure capacity adequately" and hence the infrastructure has been congested. The other charges are imposed before this point is reached, and Infrabel informed us that they

Final Report

explicitly expected to have at least some spare capacity at locations and times where the charges are applied.

Article 12 “reservation charges”

- C9.5 All of the networks except Trafikverket apply either reservation charges payable in advance or, more commonly, cancellation charges payable even if the capacity is not used. The effect of these charges will clearly depend on whether cancellation is permitted and how much notice is required to escape the charge: except in Switzerland, an efficient operator can avoid or minimise cancellation charges by giving adequate notice.

Article 22 “congested infrastructure”

- C9.6 Only ÖBB and DB Netz have declared infrastructure to be congested.
- C9.7 This also means that (according to Article 20) only ÖBB and DB Netz are entitled to “employ priority criteria to allocate infrastructure capacity” and only on the congested infrastructure.
- C9.8 Only ÖBB has identified a charge specific to congested infrastructure, the “supplement for congested infrastructure”.

The incentives to operators provided by scarcity-related charges

The table below repeats from Table 2.2 of the main report the means of increasing capacity open to operators acting unilaterally, and identified which of the scarcity-related charges have, in theory at least the potential to incentivise each of them.

- C9.9 We discuss further below whether there is any evidence that such incentives have in fact influenced behaviour.
- C9.10 Network Rail’s capacity charge varies by location, and hence may incentivise rerouting, but not by time of day, and therefore cannot incentivise retiming.
- C9.11 DB Netz’s “Auslastungsfaktor” is not clearly related to “scarcity” but can also in principle incentivise operators to reroute or withdraw trains but not to retime them to avoid “periods of congestion”.
- C9.12 ÖBB’s and Trafikverket’s charges can also incentivise retiming.
- C9.13 Only Infrabel’s charges have the potential to incentivise measures to consume capacity en route efficiently: slowing trains down, adding stops, having spare trains or buying high performance or electric trains.
- C9.14 None of the scarcity-related charges can, on its own, incentivise a complete recast of the timetable coordinated across all operators. In the case of Infrabel, we note that it appears that SNCB could in practice recast the entire passenger timetable and present it as an internally consistent and workable request for capacity.

MEANS OF INCREASING CAPACITY BY OPERATORS UNILATERALLY

Means (For further detail, see Table 2.2)	Example						Delivery timescale	Investment	
	Network Rail Capacity charge	DB Netz "Auslastung"	ÖBB "supplement"	ÖBB "incentive"	Trafikverket "passage"	Infrabel "H" and "T"		Operator	Infrastructure
Withdraw train	✓	✓	✓	✓	✓	✓	Short	x	x
Reroute train	✓	✓	✓	✓	✓	✓	Short	x	x
Retime train	x	x	✓	✓	✓	✓	Short	x	x
Slow trains down	x	x	x	x	x	✓	Short	x	x
Add stops	x	x	x	x	x	✓	Short	x	x
High density trains	✓	✓	✓	✓	✓	✓	Short	✓	x
More/spare trains	x	x	x	x	x	✓	Medium	✓	x
Longer trains	✓	✓	✓	✓	✓	✓	Medium	✓	x
High performance	x	x	x	x	x	✓	Medium	✓	x
Electric trains	x	x	x	x	x	✓	Medium	✓	x
Fixed formations	x	x	x	x	x	x	Medium	✓	x
Shorter trains	x	x	x	x	x	x	Medium	✓	x

Note: reservation or cancellation charges will also slightly incentivise withdrawal

The principal means of solving capacity allocation problems

C9.15 We set out in Section C1 three approaches to allocating infrastructure capacity within the framework of the Directive. Within this terminology:

- All the networks use "cooperative" approach of Article 21
- Most of the networks use charges related in some way to scarcity which incentivise operators to make "bottom up" decisions on where and when they should apply for capacity
- Only ÖBB is entitled to use the "top down" approach of priority criteria, and then only on the parts of the Network it has declared to be congested, where it also applies an additional charge

The ability of charges to allocate capacity in "periods of congestion"

C9.16 Auction theory suggests that open ascending price auctions, with no reserve price, can ensure that all available capacity is allocated to the user that values it the most. SBB Infrastruktur has made provision for auctions where infrastructure has been declared congested, but no SBB infrastructure has been declared congested. All the other approaches we have identified are analogous to reserve prices, which

Final Report

may deter some bidders but may also result in some capacity remaining unallocated.

- C9.17 A fundamental issue is that scarcity charges set in advance can only act as a reserve price, and cannot ensure that all available capacity is allocated. This is clearly the case with Infrabel, which explicitly expects that some capacity will remain unallocated. If the objective of scarcity charges is to ensure that all capacity is allocated, then setting a reserve price is potentially counterproductive.

Evidence of the effect of charges on operator decisions

- C9.18 ÖBB's Network Statement suggest that scarcity-related charges have an effect on operator applications, but we have not been given any specific examples of where this is the case.
- C9.19 Infrabel informed us that the scarcity-related "H" and "T" charges had influenced freight operators to avoid requesting paths at peak periods, and that a market mechanism might be more effective than a priority system because in some cases it was commercially necessary for freight trains to operate at congested locations and times. However, Infrabel also pointed out that it was not considered practicable to charge freight operators on the same basis as passenger operators, even if they consumed the same capacity on the same track.
- C9.20 Infrabel noted that SNCB was replacing much of its fleet with higher performance trains, which might result in a reduction in charges due to the "T" parameter, although it is not possible unambiguously to attribute the change to the charging structure.
- C9.21 Trafikverket informed us that their assessment was that so far charges had had no impact on operator applications, but that this was unsurprising while charges remain relatively low. We agree that the limited extent to which charges influence behaviour may reflect the relatively low absolute levels of charges.

Evidence of the effect of charges on infrastructure manager decisions

- C9.22 Infrabel informed us that the charging and investment planning functions were separate and that they were not aware of any case of charges directly influencing investment decisions. We have not identified any evidence in other networks of any linkage between scarcity-related charges and investment decisions.

Priority systems

- C9.23 At first sight, priority systems, such as that used by REFER, can be used to incentivise operators to reroute, retime or withdraw their services in the same way that charges can. We note, however, the example provided by Infrabel, that a freight service can operate at peak times and in the peak direction, provided that it is willing to pay the high charges which result. Priority systems can reflect ex ante estimates of the typical value of a type of train, but not the specific value of an any individual train, particularly for freight.

Provisional view of options which might be considered for Network Rail

Incentives to avoid periods of congestion

- C9.24 International experience suggests at least two approaches to how scarcity charges could be set and structured.
- C9.25 One approach would be to focus on particular nominated locations, as with the ÖBB's "supplement for congested infrastructure" and Trafikverket's "passage charge", although it is not yet clear whether and how such an approach could be applied where the constraint was a junction or station (as at Berlin Spandau or, potentially, Hamburg Hauptbahnhof). Our experience and preliminary research suggests that a difficulty of such a charge in Great Britain might be identifying unambiguously both the location of the bottleneck "identifiable segment of the infrastructure" and hence the operators and services which should bear the charge. In this context, we note how Trafikverket expected to need to alter the definition of the charging locations over time, which would implicitly affect the incidence and distribution of charges.
- C9.26 Another approach would be applied over relatively large parts of the network, as with ÖBB's "incentive for capacity optimisation" and Infrabel's "H" and "T" parameters. Again, it is not yet clear whether and how such an approach could be applied where the constraint was a junction or station. Our preliminary research suggests that the principal difficulty of such a charge in Great Britain is that it would be incurred by many operators in many locations and, if revenue neutral, might have little effect if the Department for Transport paid most of the charge.

Incentives to consume less capacity per train

- C9.27 We illustrated above how Infrabel's "T" parameter, which acts as a measure of how much capacity is consumed by each train, can incentivise operators to minimise their consumption of capacity, at least as measured on each section of track to which the measure is applied.
- C9.28 In principle, it would be desirable for any scarcity charging in Great Britain also to reflect the consumption of capacity. However, we note that, on the Infrabel network:
- There remains a single dominant passenger operator which is able to internalise many timetabling decisions in its application for capacity.
 - It is not seen as practical to set charges for freight operators based on deviation from maximum line speeds, even if they are competing for capacity with passenger operators who are charged on this basis.
 - The "T" parameter approach to calculating consumption of capacity on a section of route would not be applicable where the constraint was a node or station, such as the Hamburg Hauptbahnhof.
- C9.29 As we noted in paragraph B7.10, studies for Railtrack in the 1990s identified a number of other practical barriers to implementing this approach on the British network. A major issue is that calculated consumption of capacity could be highly dependent on the definitions of the sections of line on which it was calculated, and potentially on timetabling decisions outside the control of the operator, and hence potentially appear arbitrary and/or volatile.

Final Report

APPENDIX

D

AIRPORT SLOTS AND RAILWAY PATHS AND DIAGRAMS

D1 INTRODUCTION

- D1.1 This Appendix examines the phenomenon of airlines trading in airport slots and identifies similarities with, and differences from, the need of railway operators for paths and diagrams.
- D1.2 Scarcity charging is widely seen as a means to improve efficiency within the network industries. In a network such as a pipeline or an electricity grid, the unit of consumption can be relatively self-contained and self-explanatory, such as:
- Transferring one litre/barrel/tonne of gas or liquid from one node to another
 - Transferring one kilowatt hour from one node to another
- D1.3 In each case, consumption by different potential users is substantially identical. As seen by the system, the ideal outcome, at times and places where capacity is limited, is to transmit the material whose customers are willing to pay most, which will normally be a function of the difference between their costs of production and the value placed on it by their customers.
- D1.4 Transport networks differ from pipelines and electricity networks in that the material being transported, passengers or freight, must be carried in vehicles which are expensive assets and must also be used efficiently. Depending on the relative capital and operating infrastructure and vehicle, efficient investment in, and use of, aircraft and railway rolling stock may be at least as important as efficient investment in, and use of, the associated network infrastructure.
- D1.5 In the remainder of this Appendix we:
- Compare airline and railway operations, and in particular the implications for prices in setting signals for all
 - Examine examples of capacity constraints on the rail network
 - Consider the practicalities of inferring opportunity cost from the rail timetable
 - Summarise the difficulty of measuring rail capacity

D2 COMPARISON OF AIRLINE AND RAILWAY OPERATIONS

- D2.1 The table overleaf compares a number of aspects of airline and railway operations, which we discuss below in turn:
- Typical patterns of operation
 - Minimum changes to timetables
 - Access requirements
 - Dominant constraints
 - Access negotiations
 - The ownership of access
 - The need for trading
 - The means of trading
 - The means of pricing
 - Improving allocation efficiency
 - Improving investment efficiency

TRADING BETWEEN AIRLINES AND BETWEEN TRAIN OPERATORS

	Airlines	Train operators
Typical patterns of operation	Round trip from base airport, although: <ul style="list-style-type: none"> ■ May overnight elsewhere ■ Round trip may be multi-day 	Series of elements between depots and stabling points: <ul style="list-style-type: none"> ■ Single journey ■ Round trip ■ Closed diagram ■ Sequence of open diagrams
Minimum change	Round trip	Short distance: round trip? Long distance: closed diagram?
Access requirements	Airports	Stations, stabling points, depots
	Airspace (ANSPs)	Sections of route, junctions
Dominant constraint	Runways	Sections of route Junctions Station platforms Stabling
Access negotiations	Domestic: 2 airports, ANSP International: 2 airports, ANSPs	Domestic: infrastructure manager International: infrastructure managers
Ownership of access	Perpetual “grandfather rights”	Time-limited model contract
Need for trading	A few congested airports, including Heathrow and Gatwick	
Means of trading	Sale Leasing on an agreed basis	
Mean of pricing	Bilateral agreement	
Allocative efficiency	Bilateral trades over time should allocate to highest value user	
Investment efficiency	High slot value provides evidence of latent demand	

Source: Steer Davies Gleave analysis, for details see text

Typical patterns of operation

Airlines

D2.2 The majority of aircraft operated by scheduled airlines operate from a base airport and the pattern of services they provide is typically one of out-and-back return trips to that base:

- Low cost carriers may have a number of bases, which are rarely congested, from each of which a small number of aircraft operate as many as five round trips per day but return to base overnight.
- Larger carriers tend to have a single base or hub, which may be congested at peak times. Short haul aircraft make a number of out-and-back trips but may make an evening departure, overnight stay, and early morning return. Long haul aircraft may be away from their base for as much as 48 hours, in the case of flights between Europe and Australasia, and are often en route overnight.

Train operators

- D2.3 Railway rolling stock may be based at a particular depot but may rarely be there overnight. It may begin the day at a depot, stabling point, or station, and be used to provide a number of train services before ending the day at the same, or another, depot, stabling point or station. The components making up a rail service can be seen on a number of different scales:
- A single journey from one station to another.
 - A round trip from one station to another, returning to the same station.
 - A closed diagram of a series of services, ending the day at the starting point.
 - An open diagram of a services of services, ending the day at a different point. A train may be operated on several open diagrams before it ends the day where the first diagram began.
- D2.4 On short distance services, such as an airport shuttle or a commuter route, a train may make a large number of round trips a day between the same two stations (“ABABABABABABABA”). On long distance service, such as for CrossCountry or East Coast, a train may make only two or three complete journeys in a day and not return to its starting point (“ABC”) which may only occur after several more diagrams (such as “CBDE”, “EBF”, “FBA”).

Minimum changes to timetables

Airlines

- D2.5 For airlines, the minimum element of service which could be removed from a timetable will typically be a single round trip from the base airport to a destination airport.

Train operators

- D2.6 For short distance rail services, the minimum element of service which could be removed from a timetable will also typically be a single round trip (“ABABABABABABABA” becomes “ABABABABABABA”).
- D2.7 For long distance rail services, however, it may not be possible to remove any single service or pair of services in a long sequence of open diagrams without substantial reworking of the timetable affecting a number of diagrams.

Final Report

Access requirements

Airlines

D2.8 Airlines wishing to operate a service require access to two types of infrastructure: terminals or airports, and airspace en route managed by air navigation service providers (ANSPs).

Train operators

D2.9 Train operators wishing to provide a service will make use of a number of types of infrastructure. Using the terminology of aviation, these are likely to include:

- Terminal: stations, stabling points and depots (including fuelling and cleaning)
- En route: sections of route, and junctions between routes

D2.10 With the exception of limited numbers of shuttle services, rail services use stations not only at the end of each journey but also at intermediate points.

Dominant constraints

Airlines

D2.11 As airline activity has grown there has been increasing congestion of both terminal and en route facilities.

D2.12 Airports provide a number of facilities, including gate and stand facilities, terminal capacity such as check-in desks, and car parking and surface access. However, in the short to medium term, airports can normally offer airlines a service, such as a remote stand, requiring bus transfers, rather than a contact gate, and fewer check-in desks than desired, potentially resulting in longer queuing or a need to ask passengers to check in earlier. Only in rare cases is an airport so constrained that detailed analysis is required to determine whether or not an additional or alternative flight can be accommodated.

D2.13 In most cases, therefore, any constraint at an airport is the access to runway slots. Take-off and landing “slots” are largely fungible and interchangeable:

- For take-off, the time each aircraft requires on the runway depends on a number of factors including aircraft type and fuel load: a fully loaded large aircraft may take longer to take off than a lightly-loaded small one. Subject to this constraint, however, aircraft may take off at any time and in any order.
- For landing, while all aircraft land at similar speeds, small aircraft cannot safely follow close behind large ones, so the spacing between descending aircraft varies slightly and can be optimised by grouping arriving aircraft in order of increasing size.

D2.14 Subject to these variations, however, it is generally possible to replace one aircraft using a slot with another.

D2.15 Airlines wishing to visit a congested airport do not require a single slot but rather a pair of slots separated by an interval in which the aircraft can be unloaded, reloaded and possibly refuelled. Both airline and airport have an incentive to keep this interval close to the minimum necessary: for the airline to make efficient use

of the aircraft, and for the airport to make efficient use of the stands or gates which the aircraft occupies while on the ground.

- D2.16 Airlines operating from a congested base airport will also require both departure and landing slots at times matched to form a complete round trip to the destination airport. However, airlines based at a congested airport are likely to hold a large number of slots and be able to redeploy them to different services to form such out-and-back patterns.
- D2.17 However, many airports have spare slots at all times of day, and many major airports continue to expand capacity with the aim of ensuring that there is little or no shortage of capacity, even at peak times.

Train operators

- D2.18 As railway infrastructure becomes congested, capacity may be constrained by infrastructure in a number of ways referred to in Table 2.2 of the main report:
- En route, where faster trains may catch and need to pass slower ones.
 - At junctions, where movements of different trains may conflict.
 - At stations, where trains need access to platforms for times varying from a less than a minute for through commuter trains to as much as an hour for terminating international services.
 - At stabling and other facilities (the future Thameslink fleet, for example, will require around 24 kilometres of track merely to stable the trains overnight).
- D2.19 There is no single dominant constraint analogous to airport slots.

Access negotiations

Airlines

- D2.20 An airline wishing to offer a new out-and-back service from a base will need to obtain capacity at its base and the other airport. It will also need to fly en route, either using a single ANSP in the case of a domestic flight or a number of ANSPs in the case of an international or long haul flight. Unless one or both airports is slot-constrained, however, an airline will not normally face any difficulties in obtaining the terminal and en route capacity required.

Train operators

- D2.21 An operator of domestic rail services will normally only need to deal with a single infrastructure manager, but that infrastructure manager will need to offer a complete set of workable and efficient diagrams compatible with the operator's desired service pattern and rolling stock. An operator of international rail services will need to deal with more than one infrastructure manager although, in Europe at least, competition from air travel has meant that relatively few rail services (such as Eurostar and Thalys high speed services) cross more than one border.

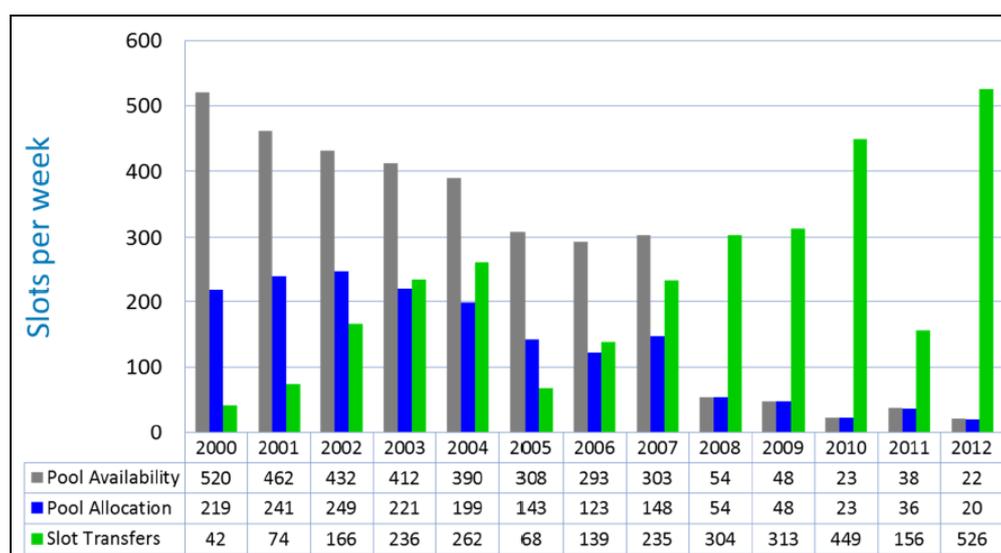
Final Report

The ownership of access

Airlines

- D2.22 The guiding rules to the allocation of airport slots in Europe under the “Slot Allocation Regulation” (SAR, Regulation 95/93/EEC) are:
- “Grandfather rights”: airlines which have used a slot in the past may continue to do so, in principle in perpetuity.
 - “Use it or lose it”: slots which have not regularly been used may be forfeit and returned to a “pool”, which also includes any newly-created slots, for allocation to other airlines.
- D2.23 In December 2011 the European Commission published its “Better Airports Package” which will, inter alia, expressly allow airlines to own, buy and sell slots. However the package has not yet come into force and the formal ownership of airport slots remains based on the combination of “grandfather rights” and “use it or lose it”.
- D2.24 At present, therefore, there is no official means by which slots can be either owned or formally bought and sold by airlines. However, at a limited number of airports, of which Heathrow is the best known example, airlines use the device of a “slot exchange”, which is permitted, typically exchanging a valuable slot for a worthless one late at night and exchanging money as part of a notionally separate transaction.
- D2.25 The figure below shows how the principal source of slots at Heathrow has been from **availability of existing slots**, through **allocation of new slots** (as additional capacity has been created by a variety of measures), to transfer or **trading of slots** between airlines, which now dominates as a mechanism.

HEATHROW AIRPORT: AVAILABILITY OF SLOTS



Source: Heathrow Airport Limited, September 2012

- D2.26 With a permitted total of 480,000 slots (or air transport movements, ATMs) at Heathrow, transfers of 500 slots per week in 2012 is equivalent to an annual turnover of around 5% of slots.

D2.27 Thus airlines may pay another airline to obtain (or lease) an airport slot as well as paying use-based airport charges to airports and air navigation charges to ANSPs.

Train operators

D2.28 Train operators’ rights to use the Network Rail infrastructure take the form of a track access contract, approved by ORR, setting out their rights. Track access contracts are for a fixed term which is ultimately limited by EU Directives, and payment to Network Rail is through regulated track access charges.

The need for trading

D2.29 As noted above, airlines only need to trade to obtain slots at a small number of congested airports such as London Heathrow.

The means of trading

D2.30 Airlines may use the slot exchange mechanism either to buy slots outright or, in some cases, to lease them. Leasing typically occurs where an airline wishes to offer services only for a single season.

D2.31 For an airline, the process of obtaining slots at an airport can be summarised as shown in the table below.

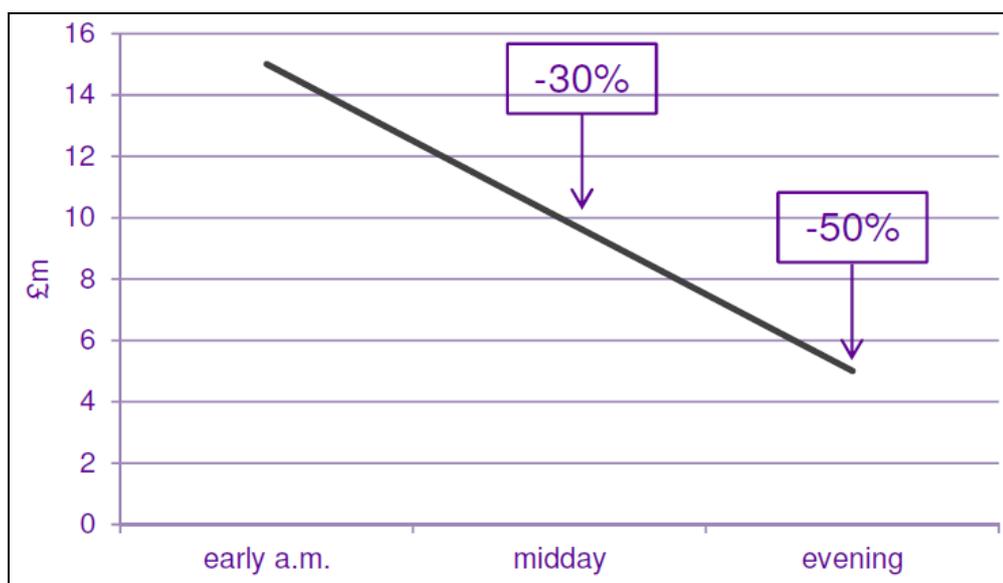
OBTAINING AN AIRPORT SLOT

Step	Action
1 Ask airport for slots at or near the desired time	If available, take the slots
2 Ask airlines with slots at or near the desired times for a swap or lease	If willing, negotiate price to mutual advantage: neither party will trade if worse off

The means of pricing

D2.32 The pricing of airline slots is primarily one of bilateral negotiation between the airlines concerned. However, Heathrow Airport Limited has published the indicative figure below, suggesting that the value of a daily slot can be as high as £15 million early in the morning, which has peaks of arriving and then departing long haul traffic, and fall to £5 million or less late in the evening.

HEATHROW AIRPORT: VALUE OF SLOTS



Source: Heathrow Airport Limited, September 2012

D2.33 For an airline wishing to value a slot, whether to sell or buy, the process is therefore as shown in the table below.

VALUING AN AIRPORT SLOT

Step	Conclusion
1 Are unused slots available at or close to the time of the slot?	Slot probably has no commercial value
2 Is intelligence available on the price at which slots at similar times have traded?	Slot value is likely to be similar to any recent trades in the same time period

D2.34 Valuation of slots may be required when actively trading, but is also relevant when reviewing the strategic value of maintaining a presence at an airport, or assessing an airline’s balance sheet for the purposes of a loan or transaction.

Improving allocation efficiency

D2.35 Evidence suggests that bilateral trades of individual slots can, over time, gradually improve the efficiency of allocation of a capacity. Airlines continue to trade and, in principle, the market value of each slot will converge on the higher of:

- Nothing, if close substitute slots are still available
- The value placed by the airline which values it the most

D2.36 Thus there will be an opportunity cost whenever an airline values a slot less than other airline would. In addition, in the case of airline slots, it is possible to estimate both:

- The typical benefit foregone by not using the slot at all, for example if the airline did not operate a flight, or the airport were to withdraw it
- The relative value of the slot to two potential users, enabling a series of bilateral trades to transfer the slot to the airline which valued it the most

D2.37 Thus the airport slot trading process is, in principle, capable of improving an efficient outcome merely by bilateral trades between airlines, but has a number of limitations:

- It will cease to be effective when slots are no longer the dominant constraint. At Heathrow, for example, it would only be possible to accept an excess of arrivals over departures for a limited period before the airport became “full”, with no remaining gate or stand space.
- It would be difficult, merely by bilateral trading, to create a large package of slots, such as if an airline required slots throughout the day to start a regular interval shuttle service. It could only achieve this by approaching the holders of a wide range of slots, at all times of day, and making contingent offers until it had assembled a combination of slots at hourly intervals. In practice, however, it is now rare for airlines to run regular interval services, especially at busy airports.

D2.38 Additionally, the slot trading process can work because of a number of particular features of airport slots:

- Runway capacity can be subdivided into fungible and independent units, the individual short periods in which aircraft use a runway to take off or land.
- Airlines can trade single slots bilaterally: there is rarely a need for a multilateral deal to assemble a large package of slots as a single transaction.
- Each airline serves a different market, and the value of a slot to it is largely independent of how other slots are used. This also means that airlines trading slots between themselves are unlikely to be direct competitors serving the same destination airport. This contrasts with the issue of contingency in the allocation of rail capacity, which we highlighted in Appendix B6.5. In practice, even if an airline which was asked to sell a slot to a direct competitor refused, or factored the cost to its own operations into its price, it could still be undercut by another airline, owning an adjacent slot, which would not demand this premium.

Improving investment efficiency

D2.39 Airports do not own slots and the value of the slots may not appear on their balance sheets. However, evidence that slots are being traded at high values can be interpreted as indicative of latent demand for runway capacity.

D2.40 However, and as noted in D2.17, the majority of European airports have built additional runways to ensure that there is sufficient capacity even for peak demand. The most conspicuous exceptions are Heathrow, the world’s busiest two-runway airport and Gatwick, the world’s busiest one-runway airport.

D3 EXAMPLES OF CAPACITY CONSTRAINTS ON THE RAIL NETWORK

D3.1 We noted in the main report (paragraph 2.30) how constraints on the rail network may be becoming increasingly systemic, and list a number of areas where there may be no single dominant constraint. We list some examples below, based on a review of the RUSes and the investment schemes proposed by Network Rail.

EXAMPLES OF AREAS WITH MULTIPLE CONSTRAINTS

Location	Issues
Central Scotland	Local, two long distance, sleeper and freight operators Multiple single track routes constraining timings
East Coast Main Line Edinburgh to Newcastle	Long distance and freight operators Local services around both Edinburgh and Newcastle
East Coast Main Line south of York	Local, three long distance and freight operators Limited platform space Conflicts at flat junctions Trains operating at different speeds
Midland Main Line south of Bedford	Local, long distance and freight operators Thameslink timings constrained by services south of London Flat junction with future Thameslink connection to ECML
Manchester area	Local, two regional and two long distance operators Limited platforms at Manchester Piccadilly and Salford Crescent Multiple flat junctions
West Coast Main Line Stafford to Rugby through Birmingham	Local and long distance and freight operators Extended two track sections with stopping services Flat junctions including crossing freight
West Coast Main Line south of Rugby	Local and long distance operators Trains operating at different speeds
Great Western Main Line Swansea to Newport	Local, long distance and freight services Flat junctions
Great Western Main Line Reading to London	Local, long distance, Heathrow Express and freight services Limited capacity at Reading before current redevelopment Future interaction with Crossrail and Great Eastern Main Line
South Western Main Line Southampton to London	Local, two long distance and freight services Limited platform numbers and length at Waterloo Flat junction at Woking
Brighton Main Line	Multiple operators including airport express services Thameslink timings constrained by services north of London Mixture of lines paired by direction and paired by purpose Conflicts at flat junctions particularly at East Croydon, but also at Redhill, Three Bridges, Haywards Heath Limited platforms at Victoria, London Bridge and Brighton
Southeastern	Complex network with multiple flat junctions

D3.2 We also examined in greater detail the East Coast Main Line south of York, which has a number of constraints, with varying degrees of severity, as listed in the table below. We compared the constraints:

- Identified in the 2006 ITS “Scoping study for scarcity charges”
- Identified in the 2008 Network Rail “East Coast Main Line RUS”
- Addressed in investment carried out to date

D3.3 The table shows which locations were described as a constraint in the 2006 ITS study or the 2008 RUS (⊗), or have since been addressed (☺), illustrating how changing demand can influence the priorities for work to enhance capacity.

“CONSTRAINTS” ON THE EAST COAST MAIN LINE (ECML)

	York	Doncaster	Newark crossing	Grantham	Peterborough	Hitchin flyover	Welwyn viaduct	Alexandra Palace	Finsbury Park	Kings Cross (P1)
ITS (2006)	⊗			⊗	⊗	⊗	⊗			
RUS (2008)					⊗		⊗	⊗	⊗	⊗
Addressed						☺		☺	☺	☺

D3.4 The constraints include York and Doncaster stations, a flat crossing at Newark, Grantham and Peterborough stations, a flat crossing (recently bypassed) at Hitchin, the two track viaduct at Welwyn, intense suburban services between Alexandra Palace and Finsbury Park, and the limited number of platforms at Kings Cross.

D3.5 The Welwyn viaduct has long been seen as a major constraint on the route, but the four kilometre long two track section includes not only the viaduct but also the station at Welwyn North and two tunnels, and it would cost several hundred million pounds to provide four through tracks. Timetabling to avoid conflicts there has meant that a series of other constraints have become as important and, in some cases, have subsequently been relieved.

D3.6 While the 2006 ITS Leeds report mentioned the flat junction at Hitchin, it did not refer to the need for increased capacity between Alexandra Palace and Finsbury Park, or an additional platform at Kings Cross, both of which have been delivered since the 2008 RUS.

D3.7 Taken together, the evidence suggests that the location and nature of constraints may evolve over time. This suggests that any system of scarcity charging may also need to be modified over time as the dominant constraint or constraints change as a result of changes both to the infrastructure and to the services operated on it.

D4 INFERRING OPPORTUNITY COST FROM THE RAIL TIMETABLE

- D4.1 The example of the rail operation described above also raises the issues of:
- Which elements of infrastructure are congested, and at what times
 - Which services should pay a scarcity charge for using congested infrastructure
- D4.2 One simplifying approach, which might be workable at some locations on the network, would be to nominate a single dominant constraint, analogous to an airport slot, attach a value or price to its use, and translate this into a charge on all trains passing through it at in periods when it was deemed to be congested.
- D4.3 Our examination of international experience identified the use of premium pricing to use short sections of track in Austria, between Mödling and Wien Meidling, and in Sweden, including the historic two-kilometre constraint between Stockholm Central and Stockholm Södra, which will soon be relieved by the Citybanan project.
- D4.4 On the East Coast Main Line, a decision would be required on the nomination of the dominant constraint. Some possible candidates, based on the example discussed above, are listed in the table below.

EAST COAST MAIN LINE: POSSIBLE DOMINANT CONSTRAINTS

Location	Argument for	Argument against
Kings Cross	Limited platforms for terminating trains and no immediate scope for expansion.	Thameslink will take up to 8 trains per hour out of the station from 2018.
Welwyn	Repeatedly shown to be extremely expensive to increase capacity, and this could not be delivered rapidly.	More trains could be diverted to use the Hertford Loop which, while slower, is an alternative.
Peterborough	Limited platforms and many flat junctions.	Current constraints should be relieved during 2014.

- D4.5 The table suggests that the nomination of a dominant constraint might be problematic but, given the recent work at Finsbury Park to Alexandra Palace and Hitchin, and planned work at Kings Cross and Peterborough, we assume for illustrative purposes that the dominant constraint, analogous to the airport runway, is taken to be at Welwyn, at least at the beginning of CP5. This suggests that any scarcity charge would be apply to users of the Welwyn viaduct, or strictly the section from Digswell to Woolmer Green Junction.
- D4.6 The next issue would be to determine whether the times at which any given train passes through Welwyn are congested at that point or not. Examination of the current Working Timetable for illustrative one-hour time periods shows the usage of the section shown in the table below.

TRAINS THROUGH WELWYN

Time period	08:00-08:59½	12:00-12:59½	13:00-13:59½	17:00-1759½
Southbound (“Up”)	17	13	13	12
Northbound (“Down”)	12	12	12	16

Note: based on trains at Woolmer Green Junction in Working Timetable

- D4.7 The pattern of service appears broadly plausible, with higher numbers towards London in the AM peak and from London in the PM peak. However, there remain at least two fundamental difficulties with this measure as a basis for estimating opportunity cost.
- D4.8 First, the periods used to count trains, based on 60 minute periods starting on the hour, are essentially arbitrary, and periods of different starting point or length would produce different relative counts. The same would be true if, instead of a count of trains, a CUI such as UIC 406 were used, and a CUI would introduce the further complication that it is measured not at a point, such as Woolmer Green Junction, but over a section of line between two end points, both of which would need to be defined.
- D4.9 Second, the fact that a lower number of trains operate in certain hours does not mean that there is spare capacity. Unlike an airport, it is not possible to infer from either the number of trains or a CUI in a period either:
- Whether it is possible to run any more trains through the constraint (although a CUI of 100% would normally mean that it is not).
 - Whether, if the CUI is less than 100%, it is possible to operate any particular train. This can only be determined by a timetabling exercise to find whether, by “flexing” the paths of other trains in the timetable, it is possible to create a complete train path.
- D4.10 In other words, the fact that there are fewer trains in a one hour period in the middle of the day provides no information on whether:
- Further trains could be added to the timetable without adjustment, or at all
 - Removing a train would allow any train, other than a direct replacement, to be added
- D4.11 The difference between airport slot and rail examples is summarised in the table overleaf. Note that, because airline slots are fungible, the cost of entry is the same as the value of exit, which would return the airport to the previous timetable. For rail infrastructure, however, it is possible that:
- The cost of entry had been high, requiring extensive timetable change, but there is little or no value in exit, because no other outstanding request for capacity could use the path created even if it became free. To recover the cost of entry it would be necessary to readjust other services, assuming nothing else had changed elsewhere on the network, to revert to the previous timetable.
 - The cost of entry had been high, but the changes made mean that spare capacity still exists and at least one more subsequent request can be met.

IDENTIFICATION OF OPPORTUNITY COSTS

Issue	Airport	Railway
Cost of entry	If slots are still free, zero.	Timetabling work when the path was created would identify the cost of changes were required to do so.
Value of exit	Otherwise, if there is unsatisfied demand, the highest offer by an unsatisfied applicant.	If there were no outstanding requests, it would be possible to recover any costs of entry by reverting to the previous timetable. If there were one or more outstanding requests, timetabling work would identify if they could now be met, and at what cost to other services.
Scope for further entry	If slots are still free, yes. Otherwise no.	If and when a further request was made, timetabling work would identify if it could be met, and at what cost to other services.

- D4.12 We note in particular that removing an existing train from the timetable would not necessarily release capacity of any use to outstanding applicants. Removing an entire round trip from a point north of Welwyn, for example, might only create an opportunity to operate an identical round trip or a (lower value) part of it.
- D4.13 The only circumstances in which timetabling work can be avoided is if an applicant can show that its service uses a subset of the path of an existing train, or return journey, but it is willing to offer more. It would, for example, almost certainly be possible to remove a round trip from Peterborough or further north and replace it with the much shorter return trip from Letchworth Garden City, reached by turning off the East Coast Main Line at the Hitchin flyover, or possibly even from Cambridge. However it seems prima facie unlikely that a shorter journey would be of higher value.
- D4.14 Rather than a direct or partial replacement, however, there might be value in re-optimising the timetable to take advantage of the additional capacity or to improve resilience and performance. With the removal of a round trip it might even be possible to carry out a major recast of the all-day timetable and achieve a major improvement in the overall service. However, any such exercises would almost certainly require multilateral coordination between all the users of the route, rather than small “capacity swaps”, and would depend on the willingness of other operators to retime their own services to take advantage of the opportunity.
- D4.15 More widely, in the absence of a workable round trip, the operator might no longer continue to use or lease the train and instead withdraw the entire diagram and all the services it provided, or even the package of services of which it formed a part. This highlights the point raised above (D1.4), that efficient use must be made not only of the infrastructure but also of the train or the fleet.
- D4.16 More widely still, the **feasibility** of a train path is contingent on what other paths exist and whether and how they can be modified. As noted in B6.4, when operators compete or collaborate with connections, the **commercial** value of a train path is contingent on what other paths exist. For example, duplicating but following a competitor’s path, or running all the services to a destination immediately after

each other, may be **feasible**, and consume relatively little additional **capacity**, but is unlikely to be **commercially valuable**.

D5 SUMMARY OF THE DIFFICULTY OF MEASURING CAPACITY

D5.1 In summary, techniques based on analysing the current timetable, such as counts of trains or CUI, may be correlated with the availability of capacity to operate additional services, but the correlation may be poor. Such a measure would be only a poor substitute for the detailed timetabling exercise needed to provide definitive information on whether:

- Adding an existing train had imposed any costs
- Removing an existing train would release any value
- Further trains can be added to the timetable

D5.2 This can only be achieved by detailed timetabling work, either:

- On a **market basis**, to identify whether and how **actual applicants for capacity** can be accommodated, in what combinations and with what compromises.
- On a **centrally-planned basis**, to identify whether and how **hypothetical new services or a redesigned timetable**, would maximise social value within the projected infrastructure configuration and capacity and demand.

CONTROL SHEET

Project/Proposal Name The practicalities of scarcity charging
Document Title Final Report
Client Contract/Project No. ORR/CT/13-52
SDG Project/Proposal No. P22673101

ISSUE HISTORY

Issue No.	Date	Details
1	31 March 2014	Final Report to ORR
2	3 April 2014	Revised to reflect comments by ORR

REVIEW

Originator Dick Dunmore
Other Contributors Lorenzo Casullo, Jim Collins, Dick Dapr , Simon Ellis, Tony
Walmsley, Philip Whitehead
Review by: Print Simon Ellis
Sign

DISTRIBUTION

Client: Office of Rail Regulation
Steer Davies Gleave: P22673101 team



P:\Projects\226\7.31\01\Work\06 Final Report\151204 Final Report no CONFIDENTIAL on control sheet.docx