An Economic Evaluation of the Swedish Prioritisation Rule for Conflict Resolution in Train Traffic Management

Emma Andersson

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:
http://dx.doi.org/10.1016/j.sbspro.2014.01.097
Copyright: Not Found
Publisher URL Missing

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-104675
EWGT2013 – 16th Meeting of the EURO Working Group on Transportation

An economic evaluation of the Swedish prioritisation rule for conflict resolution in train traffic management

Emma Andersson*

Department of Science and Technology, Linköping University, SE-601 74, Sweden

Abstract

An increase in train traffic is a politically welcomed trend, which on the other hand has led to too high capacity utilisation at times and a railway network sensitive to disturbances. Delays are easily spread, causing high cost. A mean of controlling the secondary delays is to use efficient operational prioritisation rules for trains in conflict. This paper presents an evaluation of the current Swedish prioritisation rule. For two frequent conflict situations the associated cost related to applying the rule is calculated. The result indicates a poor economic efficiency and show that significant savings can be achieved by changing strategy.

Keywords: Railway traffic, delay management, operational planning, prioritising, delay costs

1. Introduction

An increase in train traffic is in Sweden a politically welcomed trend. This has on the other hand led to too high capacity utilisation at times and a railway network sensitive to disturbances. For some areas, at periods of the day, the overutilisation often results in delays. In Sweden, trains with very heterogeneous traffic are operating on the same tracks. Disturbances often occur due to various reasons, which can be analysed and prevented separately, but in this paper we focus on secondary (i.e. consecutive, or knock-on) delays. Secondary delays are the induced effect to other traffic when a delayed train disturbs other trains. Because of the many trains and the traffic heterogeneity, delayed trains will easily disturb other traffic and the delays are easily spread (Andersson, Peterson and Törnquist Krasemann, 2011).

Train delays result in high time costs for passengers, freight owners, operators and infrastructure providers, and the induced socio-economic costs are substantial. If the secondary delays could be reduced, an extensive part of the costs could be reduced and welfare increase. A mean of reducing the secondary delays can be found in the operational train dispatching where dispatchers give priority to a certain train when there is a conflict between two trains. We have a European perspective where a master timetable is used and trains are scheduled in seconds. The current main operational prioritisation rule in Sweden dictates that the dispatchers should prioritise trains on time (i.e. trains that depart and run according to their timetable) if they are in conflict with delayed trains not running according to their timetable (Trafikverket, 2013).

This paper presents an evaluation of the current prioritisation rule. In two real world examples the paper illustrates some concrete conflict situations and the associated delay cost related to applying the rule. The aim is to show that cost reductions can be made with another strategy. The intention of this paper is to increase the awareness

* Corresponding author. Tel.: +4611363108.
E-mail address: emma.andersson@liu.se
of the prioritisation problem and it could be used as an incitement for further more comprehensive and detailed studies of how to construct new rules.

The paper is organised as follows. Section 2 introduces the problem with operational prioritisation of trains in conflict and present the Swedish rule. Section 3 describes two conflict situation examples. In section 4 the main parameters used when calculating the cost for travel time and delays are described. Also the Swedish parameter values are presented together with the formula and other parameter values needed for the delay cost calculation. The cost is then calculated for the two examples. In Section 5 conclusions and future research are discussed.

2. Operational prioritisation of trains in conflict

In previous research several ideas for railway traffic disturbance management and prioritisation of trains in conflict have been studied. Jespersen-Groth et al. (2009) have studied three main sub problems in railway disruption management; timetable adjustment, rolling stock and crew re-scheduling. In all these areas delays affect the original plan and a new one has to be constructed in real-time. For small disturbances, small modifications of the timetable are enough, but for larger disruptions overtakings, changes in stopping pattern and/or cancellation of trains could be necessary. However, according to the authors, it is hard to estimate how severe the disruptions are, i.e. if a disturbance should be classified as a small or a large disturbance.

Kliewer and Suhl (2011) have evaluated various railway dispatching strategies. Instead of focusing on the timeliness of the trains they focus on the passengers’ waiting time due to the different dispatching strategies. For example, if the dispatching strategy is to never let a connecting train wait for another train that is delayed, the passenger may get long waiting times which result in high costs.

Törnquist (2007) has used a heuristic approach to minimise e.g. the total delay, delay costs and travel times in case of disturbances. Depending on which objective is used, the dispatchers’ decisions in conflict situations differ. The choice of strategy also depends on how long in the future the operational planning horizon reaches. One conclusion of the study is that the conflict resolution strategy should be able to change and not result in the same type of trains being punished or prioritised over and over again.

Corman et al. (2012) has also used heuristic algorithms to solve a bi-objective conflict resolution goal to minimise the train delays and the passenger dissatisfaction. These two objectives are sometimes in conflict since the minimisation of train delays could result in cancellation of trains, among other things, which is a large dissatisfaction for the passengers.

From previous research we can learn that the best prioritisation rule depends on which perspective is being used and it can differ depending on situations. The current rule in Sweden is instead strict and trains should be handled in the same way in all conflict situation, as will be describes further on.

2.1. The Swedish prioritisation rule and its implementation

In Sweden the main measure of train performance is punctuality; the number of trains arriving to their final destination within 3/5/15 minutes of their planned arrival time (3 minutes for airport shuttle trains, 5 minutes for regional trains and 15 minutes for long-distance trains). Also the en-route performance is measured and a train is defined late if it is more than 3 minutes late according to its timetable. Then the cause of the disturbance must be reported to Trafikverket. When delays of more than 5 minutes occur the operator that causes the delay is then being charged by Trafikverket. Trafikverket also has to pay the involved operators if they cause the delay (Trafikverket, 2013). This gives a large incentive for both Trafikverket and the operators to decrease the amount and magnitude of train delays.
The Swedish current main dispatching rule for conflict resolution in train traffic dictates that the dispatcher should prioritise trains on time if they are in conflict with delayed trains (Trafikverket, 2013). The intention of this rule is to prevent delays from spreading. The dispatcher could deviate from the rule if there is a written request from an operator to prioritise between the operator’s own trains. A written request can also be made in an agreement between operators that some important trains from one operator should have higher priority than trains from another operator. This request must very clearly specify which trains that should be given higher priority and which trains that should be given a lower priority and it must be sent in together with the initial path request. In a conflict involving several delayed trains it is up to the dispatcher to make the best overall decisions.

In practice the train dispatchers deviate from the rule more frequently when they see a better overall solution by giving priority to a delayed train, as can be seen in the following examples. However, their decisions are then much based on previous experience and, to be strict, they are actually deviating from the rule. This leads to variations in the outcome depending on which dispatcher is making the decisions and how the dispatcher gives priority in this particular situation.

All train slots in Swedish railway timetables contain some amount of margin time (i.e. buffer time or time supplements) which the trains can use when they get disturbed. The intention with margins is to construct a more robust timetable in which trains can recover from small delays and keep the delays from spreading to other trains. However, to get the desired effect of the margins, it is essential that they are properly used by the train dispatchers. Andersson, Peterson and Törnquist Krasemann (2013) have analysed railway timetable robustness and found conflict situations where there is a special need for flexible margin use. The current prioritisation rule does not always allow for the needed flexibility and it is hard for the train dispatchers to overlook the impacts of the different decisions. A new way to prioritise between trains in conflict is a necessary condition when increasing timetable robustness according to the authors.

In Fig. 1 we can see a typical conflict situation at station B where train 2 is scheduled to enter the line close after train 1, which is delayed.

The graphs in the figure illustrate two distinct dispatching possibilities; either train 1 runs after train 2 and becomes more delayed (left graph), or train 1 runs ahead of train 2, which means that train 2 will get a small delay (right graph). In this example train 2 can recover from the delay before reaching its final destination (station E) thanks to its margins.

According to the current prioritisation rule the first solution is the correct one and we can see in the figure that this decision has a large impact for the punctuality for train 1.
3. Real world examples of a conflict situation

In this section a typical conflict situation is described by two examples. The examples are from the Swedish Southern mainline, a double-track line between Stockholm and Malmö. This is one of the most congested railway lines in Sweden where long-distance trains share the tracks with regional, commuter and freight trains. Delayed trains end up behind other slower trains and can not recover from their initial delay which makes the traffic on this line very sensitive to disturbances. The two examples illustrate a conflict situation, commonly appearing in the Swedish railway traffic. At this line, every long-distance train is exposed to 3-5 similar conflict situations per journey, when being only a few minutes delayed. Since there are 13 long-distance trains operating on the line per direction and day, the number of possible conflicts is rapidly increasing when considering all lines and train types.

For the two trains involved in each conflict situation, the en-route punctuality was collected during two month in the autumn 2011. Also the effect from other traffic was registered to see how other train influence the two conflicting trains depending on the dispatcher's decisions. However, we have excluded the two conflicting trains' possible impact at other traffic as a consequence of the dispatcher's decision.

3.1. Conflict situation example one

The first example comprises southbound long-distance train 537 and commuter train 1231 in Hässleholm (HM). Train 537 is going from Stockholm (CST) to Malmö (M), and in HM train 1231 is planned to depart only two minutes after train 537 has passed the station. This means that if train 537 has just a small delay, it will result in a conflict in HM. From the punctuality statistics it is clear that the train dispatchers often deviate from the current prioritisation rule and let train 537 be prioritised before train 1231. This happens in situations when train 537 is up to five minutes delayed. When train 537 is more than five minutes delayed, the punctuality statistics show that the dispatchers' decisions vary from day to day, see Fig. 2. In the figure we can identify three distinct performance clusters occurring after HM. Depending on which cluster a train end up in, the final arrival delay in M differ. Trains in cluster 1 overtake train 1231 in LU, trains in cluster 2 overtake train 1231 in E and trains in cluster 3 overtake 1231 already in TÖ or SG. The resulting average delay when arriving to M for each cluster is 19 minutes for cluster
1, 16 minutes for cluster 2 and 12 minutes for cluster 3. In all situations train 1231 arrive at M with an average delay of two minutes.

There are several trains that are in TÖ at the same time but then they are divided into different clusters depending on the dispatcher’s decision.

![Fig. 2. Punctuality statistics reflecting the train dispatchers decisions in the critical point in MY for train 542. The x-axis gives the space defined by stations and the y-axis gives the delay in minutes. The black lines represent three trains, equally delayed in MY, but given different priority. Trains with a delay less than 5 minutes are excluded.](image)

3.2. Conflict situation example two

In the second example the train dispatcher’s decision and the resulting delay is even more distinct. The example concerns northbound long-distance train 542 and commuter train 8774 in Mjölby (MY), where train 8774 is planned to depart four minutes after train 542 has passed the station. This means that if train 542 has just a small delay, it will result in a conflict at MY. In Fig. 3 the punctuality statistics for train 542 is shown. For three trains with the equal delay of 7 minutes in MY, the arrival delay at the end station (CST) varies relatively much depending on the dispatcher’s decision. In cluster 1, train 542 runs after train 8774 all the way to NR, which is the end station for the commuter train. This result in a 17 minutes delay for train 542 in NR, and for trains 542 that are around 17 minutes late in NR, the average arrival delay in CST is 21 minutes. In cluster 2, train 542 overtake train 8774 in LP but still gets 10 minutes delayed in NR. For trains 542 that are 10 minutes delayed in NR, the average arrival delay in CST is 11 minutes. In cluster 3, train 542 overtake train 8774 already in MT and remains 7 minutes delayed in NR. For trains 542 that are 7 minutes delayed in NR, the average arrival delay in CST is 8 minutes. However, the train in cluster 3 has the possibility to recover from the entire delay in time for CST, in contrast to trains in the other clusters, if all other trains are on time and nothing else unexpected happens. The trains in cluster 1 and 2 are delayed to such extent that they continue to CST in the wrong train slot and can not use their margins for recovering.

In cluster 1 and 2 the train dispatcher’s decision results in no delay for train 8774, but in cluster 3 train 8774 will leave MY 3 minutes late. In LP this delays is reduced to one minute and soon after LP the commuter train has fully recovered from it.
4. Economic delay calculations for the examples

Most people see traveling as an inevitable activity. Travelling in itself results in no utility, it is related to costs such as ticket purchase, fuel consumption, etc. Since a trip consume time that could be devoted to other more preferable activities it also has a time cost, which else could have been used for other, more valuable activities. To quantify this in monetary terms, the value of time (VOT) is used.

A delayed trip has negative effects for the passengers, which also are associated with costs. A parameter frequently used for estimating the value of the uncertainty in travel time is the value of reliability (VOR). VOR is a parameter that describes how much money a person would pay to reduce the travel time uncertainty.

When calculating the delay cost, VOR is the main parameter. VOR is based on VOT and therefore we start the calculations by presenting a short overview of the two parameters together with the Swedish values used today.

4.1. The value of time

Already in the middle of the 20th century researchers started to analyse VOT (e.g. Becker, 1965; De Serpa, 1971). Some of these theories were followed up by empirical studies which can be seen in a review by Hensher (1976). In 1980, the British Department of Transportation decided to review the state of the art and find quantitative measures of VOT, which lead to the first national VOT study (MVA Consultancy, 1987). This study is now the source to several other national VOT studies, including the Swedish ones.

When estimating VOT a fundamental assumption is that individuals want to maximise their utility. The most influencing factors for VOT are travel mode, travel distance, income and journey purpose (Mackie et al., 2003; Small and Verhoef, 2007). The impact of these factors has been confirmed by many studies and the research of estimating VOT has been developed through the years.

When estimating VOT for business travels there is one special aspect to consider, the travel time has also a value for the employer. One approach to estimate VOT for business travels is to set VOT to the marginal cost of the
employer (Fowkes, Marks and Nash, 1986). The authors also argue about some factors that indicate that this is not an accurate estimation of VOT, e.g. the business trip might occur on private time and the employee might work during the travel.

4.2. The value of reliability

When there is an uncertainty in travel time the traveller has to add a “safety margin” to the travel time which has a cost assigned to it (Gaver, 1968). There is also a discomfort related to the uncertainty itself, i.e. the risk of being late. Jackson and Jucker (1982) introduced the concept of utility as a function of expected (mean) travel time and variability (standard deviation) in travel time and where travellers seek the best trade-off between them.

Bates et al. (2001) bring together a large amount of theoretical and empirical results to find a general theory for travellers’ VOR. They study public transport especially where the timetables often are fixed. This means that the possibilities for departure are discrete which could lead to further disutility when travellers want to add a “safety margin” to the travel time. If the travel time becomes unreliable the generalised cost of using the service increases and the traveller might change to other transport modes. For scheduled travels with long intervals between departures and relatively low travel time variability, such as long-distance trains, the most common approach is to use average delay as a variability measure (Börjesson and Eliasson, 2012).

4.3. Values used by Trafikverket

The first national Swedish VOT study was performed by Algers, Dillén and Widlert (1995). One of the aims of this study was to co-ordinate resources from different transport sectors and estimate general VOT:s. Another aim was to provide more insight in VOT for business trips, since previous single Swedish studies had no satisfying general values. Algers, Dillén and Widlert (1995) used the Hensher-approach as a base for the business travel VOT and the following components to define VOT for business travels:

- The marginal productivity of labour
- The share of saved travel time used for leisure
- The share of saved travel time used productively
- The relative productivity of saved travel time that was used for work
- The value to the employee of saved travel time despite work or leisure (VOT for private traveling)

The values for public transport by rail was classified according to trip purpose (commuting, business or private), trip length (regional or national) and train type (regional, Inter-City or fast train). The values from the study fit well to other national VOT studies in Western Europe at that time (Balcombe et al., 2004). Since 1995, these values have been revised several times to represent the present year.

In 2007/2008 a new VOT study was performed, leading to new national VOT:s. As before, the estimation was based on the travellers’ wish to maximise their utility. However, a new method for VOT estimation was chosen (Börjesson and Eliasson, 2012). The present VOT:s used for Swedish train cost-benefit analyses are shown in Table 1.

VOT in the 2007/2008 study is not classified with respect to train type (Trafikverket, 2012) but it is estimated with respect to the length of the calculation period. If the calculation period is 40 years or more, the long-run values should be used, else the short-run values should be used. The long-time values are scaled up short-run values due to expected income increase over the years, which will also result in increased VOT.

There was also a new assumption made regarding VOT for railway business travels, namely that the share of saved travel time used for leisure should be zero (Trafikverket, 2012). This means that all travel time will accrue the employer. The relative productivity was set to 1, which means that the productivity is the same when travelling
as working in the office. However, the share of saved travel time used productively, i.e. the time actually used for work, was set to only 15 %.

<table>
<thead>
<tr>
<th>Private</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commuting</td>
<td>Regional</td>
</tr>
<tr>
<td>Business</td>
<td>Regional</td>
</tr>
<tr>
<td>Short-run</td>
<td>69</td>
</tr>
<tr>
<td>Long-run</td>
<td>92</td>
</tr>
</tbody>
</table>

The VOR estimation for public transport was based on average delay. This is an established way to estimate travel time uncertainties and just as theoretic well-motivated as variance or standard-deviation methods (Börjesson and Eliasson, 2011). The chosen VOR is 3.5 times VOT, which is in line with estimations by other studies (Trafikverket, 2012).

4.4. Delay cost calculation formula and parameter values

When calculating the delay cost for the different dispatcher decisions we only consider the costs for the two trains involved in the conflict. The delay cost for each train is calculated individually to be compared in the next step. To calculate the delay costs for train $t$ we use the following formula:

$$Cost^t = d^t * p^t * \sum_{\alpha=1}^{3} VOR^\alpha * s^\alpha t,$$

where $\alpha$ specifies the passenger type (commuter, private or business), $d^t$ is the delay for train $t$ in hours, $p^t$ is the number of passengers on-board train $t$, $s^\alpha t$ is the share of passenger type $\alpha$ at train $t$ in percent and $VOR^\alpha$ is the value of reliability for passenger type $\alpha$ in SEK/hour ($VOR^\alpha = 3.5 * VOT^\alpha$).

For long-distance trains an average number of passengers on-board between every station for the studied time period was provided by the operator, along with the share of passenger types, see Table 2. For the commuter trains, no precise information could be given. Therefore the occupancy on-board and share of passengers are based on assumptions. In a minor sensitivity analysis we have made small adjustment of the assumed figures and found that they have a small impact on the result.

<table>
<thead>
<tr>
<th>Train</th>
<th>Business</th>
<th>Private</th>
<th>Commuter</th>
</tr>
</thead>
<tbody>
<tr>
<td>537</td>
<td>35 %</td>
<td>55 %</td>
<td>10 %</td>
</tr>
<tr>
<td>542</td>
<td>34 %</td>
<td>57 %</td>
<td>9 %</td>
</tr>
<tr>
<td>1231/8774</td>
<td>-</td>
<td>30 %</td>
<td>70 %</td>
</tr>
</tbody>
</table>

Both examples involve commuter trains that begin their journey around 6 p.m., which is in the afternoon traffic peak. In both commuter train systems they use the same type of train model, i.e. X61 with 210 seats, and the assumption made is that 90 per cent of the seats are occupied. This means that there are an average of approximately 190 passengers on-board commuter train 1231 and 8774.
For the calculation we also need the share of each passenger type. In the examples the commuter trains operate in the afternoon peak we therefore assume that the share of commuters is high, 70%. The rest are travelling privately. The share of passengers for all involved trains is shown in Table 2.

4.5. Result of the delay cost calculation

The costs for the delays are calculated for the two examples and the result for the long-distance trains is shown in Table 3. We present the delay costs using the short-run VOT from Trafikverket (2012). The long-run values result in the same result but with figures scaled up.

| Table 3. Delay costs per long-distance train for the typical delays in the two examples |
|---------------------------------|-----------------|-----------------|
|                                 | Delay (min)     | Cost (SEK)      |
|                                 |                 | (Short-run)     |
| Example 1                       |                 |                 |
| Cluster 1                       | 19              | 27 803          |
| Cluster 2                       | 16              | 23 413          |
| Cluster 3                       | 12              | 17 560          |
| Example 2                       |                 |                 |
| Cluster 1                       | 21              | 36 182          |
| Cluster 2                       | 11              | 19 517          |
| Cluster 3                       | 8               | 14 194          |

In the table we can clearly see a difference in costs depending on how the train dispatcher has prioritised. For those trains, in example 1, that have the same delay in TÖ there is a possibility to save 4 400 – 5 900 SEK per train. This can be achieved by giving the long-distance train a higher priority and let it end up in a cluster with smaller secondary delays.

For trains in example 2 the dispatcher’s influence on the delay cost has a larger impact. If the dispatcher would have let the train from cluster 1 run like the train in cluster 2, 16 700 SEK would be saved and if the dispatcher would have let it run like the train in cluster 3, 22 000 SEK would be saved. However, according to the current prioritisation rule the train dispatcher made the right decision for the train in cluster 1. The train was delayed and should not be prioritised before a train on time.

To determine whether this is the best overall solution or not, the cost resulting from the commuter train delays must also be calculated. In example 1 the commuter train was two minutes delayed when arriving to M regardless of the train dispatchers’ prioritisation. This results in a delay cost of 1 400 SEK, which is much less than the cost for delaying the long-distance train. In fact, the commuter train could be delayed even more to let the long-distance train have priority. Since a train, according to Trafikverket, is delayed when it is more than 3 minutes behind its timetable, the small delay brought on the commuter train is not even considered as a delay.

In example 2 the commuter train gets delayed in cluster 3. The three minutes departure delay in MY is soon absorbed and the cost for the small delay is 700 SEK. Compared to the savings of 22 000 SEK that can be made for the long-distance train, delaying the commuter train one minute must be seen as an acceptable loss. In this case the commuter train delay must be reported as a delay to Trafikverket, since it is three minutes, but no further reprimands or compulsions are needed.

Reliability is an important factor that also has an effect on the travellers choice of transportation mode. In the above examples the delays could be reduced if the dispatchers deviate from the current prioritisation rule. In such case also the reliability will increase. This would probably lead to a higher share of travellers choosing railway as transportation mode. The benefit for these new travellers is not included in the calculation, which means that the result is to some extent underestimated.
5. Conclusions and future research

In this paper the current prioritisation rule for trains in conflict has been studied. The current rule is easily communicated and implemented and implies that an on-time train should always be prioritised before a delayed train. However, for a frequently occurring conflict situation, we show that this rule leads to a poor economic efficiency that brings large costs that could have been avoided with another strategy. We also show that it is possible to reduce costs by delaying the on-time train and by such means prevent the already delayed train to get even more delayed. Often the train dispatchers make good decisions that result in overall reduced delays, but there is a need for decision support that encourages them. The main contribution of this paper is to show that cost reductions can be made with another prioritisation strategy.

How to establish new prioritisation rules is a challenge for future research. Intuitively, a more natural rule for prioritising between trains in conflict should be to let the train with the largest cost be prioritised. This requires a more extensive analysis of all cost involved in a certain decision.

When delaying an on-time train it could result in more negative effects such as important transfers could be missed or rolling-stock circulations could break. In real traffic, there is more traffic and a delayed train can end up after several other trains, far from its own timetable slot, and with no possibility to recover. Therefore it is important to make the prioritisation from a broader perspective than to just give priority to the on-time train. To get the best overall solution it is necessary to include more trains in a comprehensive analysis since a decision at one point in the network could result in more or less successful chain reactions further on. This paper serves as an incitement to proceed with further more thorough studies.

When dealing with larger networks there is a need for other methods, such as optimisation and simulation. The induced economic costs, as presented in this paper, could be a suitable objective for such an optimisation procedure.

For an acceptance among the operators, however, the strategy should be combined with some type of economic compensation controlling equity between the operators and Trafikverket. In such way we would achieve a higher reliability, which would also lead to increased railway travelling.

Acknowledgements

This research was conducted within the research project “Robust Timetables for Railway Traffic”, which is financially supported by grants from VINNOVA (The Swedish Governmental Agency for Innovation Systems), Trafikverket and SJ AB. The author is grateful for all data provided by Trafikverket and SJ AB. Also special thanks for support and feedback provided by Johan Holmgren, Anders Peterson and Johanna Törnquist Krasemann at the Department of Science and Technology, Linköping University.

References


