A Mixed Method for Railway Capacity Allocation

Emanuel Broman*[1,2], Jonas Eliasson[2] and Martin Aronsson[3]

[1]Swedish National Road and Transport Research Institute VTI, Malvinas väg 6, SE-114 28 Stockholm, Sweden
[2]Linköping University, Luntgatan 2, SE-601 74 Norrköping, Sweden
[3]RISE SICS, Isafjordsgatan 22, SE-164 40 Kista, Sweden

Abstract

Railways in several countries are now operated through open access competition. In principle this allows for anyone to run train services at any time, although in practice a process is needed to coordinate and prioritise between requests, especially if capacity is scarce. Central to this process is usually a national infrastructure agency, which coordinates and prioritises according to some process framework, including explicit or implicit prioritisation criteria. However, such frameworks are often insufficient to differentiate between competitors in an open-access environment, where two or more operators may wish to offer similar services on the same line at the same time of day. An allocation method that improves social welfare would be desirable, but more exact calculations of the welfare contribution of individual departures is probably unfeasible in a deregulated environment, as they require commercially sensitive information such as number of passengers, fares and cost structures.

This paper proposes a method designed to allocate capacity more efficiently between commercial and public operators, as well as within segments. Market prices for capacity are introduced in a limited form. The mixed method uses auctions to allocate pre-defined paths to commercial operators on specified, capacity-constrained lines, while calculating the value of capacity for non-commercial public transport through social cost-benefit analyses (CBA) of timetables. To balance between commercial and non-commercial traffic, these CBA results are compared to the bids by commercial firms in the auction.

The result is a feasible method for more efficient capacity distribution between and within market segments, and also gives incentives for more efficient use of infrastructure capacity by operators.

Keywords: railways; capacity allocation; deregulation

1. Introduction

Open access competition has been introduced on several European railway markets over the past two decades; first for freight transport, then for passenger operators as well. It is an advanced form of deregulation that allows for an unlimited number of operators to enter the market and offer their services. Where open-access has been introduced, it has tended to lower prices, increase the number and diversity of operators, and increase service frequency.

As most railways (outside North America) are government-owned, governments control access to the infrastructure. This allocation of track capacity is a complex task that by necessity needs to follow some procedure in order not to become unfair or unmanageable. The outcome of the allocation procedure is a timetable, i.e. a document that specifies exactly which trains will run on the network, when they will run and how long the trips will take. All of these aspects – the composition of traffic, the timing of departures and the travel times – affect the railway system's welfare contribution. Ideally, the welfare contribution should be as high as possible (in Sweden this aim is even stated in law.)

The efficient allocation of capacity is made more complex by open access competition. First, it increases the number of conflicts, i.e. situations when two or more operators have applied for access to the same track at the same time. The number of conflicts correlates with the amount of traffic on the network, as more traffic makes capacity scarce. It also correlates with the diversity of traffic, as fast trains catch up with slower or stopping trains, making larger headways necessary. And it correlates with the number of active operators on the market, as operators plan their own traffic to be conflict-free but typically do not cooperate to avoid conflicts between

* Corresponding author. Tel.: +46 70-237 19 24.
E-mail address: emanuel.broman@abe.kth.se

themselves ahead of making their requests. As all three developments are likely results of open access competition, the number of conflicts can be expected to rise substantially.

There are other ways as well through which open access competition makes capacity allocation more complex. Because of heterogeneity in traffic, there is no longer any one type of allocation process that is most suitable for all traffic. Government run traffic can reasonably well be planned through administrative procedures, if these are suitably designed and there is access to all the information that is needed to properly assess the welfare contribution of alternative timetables. For commercial traffic, however, such procedures are unlikely to be efficient, largely because private operators will not share the necessary information. For this segment, it may be more suitable to allocate capacity through some market mechanism.

This paper aims to describe these challenges in sufficient detail, and then to propose a method that deals with the challenges and improves the railways' welfare contribution. Based on interviews and seminars with operators, regulators, public officials and politicians, we map out and explain the problems that need to be resolved and the conditions are for different types of solutions. We then discuss different methods for conflict resolution, with their respective strengths and weaknesses. Based on this, we then propose a way to combine different allocation procedures for different segments and types of situation, into a framework that feasibly and more efficiently allocates capacity.

Currently, infrastructure managers in many countries favour pure administrative procedures for capacity allocation, such as "first come, first served" or prioritisation criteria based on market segments (passenger trains have priority over freight trains, for instance) and simple metrics such as distance travelled. The advantage of this type of method is that they are simple enough for infrastructure managers to use, they are transparent and they produce predictable outcomes. Their drawback lies in the outcomes; "first come, first served" makes no effort to improve welfare above what one would achieve through allocating randomly, and prioritisation criteria risk producing corner solutions, where one market segment is favoured endlessly despite diminishing returns from adding capacity. Administrative processes do, however, mix well with collaborative approaches.

Collaborative approaches are another common set of allocation methods. In discussions between operators, they agree to voluntary adjustments of their requested timetables to resolve conflicts. Often enough, a small adjustment by one operator is sufficient to allow for a large improvement to another; and the good-will earned turns can be used to gain similar advantages in the next meeting, when the other operator needs to make a small adjustment.

This type of approach can be combined with administrative procedures by proclaiming that if operators cannot agree, the infrastructure manager will resolve the conflict unilaterally based on prioritisation criteria. This combination of methods is a feasible way to resolve conflicts between operators that belong to different market segments; they have incentives to agree, and if they do not the conflict will be resolved anyway through a transparent procedure. The method is more problematic when dealing with operators of the same type, especially if these are commercial firms in direct competition. In such a case, the operators may have no incentives to improve the situation for their competitor however small the required adjustment is, and there is no transparent procedural method to fall back on that can distinguish between same-type services that operate the same line.

A possible, but less used, method is the social cost-benefit analysis (CBA). The social CBA is much used elsewhere in the transport sector, especially to evaluate proposed infrastructure construction projects. By quantifying the potential welfare contribution of each capacity request, and the amount of capacity – however defined – that is needed, it is theoretically possible to produce a timetable that is optimal in terms of social welfare. But to do these calculations one must have a lot of information about the proposed traffic, including of cost structures, fares and volumes (of passengers or of goods). This could be made available for government-run services, but probably not for commercial traffic, as operators are wary of revealing business secrets.

To make the prioritisation between commercial operators more efficient, it would be useful put a price on capacity that varies with supply and demand, and thereby make operators 'reveal' what they are willing to pay. This could be accomplished either through a dynamic pricing mechanism for track capacity controlled by the infrastructure manager, or through an auction in which operators bid for departure slots.

The two types may have different uses. Dynamic pricing forces overall demand down to the level of supply. This is probably useful when freight traffic dominates, as departure times can usually be moved a few hours or more without problems, so conflicts can be avoided as long as demand does not exceed supply. Auctions are better at resolving conflicts, so may be more useful when dealing with commercial passenger operators, as they tend to request specific departure times that shall not be moved more than 10 minutes or so. The drawback with auctions is that they increase complexity on the operator side. If an operator does not win the slots that they had planned, they might be left with too many or too few train-sets, or staff-schedules and train-circulation schemes that do not work. These problems can be dealt with if the auction only covers a few hours, but become much bigger if all capacity on a line is allocated by this means; this is yet another reason to use this method for lines with passenger traffic, not freight traffic.

Market-pricing mechanisms are thus potentially useful for capacity allocation on lines with predominately commercial traffic, whether of passengers or freight, but they are probably less useful for non-commercial traffic. This is because the willingness-to-pay of a government-owned operator is not necessarily a good proxy for the "value" of the train service. Governments have aims that are less-easily quantifiable than those of commercial firms. It is less obvious how the welfare of the citizenry of a certain area, say, is affected by the adding or cancelation of individual departures, compared to how the profit of a commercial operator is affected. Governments would thus not get the clear feed-back on their bidding-strategy that commercial firms would. Also because fares on government-run services tend to be low and flat, they cannot be used as a proxy for passenger's value of making the trip. There is a declining marginal value of adding more departures on a train line, but for a government-run service it is hard to know at what point this value becomes lower than the marginal cost. For all of these reasons, market-pricing mechanisms for capacity allocation are more suitable for commercial traffic and less so for non-commercial traffic.

We propose in this paper to use a combination of capacity allocation procedures, each suitable for a specific market segment. The interfaces between procedures are designed to shift capacity in order for the total outcome to be more efficient in the social welfare sense. The procedures include an auction mechanism for commercial traffic on the most congested hours and parts of the network; a tool that calculates the respective welfare contribution of alternative timetables for commuter traffic; a short-term allocation procedure that uses dynamic pricing; and a procedure based on negotiation and priority criteria.

2. Traffic mix

An efficient capacity allocation method must be suitable to the traffic that operates the tracks. The traffic can be divided into three distinguished segments; commercial passenger traffic, tendered traffic and freight traffic. (This segmentation is used throughout the paper.) The segments differ in terms of the public's role in planning the traffic, whether they have commercial target functions or other aims, and if enough is known about the traffic to calculate its welfare contribution. For sure, there are examples of traffic that borderlines between these segments or that don't fit their segment's characteristics perfectly. But for the most part it is useful.

2.1. Commercial passenger traffic

The *commercial passenger traffic* segment covers long-distance lines between the major cities. It is operated by commercial firms on a deregulated, open access market. The largest operator on the market is the government-owned former monopolist (SJ), though both it and its competitors operate on strict commercial terms and can be assumed to be profit-maximising. These operators are allowed to set prices freely, and they use flexible pricing to achieve their objective to maximise profit. They are assumed to be less flexible in increasing or decreasing capacity, and to have a medium-long planning horizon. Because of competition, operators are thought to keep data on cost structures, revenue, fares and passenger numbers secret. As we shall see, this has consequences for what types of capacity allocation mechanisms will work for this paper.

2.2. Tendered traffic

Much of passenger train traffic is subsidised, for several possible reasons. One reason is the Mohring effect, which predicts increasing returns to scale of public transportation. Because increasing passenger numbers justifies a more frequent service, they cause waiting times (or scheduling times) for other passengers to drop. Travelling with public transportation hence has positive externalities, and the behaviour should therefore be incentivised through subsidies, in order to increase total welfare. Another possible reason is redistribution. Public transportation is more-often used by the less-well-off, so subsidising it is a progressive policy. A third reason for subsidies could be regional politics. Railway lines that need subsidising to stay profitable often run to or between smaller cities. Subsidising them is hence a way to transfer funds from large cities to more rural areas.

In Sweden, a common practice is for regional or national government to plan railway lines, and then offer it as a time-limited concession to the lowest-bidding commercial operator, while retaining detailed control at the political level. Commuter traffic, regional traffic and long-distance traffic outside of the main corridors are all operated this way. Forward planning for this category of traffic is long. Pricing is comparatively inflexible since fare-structures are typically set yearly for all departures on a political level, and they frequently form central parts of election campaigns on the regional level. Capacity is as inflexible as for long-distance traffic.

2.3. Freight traffic

Freight traffic is operated by commercial firms on a deregulated market. The segment is heterogeneous, but some characteristics ought to be mentioned: Much of the capacity for this segmented is allocated through the "ad-hoc" method rather than through the yearly time-tabling process. Freight traffic normally has much lower scheduling costs than passenger traffic; to be offered a time-slot one hour later than requested is normally seen as okay by operators. The planning-horizon is usually short. The socio-economic value of a specific service is even harder to assess than for commercial passenger traffic, as both the value and time-dependency of goods vary greatly, and as the second-best alternative (usually lorry) is a good alternative for some types of shipment but a very poor one for others.

3. Possible allocation methods

It normally lies upon the infrastructure agency to create a timetable for the railway network, i.e. to allocate track capacity to the active operators on the market, through some procedure. One important requirement of a timetable is that it must be feasible, i.e. no more than one train shall be planned to occupy the same piece of track at the same time. This implies that there must be a mechanism for conflict resolution; when two operators have timetable requests that cannot both be feasibly approved, one of them has to make amends.

Another requirement of capacity allocation procedures is that they should generate timetables of high social welfare. As discussed above, there may be many types of trains requesting to travel on the same tracks. Inevitably, some will have a higher welfare contribution than others, either because many more people travel on them, or because there is no good alternative way to solve the same transport-need, or some other reason. It is desirable that the used allocation method produces as high a welfare outcome as possible.

A number of different capacity allocation methods are either in use somewhere in the world or are proposed in the literature. In this section we discuss five different allocation methods. All five have influenced the system of combined allocation methods that is proposed in this paper.

3.1. Administrative procedure

The simplest way to resolve conflicts for capacity is through an administrative procedure that simply decides which operator needs to make amends based on some pre-defined criteria. An example of this is Belgium, where the infrastructure agency categorises traffic into inter-city trains, commuter trains, freight trains, etc. When two operators are in conflict, it is the one belonging to the least-prioritised category that needs to make amends. The benefits of this type of procedure are clear: it is simple, and the most-prioritised types of trains almost always get their requested timetables granted.

But there are drawbacks too. The procedure risks producing so called corner solutions, which means that all time-slots are given to the prioritised category of traffic and none to the less-prioritised types. But because of diminishing returns to scale, the seventeenth commuter train within the same hour, say, is worth less than the average of the first sixteen. If this 'last' commuter train replaces the most valuable freight train, this may be damaging to welfare even if commuter trains on average have a higher welfare contribution than freight trains. Corner solutions are thus most-often less than welfare-optimal.

Another problem concerns conflicts between operators within the same category of traffic; they cannot be resolved through this method. This may not be a problem in highly regulated markets, which typically have national monopolies for freight and long-distance passenger traffic, and local monopolies for commuter traffic. But on de-regulated markets, this method must be complemented with some other means of conflict-resolution.

3.2. Negotiations and priority criteria

The allocation method used in Sweden today uses a combination of negotiations and priority criteria. It is a planned process, controlled by the infrastructure agency (Trafikverket). A central philosophy is that operators meet and negotiate to solve conflicts; often-enough one part resolves the conflict by making a small adjustment in its timetable, just to earn good-will for the time when they need their counterpart to make adjustments. If negotiations fail, the infrastructure agency has the power to unilaterally solve the conflict based on priority criteria, which lay out which type of traffic is more valuable. The priority criteria are rarely used – only a few times per year – but it seems likely, as argued by Eliasson & Aronsson (2014), negotiation outcomes are heavily influenced by how the infrastructure agency would resolve the conflict, should it come to that.

The method's first part – negotiation – is usually good at solving minor conflicts. Solinen (2017) stress that creative solutions are sometimes worked out in this setting, for instance when regional traffic operators agree to cancel one train in favour of an express train, provided that their passengers are allowed to travel with the express service. Not all conflicts can be solved through negotiation, however. Especially for conflicts between same-type commercial operators, there is little incentive for either party to make amends that benefit its direct competitor. And small adjustments are not enough, such as when at least one requested departure must be denied to get a feasible timetable; such big conflicts are usually not resolved in the negotiation phase.

Another problem is lack of transparency; a new entrant on the commercial passenger market will struggle to understand how conflicts with its competitors are resolved, and hence what it can do to gain access to the most valuable time-slots. The practice may also be too inflexible. Normally in e.g. Sweden, each year's timetable bears a large resemblance to that of the year before, but if a large change was forced by the entry into the market of an operator that tries to take a dominant position, the number of resulting conflicts may be too big to handle.

The second part of the method – the priority criteria – have problems similar to those of administrative procedures. They risk producing corner solutions, which hurt welfare. They also fail to distinguish between requests within the same category of traffic, including of the type discussed above where two commercial operators request the same time-slot. This means that there are certain types of conflict that is unlikely to be resolved through negotiation and cannot be resolved through use of priority criteria. Such conflicts need to be resolved through other means, for instance a proper social cost-benefit analysis (social CBA).

3.3. Social cost-benefit analysis

It is possible to use social CBA to measure and compare the welfare contribution of timetables, at least in principle. Social CBA is an established framework within transport science (and other fields), predominantly used to evaluate proposed infrastructure investments, but also other policies. It sums the contributions to consumer surplus, producer surplus and net external effects of the evaluated policy. In the case of timetables, consumer surplus would for example increase if the increased frequency of service was raised as waiting times would be lowered, but producer surplus would instead decrease because of the additional cost of operating more trains (in addition to other effects). Social CBAs are especially well suited to make comparisons, such as between two possible timetables; it is less exact in gauging absolute values. The possibility to extend this framework to use for timetable comparison largely rests on the availability of data.

Because social CBAs are already used extensively within the transportation field, some of the required data is readily available in many countries. This includes estimates of travellers' value of time and value of delays, and of demand elasticity. But to evaluate timetables one would need additional data. For passenger traffic this includes: operators' fixed and variable costs; passenger numbers (per category – leisure or business); travel times, both planned and delays; and a measure of quality, e.g. fares. For freight traffic, data is needed on: operators' fixed and variable costs; prices paid by customers, a measure of punctuality and of its importance; prices paid by customers for their second-best option (often lorry); and a measure of punctuality for the second-best option.

The availability of data depends on the segment. For government operated or tendered traffic, the required data is either already available or could be demanded by the public transport agency. But for commercial passenger traffic, much of the required data is either corporate secrets of unknown to the operators themselves. This includes cost structures, passenger numbers and fares. It is even more difficult with freight traffic; it is probably not possible for the regulator to acquire useful data on the value of punctuality and prices for second-best options. This means that while it may be possible to allocate track capacity for government tendered traffic by use of social CBA, it cannot be used for commercial operators.

3.4. Dynamic track charges

A possible fix to the problem is, at least in theory, to introduce *dynamic track charges*, which are iteratively raised or lowered until demand equals supply. In each iteration, operators decide whether the track charges on offer allow for them to profitably operate the time-slot. For conflicts between competing commercial operators, the outcome of such a process can be expected to closely resemble the welfare-optimising outcome (ref Papper 1). It is, however, less suitable for tendered traffic. The decision-making bodies are not profit-maximising but welfare-maximising (or something worse), and that target-function may be more difficult to optimise, which makes their behaviour harder to predict. The iterative approach may also be problematic, as it presupposes that operators can repeatedly make decisions on complex and important issues, regarding its timetables and the profitability of different offerings.

Many important decisions regarding the traffic, including frequencies and fares, are decided not by the operator but by the public transport agency. These are not profit-maximising; instead, their objective function is possibly welfare maximisation, or some other political motive.

There are reasons to believe that the regional traffic authorities are less well suited to judge the value of different timetables, compared to commercial operators. First, they may have the wrong objective. Although arguably any government body should strive to maximise overall welfare, it is possible that they instead aim to maximise the benefits of some specific interest group. Secondly, welfare effects are not as immediately visible as revenue is. Commercial operators receive feedback in the form of changing revenues and costs when their service offering changes, but there is no analogous feedback for tax-funded operators.

Thirdly, regional authorities are not subject to the “survival of the fittest” mechanism of markets. This is a fundamental difference between firms operating in competitive markets and government entities; commercial firms that efficiently cater to their customers’ needs can increase market presence to the detriment of less efficient competitors, but the size of government entities is mostly decided by other factors, such as geographic borders. Fourthly, regional governments have small to non-existent incentives to take account of the cost that they impose on the national traffic. They have for instance been unwilling to make adjustments to enhance overall capacity utilisation, such as skipping small stations on every other tour, or moving some traffic from train to bus to save capacity for long-distance trains.

3.5. Auction of pre-defined train-paths

There would probably be fewer iterations in an *auction of pre-defined train-paths*. The auction is also a more efficient technique than dynamic pricing at simultaneously allocating several slots to two competitors, as prices are set directly by those who know more about the slots’ value – the operators. The complexity of decisions that operators need to make remains large, however, as staff-schedules, connections and train circulations need to be accounted for throughout the bidding process. The method would not solve the transparency issues of the *negotiations/priority criteria* method, as the train-paths would anyway need to be defined in collaboration between the infrastructure agency and operators.

4. A better way to allocate capacity

The proposed capacity allocation method builds on separate, but linked, mechanisms for commercial and tax-funded operators. The loss imposed on regional traffic operators by altering their timetables, compared to their requests, is calculated with social cost-benefit analysis (social CBA). Commercial operators are allocated train slots through an auction mechanism, and the balance between the two groups of operators is guaranteed by using the loss imposed on regional operators as reserve price in the auction. The reason for this separation is that too little data is available on commercial operations to make a proper CBA, while for public transport there are reasons to believe that an auction mechanism would not produce an efficient outcome (as described more thoroughly in section **Fel! Hittar inte referenskälla.**).

In summary, the method works as follows: First, regional operators send their applications. Secondly, the infrastructure agency plans corridors for long-distance commercial trains, making adjustments to the regional trains’ timetables as necessary. It calculates the loss imposed on regional operators and their passengers by the change. Thirdly, the long-distance commercial slots are auctioned to operators, with the calculated loss imposed on the regional traffic used as reservation price.

In the first phase, the infrastructure agency selects the parts of the network it deems suitable for the proposed method. In addition to setting the geographic boundaries, it is decided at which times of day to use the method. An important property of the selected tracks is that they are capacity constrained. This means that only parts of the network with high capacity utilisation shall be included, and that they shall only be included at times of day, and on those days of the week, that they are especially heavily utilised. This may for instance be during morning and evening rush-hour. On double-track stretches the time-limit may apply differently for the two tracks, for instance if the outbound track from a city is congested earlier in the morning than the inbound track is.

Operation on selected lines consists at least in part of commercial passenger traffic, as it does on the Swedish main lines from Stockholm to Göteborg and Malmö, respectively. At the heart of the method is an auction, where commercial operators bid for departure slots, which implies that it is not useful for those parts of the network that are exclusively used by tax-funded traffic and freight traffic. As for regional traffic, a method is used to calculate the welfare difference between alternative timetables.

4.1. The welfare of timetables

A vital component of the proposed allocation method is the comparison of different timetables for the tendered traffic. We will use a tool developed by Eliasson, Ait Ali and Warg (2018?) to calculate differences in social welfare between competing possible timetables; it is briefly explained in this section. The tool is programmed in MATLAB© and integrated with RailSys©.

The first step is to produce a so-called OD-matrix that shows how many passengers travel between any given pair of origin and destination, at any given time. Such a matrix is not always readily available however, as e.g. commuter operators typically track only boarding and not exits. But as commuters normally travel in one direction in the morning and in the opposite direction in the afternoon, it is possible to produce an approximation of the OD-matrix using only boarding data. The same OD-matrix is used for all compared timetables, implicitly assuming that changes to the timetable does not affect demand; this is probably a reasonable simplification when changes to the timetable are small.

Timetables are registered via RailSys. Based on these, travel patterns can be calculated. It is assumed that travellers choose to travel as fast as possible to their destination given their boarding time and the timetable, and that they do not unnecessarily switch trains. From the travel patterns, the tool calculates total waiting time, in-vehicle time, time spent on crowded trains, and the number of train switches; these are then used to calculate travellers' generalised cost. Because demand is assumed to be constant, it is sufficient to know the generalised cost to calculate the difference in consumer surplus between two timetables.

The tool also calculates producer surplus. From the timetables, it is possible to calculate the number of trainsets needed, total running time, and staff time. Together with information about the cost of owning, operating and maintaining trains, and of paying salaries to staff, this is sufficient to calculate producer costs. Again, the assumption of constant demand simplifies calculations, as ticket sales are also constant; the change in producer costs therefore directly gives the change in producer surplus. Welfare is then calculated as the sum of consumer and producer surplus.

Based on the OD matrix, the network model in RailSys and the inserted timetable, it is possible to calculate waiting times, travel times and a crowding factor for each traveller. Together with public data on the value of travel time, this is used to calculate the consumer surplus. Public data is also used to estimate operating costs. The sum of changes in consumer surplus and operating costs when replacing one timetable with another is a good-enough estimation of the difference in social welfare between those timetables, for our purposes.

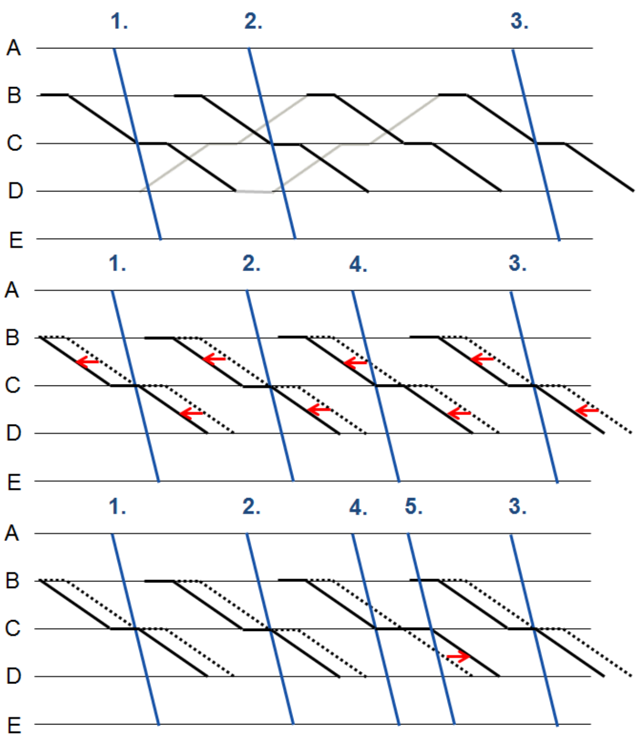

Figure 1. Adjusting commuter train timetable to make room for inter-city train-paths

--

Beskriv hur man skulle kunna testa modellen
Plocka in en riktig tidtabell någon gång när det är tajt, till exempel något av pendlarstråken runt Stockholm i morgongrusningen. T.ex. sträckan Stockholm C/Stockholm City-Uppsala. Lägg in fler ”X2-lägen”. (Jennie eller nån kan lägga in, det måste inte göras med hela Trafikverkets organisation och världens hallabaloo för att visa konceptet.) Sedan räknar man ut skillnaden i welfare för de olika tidtabellerna. Kan bl.a. bli en snygg graf som visar hur reservationspriset stiger med antalet kommersiella tåglägen (exponentiellt?).

4.2. Preparations for the auction

Regional traffic departments are the first to send their timetable requests. It is important that these initial requests are in their view the most ideal traffic and not, as they are sometimes believed to be today, an adaptation to what they believe will be granted.

The infrastructure agency then decides an interval for the number of slots that will be sold to commercial operators. This interval shall be wide enough so that it can be said with certainty that the number of slots sold will fall within it. As with the definition of lines, this will likely be easier with experience. A first version of the regional timetable is made that exactly resembles the wishes of the regional public transport agencies.

The first version of the timetable for regional traffic is then successively adjusted to make space for slots in the commercial timetable (see Figure 1). As the number of commercial slots increases, more intrusive changes need to be made to the regional timetable. When this process is complete, there is one version of the regional timetable and one corresponding version of the commercial timetable for each number of commercial slots within the defined interval. Slots that are defined in one version of the commercial timetable also appear in other versions, so that a version with eleven slots exactly resembles the one with ten slots, except for the one additional departure.

When this process is complete, the social welfare of each version of the regional traffic departments' timetables is assessed, using the tool described in the previous section. This method uses an origin/destination matrix for the number of passengers within the network, and then calculates the value of waiting time and in-vehicle time for all passengers. From this we receive a prediction of the socio-economic loss of each version of the timetable that is useful to our purposes, which is to compare the different timetable-versions to one another. For the timetables that correspond to few commercial slots this loss will be small, but as the number of commercial slots increases and the changes to the regional timetables become more severe, the socio-economic loss becomes greater. The change in socio-economic loss between two consecutive versions of the timetable is then used as a reserve price in the auction.

The method presented in this paper makes use of a model developed by Ait Ali et.al. for calculating the welfare contribution of timetables for public traffic. It is possible with this model to compare the welfare contribution of two alternative timetables. If one of the timetables allows for a commercial train to be scheduled on the same track, the difference between the two timetables can be used as a measure of the commercial train's externality due to crowding.

The model includes an estimation of the number of travellers between any origin-destination (OD) pair in the network at any time of day. This OD matrix is constructed from data on the number of boarding passengers at all times and stations. Destinations are derived from boardings on the return-trip; since passengers on public transport typically used pre-paid electronic tickets, such as monthly travelcards, the same ticket is registered for several trips. The timetabling software RailSys is integrated with the model in order to provide a usable user interface for inserting timetables. RailSys also includes a so called microscopic (i.e. detailed) model of the network, with travel times for all OD pairs.

4.3. Auctioning the slots

The infrastructure manager brings all possible versions of timetables to the auction; one for each number of commercial train slots that might be sold. For each pair of timetables, the welfare difference for tendered traffic is calculated; this is used as reservation price in the auction. Timetable adjustments are also prepared for operators that use rolling stock with slightly different characteristics, including small differences in cruising speed and acceleration and breaking performance.

The auction is organised by the infrastructure agency, and participating are representatives of all commercial operators that plan to have operations on the defined lines during the forthcoming timetable year. There are several auction rounds with up to a few days in between; at the date of the first auction round representatives of all bidding (commercial) operators show up at the premises of the infrastructure manager.

All slots for all defined lines on the network are auctioned simultaneously. This means that all operators place one bid for each auctioned time-slot on every part of the network that is included in the auction process. The reason is for operators to have train circulations work, to get the right quantity of time-slots (corresponding to how many train-sets they have), and to be able to make a cohesive offer to travellers, such as equally many departures in each direction of a line.

In order to win a time-slot, an operator must place a bid that is both higher than those of its competitors and higher than the reservation price. A reservation price is calculated for each time-slot, and it is equal to the loss imposed on the regional traffic by making room for that commercial timeslot (as described in 1.2). One of the outcomes of the auction is hence to decide which version of the timetable will be used, which affects both commercial and non-commercial operators. The purpose of making the reservation price equal to the welfare loss imposed on regional traffic is to ensure an efficient capacity allocation between the segments of tendered and commercial traffic.

The auction takes the form of a combinatorial clock auction, as designed by Ausubel, Cramton & Milgrom (2006). This is a version of the Vickrey auction that has been extensively used to auction radio spectrum (Mochon & Saez, 2017) (Ausubel & Baranov, 2017), and that is widely regarded to be more practical in real-world applications than a conventional Vickrey auction (Levin & Skrzypacz, 2016). Yet it retains the typical features of a Vickrey auction: that winners typically pay a lower price than their winning bid, and (hence) that the auctioneer's revenue is lower than profit-maximising.

In short, a combinatorial clock auction works as follows. It begins with a clock-phase, in which the auctioneer (infrastructure manager) sets and step-wise increases the price. In each step, the bidders (operators) respond with their demand, i.e. how many departure slots they would buy if this was the price. This bidding-phase is open, and largely exists in order for the bidders to learn about what a likely equilibrium will look like in the final round. In the final round, bidders place sealed combinatorial bids for the combination of departure-slots that they want to purchase. Based on the information that this provides, the infrastructure manager allocates departure-slots so as to maximise the value for operators, using their bids of expressions for their true valuations. The prices paid are then calculated as the opportunity cost of taking up specific departure slots (typically lower than the winning bid). A more thorough description of the combinatorial clock auction process can be found in (Levin & Skrzypacz, 2016).

4.4. The rest of the network

After the last round of the auction is completed, there exists a definite timetable for both commercial and tendered traffic on the parts of the network and the times of day that are handled through that process. It is now time to lay the timetable for all other times of the day and all other parts of the railway network. This is done according to the *negotiations and priority criteria* method (as described in section 3.2), which is a good method for small conflicts, and for combining requests from different types of traffic. In this process, it is ensured that trainset circulations work, taking into account the needs given by the auction process.

After the timetable for the residual network has been set, a process begins to make small adjustments to the auctioned slots. These changes include adjusting the times at intermediate stops slightly to give adequate time for boarding depending on if there is a connecting train that brings many passengers, and similar small changes. These changes are only allowed if they can be made without major consequences for other traffic. The changes are made after negotiation with the infrastructure agency, and subject to a fee.

4.5. A system of methods

We propose in this paper to use a system of linked allocation methods on the Swedish railway network. The different methods serve different purposes or different segments of the market. Described in section 3.2 is the yearly process which mostly serves operators of passenger traffic. It is itself a combination of methods, with an auction mechanism, a planned method, and some track capacity that is allocated based on calculations of socio-economic efficiency. Freight operators are mostly not served by the yearly process but by the ad-hoc method, which allocates capacity on a first-come-first-served basis. On top of this, capacity is reserved for planned track maintenance and for buffer and robustness reasons.

Table 1. Proxies for socio-economic value for different capacity allocation methods.

Allocation method	Proxy for socio-economic value
Yearly allocation	Price per departure-slot payed in auction; average over all auctioned slots (specified per track-segment and time-of-day)
Ad-hoc allocation	Price per departure-slot; average over all slots sold through dynamic-pricing measure (specified per track-segment and time-of-day)

Maintenance windows	Welfare-gain of scheduled maintenance windows compared to second-best alternative, as fraction of number of replaced departure slots in mixed traffic; average over all scheduled maintenance windows (specified per track-segment and time-of-day)

The different methods are used at different times; the yearly process takes place earlier than the ad-hoc process, for instance. Because of this, each method is apportioned a certain amount of track-capacity to allocate. As the overarching aim of the allocation system is to achieve socio-economic efficiency, it is important that the apportioning of capacity to the different methods is efficient.

The apportioning of track-capacity to the different methods is done yearly. The infrastructure manager decides on the matter using different proxies for socio-economic value in combination with experience and rules-of-thumb. We list in Table 1Table 1. Proxies for socio-economic value for different capacity allocation methods. possible proxies for socio-economic value in the different methods.

5. Practical implementation

As the allocation method proposed in this paper has not been tested, there are many questions to consider. Some concern the produced timetable – will it be feasible, and improve on welfare compared to other methods? Other concerns are practical, and regard how the operators and infrastructure manager will manage their respective tasks in the proposed procedures.

Start with the practical. An important question is if it is possible to create train paths are that allow operators with different types of rolling stock to make use of speed, acceleration and braking abilities. According to Solinen (2017) this is possible for the two fast train operators on the Stockholm-Göteborg line, SJ and MTR Express. Svensson and Solinen (2017) mention that there is a postal (PostNord) train that departs from Stockholm in rush hour, and that this train is fast enough to use a corridor designed for the passenger trains. As PostNord is a commercial firm, this may be possible

However, the sequence may be important if both operators use slots that are close in time, as SJ's trains are faster than those of MTR Express, its competitor. However, Westin (2017) argues that there is a large psychological barrier within the agency against using the overloaded declaration before operators have been consulted, and that this is a serious hindrance to the feasibility of the solution.

Regarding the process, Solinen (2017) claims that it would not be a problem for the regional traffic departments to send their capacity requests earlier than today, as is needed in the proposed process, since they know far ahead how they want to operate their traffic and they do not make large changes from one year to the next. However, both Svensson and Unebrand (2017) argue that the auction process should not be allowed to take too much calendar time, certainly not as much as a month. This is because the date when operators send in their requests is determined through international deals – it is currently the middle of April – and the process must therefore not be much longer than it is today.

Westin and Nordlöf (2017) ask if it is possible to 'game' the proposed solution. Since there are naturally few operators – typically only two – that compete on a single line, they may choose to decide in between who gets which slots, and thereby pay too low prices.

Regarding the legal worries, Unebrand and Svensson (2017) mention that capacity charges, including through auctions, can only be collected after a section is declared to be overloaded. But making such a declaration early in the process is not necessarily a bad idea, according to her

Interviews have been conducted to understand whether the proposed capacity allocation method is feasible.

5.1. Auction outcomes

A central part of the system of allocation methods proposed in this paper is the auctioning of time-slots for long-distance commercial passenger trains. As such an auction has not been implemented anywhere, we do not know for sure what its outcome will be; one of fierce competition and rising consumer surplus, or of one operator buying all time-slots in order to restore monopoly and raise fares. It depends on a number of influencing factors, whose characteristics are debatable.

It is possible that the outcome of the auction process is a restored monopoly, for the following reason: As fares are higher in monopoly (Beria, Redondi, & Malighetti, 2014) (Vigren, 2016), a would-be monopolist can afford to out-bid a smaller rival. The profit of a monopolist is higher than the combined profits of two competing operators, if their cost per departure is identical and non-increasing (Broman & Eliasson, 2017). This suggests that it is profitable for a would-be monopolist to outbid its competitor in an auction for track-capacity even when the competitor is prepared to operate at zero profit. Indeed, Nordlöf (2017) argues that the ‘natural’ outcome of the auction may be that one operator buys all slots, as monopoly is more profitable than competition.

Also possible, however, is an outcome with many small, single-departure operators and profits close to zero. Driving this outcome is the fact that ‘existing’ departures loose passengers when supply increases. This means that even if an additional departure is profitable in itself, an operator that is already present on the market might still decide that increasing supply is unprofitable when looking at all its departures. A new entrant, to the contrary, could profitably enter the market. Small operators will therefore continue to enter the market as long as single departures are profitable. Larger operators will, at the same time, dispose of departures in a quest to reduce overall supply and thereby push the general price-level upwards.

Which of these outcomes (or which combination of the two) is more likely depends on a number of factors. One factor is the cost per departure. Operators’ capacity can plausibly be described as fixed in the short to medium term. Lead-times for newly produced rolling stock are long, making it all-but impossible for operators to increase capacity with new train-sets in the short term. At the same time, markets for used rolling stock are rather illiquid because of differing standards between countries and too few potential buyers and sellers within countries. This leads to the assumption that capacity is (more or less) fixed in the short term. If capacity is fixed, then the cost per departure is when operating below the capacity-limit, because ownership of the rolling stock is a sunk cost that does not change by marginally changing the number of departures. But above the capacity-limit, the cost per departure may be high, progressively increasing or even infinite (depending on more detailed assumptions on how the market for rolling stock works). The assumption that the cost per departure becomes very high beyond some point leads to the conclusion that a monopoly outcome is not so likely, because it reduces the profitability of the ‘last’ departures for the would-be monopolist and hence its incentive to out-bid its competitor.

Another question is if operators prefer their departures to be grouped together or spread-out. We argue in (“Welfare effects”) that operators with grouped departures enjoy pseudo-monopoly situations that allow them to charge higher fares. There is hence additional revenue to be made by grouping departures together. On the other hand, there may be practical reasons for operators to prefer more spread-out departures. They may wish to offer travellers options throughout the day; there may be connections to other lines and other modes that are spread-out, and that are important sources of demand; and it may be more cost-efficient to make trainsets departure-ready with some time between, by simplifying staff-scheduling. Such practical concerns imply that grouping departures is implied with additional costs to the operator. Assumptions on the revenues and costs associated with grouping departures are likely to affect the prevalence of pseudo-monopolies, and hence market structure and welfare effects.

The overall demand/price elasticity on the market is likewise important and at the same time not obvious. One might argue that total demand is rather static, as the need to travel between any two cities is influenced by factors outside the control of actors on the transportation market. But there may be other modes than train that offer an alternative, such as aeroplane, coach or private car; the demand for train-travel is then more likely to vary with the fare level. If overall demand is price-sensitive, then the incentive to form a monopoly on the train market is reduced, as there is anyway less scope to increase fares.

There are thus many factors that affect the outcome of a track-capacity auction as the one proposed in this paper, and differing assumptions on those factors leads to different conclusions regarding outcomes. Some of these factors – including the cost per departure, if operators prefer departures to be grouped together or spread out, and the overall demand/price elasticity of the market – are far from obvious how to make, yet are likely to have big implications for results.

6. Conclusions

This paper proposes a method for allocating capacity on the Swedish railway network that combines market prices through auction with a planned process, and a method for estimating the socio-economic value of subsidies regional traffic. It attempts to solve the problems in today’s method, including its inflexibility, its opaqueness and its lack of method for solving conflicts between similar competitors. Also, it provides a measure to allocate capacity between commercial and subsidised traffic efficiently.

Interviews have been conducted with staff at the infrastructure agency in order to assess the feasibility of the solution. This has provided insights into the details that need to be addressed before the proposed method can be implemented, as well as some questions of more fundamental nature that need to be addressed.

In all, it seems however that no serious obstacles have been found that should not be possible to overcome. The proposed method thus seems to offer a feasible path to introduce market prices as a means to solve timetabling conflicts on congested stretches of the Swedish railway network. Today’s method could thereby be replaced with one that is more efficient and more suited to a deregulated environment.

References

Aronsson, M. (2017, January 25). PhD, SICS Swedish ICT. (E. Broman, & J. Eliasson, Interviewers)

Ausubel, L., & Baranov, O. (2017). A Practigal Guide to the Combinatorial Clock Auction. *The Economic Journal*, *volume 127*, *issue 605*, F334-F350.

Ausubel, L., Cramton, P., & Milgrom, P. (2006). *The clock-proxy auction: A practical combinatorial auction design.*

Beria, P., Redondi, R., & Malighetti, P. (2014). The effect of open access rail competition on average prices. The case of Milan-Ancona. *XVI Riunione Scientifica SIET, October 8-10*, 1-23.

Broman, E., & Eliasson, J. (2017). Market dynamics in on-rail competition. *Transportation Research Procedia 22*, (pp. 232-244).

Dahlberg, H. (2017, January 25). Market & Planning, Business Development, Section for Service Management. (E. Broman, & J. Eliasson, Interviewers)

Eliasson, J., & Aronsson, M. (2014). Samhällsekonomiskt effektiv tilldelning av järnvägskapacitet. *CTS working paper*.

Levin, J., & Skrzypacz, A. (2016). Properties of the Combinatorial Clock Auction. *American Economic Journal 106(9)*, 2528-2551.

Mochon, A., & Saez, Y. (2017). A review of radio spectrum combinatorial clock auctions. *Telecommunications Policy, volume 41*, 303-324.

Nordlöf, P. (2017, February 2). manager Expert Centre, Social CBA. (E. Broman, Interviewer)

Svensson, A. (2017, February 23). Strategic Development. (E. Broman, Interviewer)

Unebrand, J., & Solinen, E. (2017, February 23). planning, Expert Centre, Capacity Central. (E. Broman, Interviewer)

Westin, P.-E. (2017, February 6). Strategic planning, infrastructure charges. (E. Broman, Interviewer)

Vigren, A. (2016). *Competition in Swedish Passenger Railway; Entry in an open-access market and its effect on prices.* CTS working paper 2016:18.