Coordination of planning processes for traffic operators on rail networks

Annual Report 1
Swedish national railway administration
(Banverket) R&D-project — SPOK

Martin Aronsson, Jan Ekman, Per Kreuger

30th June 2003

Swedish: SAMORDNING AV PLANERINGSFLÖDEN OCH KRAVSPECIFIKATIONER FÖR TRAFIKUTÖVARE PÅ JÄRNVÄGSNÄT (SPOK)

Sammanfattning

This project is concerned with how to improve the capacity allocation process. In particular the project aims at proposing an enhanced format for train path applications, study the technical limitations for timetabling support tools and in the longer term to implement support systems for the train path allocation process. This report describes the various factors that affect the application process, and report the opinions from several actors in the field.

Since the deregulation of the Swedish railway, and with new EU directives, the foundations for the capacity allocation process is changing rapidly. There is a strong need for clear and predictable principles that are fair and operator neutral and implements the prioritisation given to different types of traffic and for improved methodologies and decision support tools in the capacity allocation process. This is crucial to support both the possibility for the traffic operators to state demands and the timetable designers in judging conflicting train paths. A conclusion is that almost all the requirements on the timetable presented in this report can in fact be stated as properties and relations of events in the timetable.

Nyckelord Train path application, timetable, railway, schedule, capacity allocation
## Innehåll

1 Preface 5
2 Introduction 5
  2.1 Main question and project aims 5
  2.2 Principles for allocating train paths (the PTAT-project) 5
3 The capacity allocation process in brief 5
  3.1 The parties involved in the capacity allocation process 6
  3.2 The importance of how capacity is allocated 6
  3.3 The complexity of allocating capacity 7
4 The overall transportation goals 8
  4.1 Introduction 8
  4.2 The overall goal and overall good 8
  4.3 Assessment of good for planned traffic and real 10
  4.4 Three tasks for improving the capacity allocation process 10
  4.5 The measure of good 10
  4.6 The vagueness of the overall goal 11
5 Potential for improvement 12
  5.1 Introduction 12
  5.2 Global goals versus local decisions in the process 13
  5.3 The predominance of the traffic operator SJ 14
  5.4 The information needed in timetabling 14
  5.5 The time needed in the process of allocating capacity 14
  5.6 The timetabling scheduling problem 15
  5.7 The network statement, slots and train paths 15
6 Issues of concern for the parties involved

6.1 Issues of concern for traffic operators

6.1.1 Some general remarks

6.1.2 Passenger traffic

6.1.3 Freight traffic

6.2 Issues of concern for the allocation body

6.2.1 Train path application format

6.2.2 Transportation tasks

6.2.3 Rules of thumb

6.2.4 Interpretation of the train path application and dependencies between trains

6.2.5 Heuristics in timetabling process

6.2.6 The TrainPlan planning tool

7 Requirements

7.1 Identified requirements

7.1.1 Customer requirements (transportation task)

7.1.2 The operators requirements

7.1.3 Connections and exchanges

7.1.4 Weight, length, and cargo requirements

7.1.5 Infrastructure constraints

7.2 Mathematical formalisation of requirement primitives

7.2.1 Elementary concepts

7.2.2 Basic mathematical properties and relations

7.2.3 Weights on properties and relations

7.3 Mapping the requirements to mathematical primitives

7.3.1 Notation

7.3.2 Trains

7.3.3 Relations

8 Outline of proposed approach for improvement of allocation process

9 Conclusions, results and future work
1 Preface

This document reports some results in the project “Coordination of planning processes for traffic operators on rail networks: part 1.” The project is carried out by SICS and the first phase of the project reported here spans a period from July 2002 to June 2003. The project is part of Banverket’s FoU-programme (Banverket’s programme for research and development). Contact at Banverket is Thomas Franzén. Participants at SICS are Martin Aronsson (project manager), Jan Ekman and Per Kreuger. A separate result of the project is the State of the art report, which is written in Swedish.

2 Introduction

2.1 Main question and project aims

This project is concerned with how to improve the capacity allocation process. In particular the project aims at suggesting an enhanced format for train path applications, study the technical limitations for timetabling support tools and in the longer term to implement support systems for the train path allocation process.

2.2 Principles for allocating train paths (the PTAT-project)

The project “Principles for allocating train paths”\(^1\) is carried out by Banverket and runs in parallel with this project. The aim of the PTAT project is to propose principles for prioritisation when allocating train paths to competing traffic operators. The PTAT project group is formed from Banverket and representatives from some of the operators. A project goal is to gain approval for and, in the year 2004, include these principles as a part of the network statement\(^2\).

This project partially overlaps with the PTAT-project. However this project, i.e. SPOK, has a more far-reaching goal and is more oriented toward clarification of the demands of operators and end users, what may satisfy these demands and what requirements are manageable in the timetabling process. PTAT is centred around gaining support for the principles for capacity allocation put forward and the principles for prioritisation finally decided on will be put in operation in the near future.

3 The capacity allocation process in brief

The project deals with the question of how to improve the capacity allocation process, i.e. the process that results in the timetable. What is meant by improving the capacity

\(^1\)In Swedish: Principer för tilldelning av tåglägen (PTAT)

\(^2\)In Swedish: Järnvägsnätsbeskrivning
allocation process and how this project aims to contribute to such an improvement is made clear in the sequel. We will start with a brief introduction to the capacity allocation process, illustrated in figure 1.

3.1 The parties involved in the capacity allocation process

Tågtrafikledningen is responsible for the capacity allocation

Today the (railway) infrastructure is owned and maintained by Banverket, i.e. The Swedish national rail administration, and the government body Tågtrafikledningen (Dispatching authority) is responsible for allocating railway capacity to the traffic operators. This is the implication of the deregulation of the Swedish railways. Among the traffic operators SJ and Green Cargo still dominate the passenger and freight services, respectively. In the capacity allocation process the concerned parties interact in order to establish a timetable. As a part of this interaction the traffic operators apply for train paths from Tågtrafikledningen, which determine the timetable. The process is repeated twice a year. Although Tågtrafikledningen is responsible for establishing the timetable, it is prepared by timetable designers at Banverket Trafik, a subdivision of Banverket.

3.2 The importance of how capacity is allocated

The capacity allocation process is important for trade and industry in Sweden.

---

Redrawn from [Fra03].
The capacity allocation process is important for trade and industry in Sweden. This concerns both passenger and freight services, although the needs for different transports differ a lot. An efficient utilisation of railway infrastructure is an essential component in a high quality transportation system as a whole. A well-functioning transportation system is necessary and critical for a region to be attractive and available for trade and industry and for people to work and live in. This is especially true for countries like Sweden and Finland in which industry greatly depends on transports.

Capacity allocation is concerned with large amounts of money. For instance, Green Cargo estimates that if the time that their trains have to wait for other trains (about 40000 hours over a year) could be halved, then they would earn 80 million SEK, just in production costs. The possibility to carry out certain transportation tasks are also significant to the profit of customers of freight services.

For any (profit making) company, it is very important to estimate their revenue and costs in order to make their business profitable. For traffic operators their main cost is the cost for running the trains, in principle that is the cost of the rolling stock and the personnel. Their revenue depends on how well they satisfy the demands of passengers and freight customers. Hence, the allocation of capacity affects both the costs and revenue for the traffic operators. Therefore, it is of great importance for the traffic operators to have a clear understanding of the principles and practice governing capacity allocation.

3.3 The complexity of allocating capacity

Timetabling is at the heart of the capacity allocation process

The capacity allocation process is complex in that it concerns issues ranging from the objectives for the national transportation policies to rules of thumb for how to deal with train paths that overlap in applications for capacity. The process also comprises a great deal of activities, from the traffic operators application preparations to timetabling. Important to the capacity allocation process is also how the involved parties interact and cooperate.

Figure 2 attempts to illustrate the most important concerns in the capacity allocation process.

Timetabling is at the heart of the process, all the other concerns affects the timetabling process; different kinds of regulations, needs of and demands from traffic operators and the infrastructure management, etc.

In this project we pay particular attention to two of the above concerns: train path applications and technical limitations. The project study what information should be submitted in the application from two perspectives: what are the needs of the applicants, and what can be efficiently handled by the allocation body. The project also aims at studying timetabling support tools. When aiming at a support tool for timetabling, none of the concerns in the picture above can be ignored.
For the moment we will consider one of the concerns, i.e. the overall transportation goals. After that we shall look at the capacity allocation process from a different angle, namely the need of change in the process at present.

4 The overall transportation goals

4.1 Introduction

The aim of this section is to discuss the possibilities for the capacity allocation process to meet the overall goal described by regulation. We mainly consider the capacity allocation process as described by the instructions of Tågtrafikledningen, the authority responsible for allocating railway capacity.

4.2 The overall goal and overall good

The overall principle for allocating capacity is that the infrastructure is used efficiently. Let us quote beginning of the regulation (förordning) (1996:734) 22 §:

“Efficient utilisation of the national railway infrastructure shall be taken into consideration when allocating track capacity.”

According to Banverket this is the only clear instruction in the current regulation concerning capacity allocation. Let us cite [Ban02]:

4Quotation in Swedish: “Vid banfördelningen skall beaktas att statens spårutbyggnader utnyttjas effektivt.”
"The present principle for allocating train paths is is that it shall result in an efficient utilisation of the national railway infrastructure in terms of public common good. In addition to this goal there are no clear instructions in the current regulation on how to allocate capacity and on what this allocation shall be based."\(^5\)

In some sense it is the objective for both this project and the PTAT-project to (partially) fill out the large gap between the abstract and ambitious overall principle of efficient utilisation and the practical procedures of capacity allocation.

That infrastructure is used (or utilised) efficiently is the responsibility of Tågtrafikledningen that determines the timetable. The way to assure that infrastructure is utilised efficiently is to take into account the overall good, as expressed in the instructions for how to allocate capacity during autumn 2002 and spring 2003.

“The planned railway traffic shall according to the regulation result in efficient utilisation of the national railway infrastructure. This goal is a guiding principle when planning the railway traffic, at which the authority has to balance a number of, sometimes conflicting, applications and view the overall good of the traffic as a whole.”\(^6\)

That is, the timetable designers are instructed to base their decisions on assessments of the overall good for the traffic as a whole. This principle also is abstract and ambitious, just as the overall principle of efficient utilisation. It is difficult to follow the principle in practise, especially since the instructions of Tågtrafikledningen for how to allocate capacity do not supply any guidance on how to apply the principle. Therefore the timetable designers are more or less left on their own, with their experience and skill and within a limited amount of time available, to form an opinion on which solution best utilises the infrastructure. In the following we, however, presuppose that the aim of Tågtrafikledningen and the government is that the instruction shall be followed.

As far as we know, timetable construction at present differ somewhat from the principle in the instructions of Tågtrafikledningen. In fact, by the nature of the principle, it would be remarkable if this was not the case. For example the decisions made at timetabling are not based on assessments of the overall good for the traffic as a whole. Let us therefore distinguish between the capacity allocation process as described by the instruction and the present practise and let us use allocation by instruction and allocation in practise, respectively, to refer to each of these two in the following.

\(^5\)Quotation in Swedish: “Den nuvarande principen för tilldelning av tåglägen är att den ska leda till ett samhälls ekonomiskt effektivt utnyttjande av spåranläggningarna”, (förordning (1996:734) om statens spåranläggningar 22 § 1)). Utöver detta mål finns det inga tydliga direktiv i gällande lagstiftning om hur tilldelning av tåglägen skall gå till och vad den skall baseras på.”

\(^6\)Quotation in Swedish: “Tågplanen skall enligt förordningen resultera i ett effektivt utnyttjande av statens spåranläggningar. Detta mål är vägledande vid tågplanekonstruktionen varvid myndigheten har att sammanväga ett antal, ibland konkurrerande, ansökningar och bedöma den samlade nytta av den totala trafikbildan [BL02].”
4.3 Assessment of good for planned traffic and real

We have observed that, by the instructions of the responsible authority, time-tableing is based on assessments of the overall good for the traffic as a whole. In this context we must not fail to observe also that the infrastructure utilisation does not concern the timetable but the real traffic resulting from putting the timetable in operation. That is, in order for the timetable designers to follow the instructions they need not only to consider the good of the planned traffic but also to take into account the predicted deviations from the timetable. Hence, the timetable designers need to be concerned with the loss of infrastructure utilisation (or good) resulting from of the predicted deviations from the timetable. Notice that this loss may not at all correspond to the total minutes of delay resulting from putting the timetable in operation, but on some valuation of what the deviations resulted in.

4.4 Three tasks for improving the capacity allocation process

The task of improving the capacity allocation process concerns is dived in three as follows:

- Decide on precise and assessable goal for the capacity allocation
- Make sure that the right information exists at timetabling
- Find useful principles and build tools to support the creation of solutions that meets the goals

If the goals with the capacity allocation are determined, then it is clear which information is unnecessary for the allocation. For instance, if it is decided that traffic operator’s costs for personnel is taken under consideration, then some information on working schemes might be relevant for capacity allocation and can be part of an application.

For allocation by instruction the assessments of the overall good for the traffic corresponds to the first task and leads to the question: Is the measure of good appropriate?

For allocation in practise rules of thumb are used to resolve conflicting train paths in the applications for capacity. This corresponds to the third of the three tasks above. This lead to the question: Do the rules of thumb used in timetabling result in sufficiently efficient utilisation of the railway infrastructure? For some rules of thumb or support tools used at timetabling in allocation by instruction we are instead led the question: How well do the used rules and tools fit the measure of good?

4.5 The measure of good

As mentioned previously, in allocation by instruction the timetable designers are instructed to base their decisions on estimations of the total good for the traffic as a
whole, as a means to achieve an efficient utilisation of the infrastructure. The instructions [BL02] for how to allocate capacity express some general concerns in timetabling, but do not really give any hint about the estimation of good. As an example let us cite these instructions:

“When determining the timetable the main concern is the purpose of the slot, i.e. transportation task. It is presupposed that each traffic operator has a fixed plan for their own traffic. This one is probably based on optimal resource usage for their own transportation tasks.”

Efficient utilisation of infrastructure is commonly interpreted as a utilisation with a high common good, as in [Ban02]:

“The present principle for allocating train paths is that it shall result in an efficient utilisation of the national railway infrastructure in terms of public common good.

[Ban02]. There exists formulae for computing the common good, see for instance [SW01], but the general opinion among the parties involved in the capacity allocation process is that the use of these formulae for allocating capacity do not, in many cases, result in an efficient utilisation of the infrastructure. The formulae are mainly tailored for infrastructure investments, and not for judgement in the timetabling process.

4.6 The vagueness of the overall goal

In documents from Banverket the opinion expressed is that it is not clear how to interpret the law. Let us once more cite [Ban02].

“What the goal means in practise is not clear, neither which factors shall be given any value when determining the public common good.”

---

7 Quotation in Swedish: Vid tidtabellkonstruktionen utgår Banverket från ändamålet med tägläget, d.v.s. transportuppgiften. Utgångspunkten är att det vid täglägesansökningsstillfället hos varje trafikutövare finns en uppgjord plan för den egna trafiken. Denna bygger troligen på en optimal resursåtgång för det egna trafikupplägget.

8 Quotation in Swedish: “Den nuvarande principen för tilldelning av täglägen är att den ska leda till ett samhälls ekonomiskt effektivt utnyttjande av spåranläggningarna”, (förordning (1996:734) om statens spåranläggningar 22 § 1)).

9 Quotation in Swedish: “Utöver detta mål finns det inga tydliga direktiv i gällande lagstiftning om hur tilldelning av täglägen ska gå till och vad den ska baseras på. Det anges endast att företräde till täg lägen får ges till vissa kategorier (samma förordning 22 § 2)). I praktiken utnyttjas denna möjlighet till prioritering inte (tillämpning av s.k. grandfather’s right har till exempel bedömts hindra utvecklingen av konkurrens på banan). Vad målet innebär i praktiken är inte självklart, inte heller vilka faktorer som ska tillsättas betydelse i den samhällsekonomiska bedömningen.”
Although there is no precise goal of the capacity allocation process it is assumed, in this document, that there is a common understanding that the overall goal aims at regions to be attractive and available for trade and industry and for people to work and live in and possibly with some relation to profit made by some industry.

The vagueness of the overall goal is, of course, unproblematic when one solution, i.e a timetable, to the capacity allocation problem is better than others in all respects. A better satisfaction of the demands of passengers and customers freight services implies a more efficient utilisation of the railway infrastructure. The same holds for a decrease of the costs for railway operators. The vagueness of the overall goal only becomes cumbrous when there are mutually exclusive choices between factors, all of which would make the utilisation of the railway infrastructure more efficient. This is at least true if we are not faced with the possibility to close down unprofitable traffic with public funding.

It is probably the case that there is enough room for changes in the capacity allocation process for making essential and rather uncontroversial improvements with respect to the common understanding of the overall goal. Hence, as long as the project concerns this scope of improvements, the exact formulation of the overall goal of the capacity allocation process is not essential. Moreover if those improvements exist, it is likely that some of them are really good too.

Following this discussion, we may formulate the aims of this project as to find an enhanced format for the applications for train paths that supports availability of necessary information for timetabling, by first and foremost consider improvements that are uncontroversial with respect to the common understanding of the overall goal. For the development of support tools in the continuation of this project, it ought to be the case that more precise formulations of the overall goal can be met by calibrations or minor adjustments.

5 Potential for improvement

5.1 Introduction

The following sections will each point out a separate cause for a need of change in the capacity allocation process. The needs of change are concerned with the following items:

- Global goals versus local decisions in the process
- The predominance of the traffic operator SJ
- The information needed at timetabling
- Connections and flow of goods and passengers
- The time needed in the process of allocating capacity
At a general level the need of change considered here is concerned with the relation between the overall capacity allocation goals and the timetabling practise. To begin with we just consider the arguments for that there exists a need for a change. Later on we will sort out what is needed to accomplish a change.

### 5.2 Global goals versus local decisions in the process

For the purpose of timetabling, the Swedish railway network is divided into regions. Different timetable designers prepare the timetables for separate regions. This implies that, in order to eliminate some bad solutions for the entire timetable, the timetable designers of adjacent regions must exchange information. At present, however timetable designers of different regions do not exchange a lot of information.

Let us consider the following example of a bad solution resulting from absence of cooperation among timetable designers of different regions (see figure 3). Assume for instance that a train enters the region $B$ and leaves the region $A$ at time $T$. Assume that it is of great importance that $T$ is within the tight interval $I$ of time for achieving a good solution for the region $B$, whereas achieving a good solution for the region $A$ only requires that $T$ is within a larger interval containing $I$ as a subinterval. In this case, if the timetable designers do not speak with each other, a bad solution may of course be chosen, i.e. the timetable designer of region $A$ chose $T$ to be outside the good interval $I$ and the timetable designer of region $B$ does not check if that is possible to change.

The current scheme with regions that have the complete responsibility for making a timetable for their part of the network is currently under discussion. It leads to problematic hand-overs at the borders. Another approach is to focus on the trains, and letting
different timetable designers focus on the complete journey for the train. The advantage of this approach (sometimes referred to as corridors or channels) is that someone is taking the responsibility for the complete picture for the train, but the drawback is the quite complex situation where many designers are involved when many trains passes through a small area. The best approach may be something in between, e.g. some kind of customer support for the whole journey, and regional responsibility (or regional moderator) where it is needed.

5.3 The predominance of the traffic operator SJ

Although Tågtrafikledningen is responsible for allocating railway capacity and Banverket Trafik are commissioned to prepare the timetable, the operators of passenger traffic, particularly SJ, at present and in practise predominates timetabling. Partly it is a consequence of the high priority given to passenger trains and partly because SJ’s application contains a lot of detailed information which makes the designers having difficulties to overlook the effects of changes, especially when the timetabling process gets short on time.

Another reason is that the timetable designers cannot easily foresee the consequences, as discussed above, of allocating train paths to SJ other than those asked for. A reason for this, in turn, is that SJ’s transports makes up a large and complex net of connections.

Moreover, although the true requests on the timetable concerns connections between trains these are at present mainly left out of the application for train paths and have to be rediscovered by the timetable designers, as implicit request in the applications. Hence, at present, it is not easy to use the full strength of a hypothetical powerful tool to optimise the timetable for e.g. efficient resource utilisation.

5.4 The information needed in timetabling

At present the timetable designers inevitably need to evaluate the importance of requests received in order to understand what solutions are possible and desirable. Hence, it is important that the designers are aware of the underlying causes of the requests. Those causes are not always stated in the applications. As an example, a request on a very tight interval of time for an arrival of a train might be caused by the fact that the train is part of a customers production line. Today the ability for timetable designers to make correct judgements relies heavily on implicit knowledge on Swedish railway transports and experience of timetabling.

5.5 The time needed in the process of allocating capacity

The timetable designers use the support tool TrainPlan to complete the work with the timetable. There are no optimisation tools or tools for analysing the overall good for the traffic as a whole, which is the basis for making decisions in allocation by instruction,
and TrainPlan doesn’t provide that either. There is more or less only time for one attempt to create the timetable, although changes are made when the conditions are changed.

5.6 The timetabling scheduling problem

Timetabling is a hard scheduling problem

The capacity allocation process is complex also in that it inherits a hard scheduling problem. That is, if all the true requests are known and it is clear and simple how to evaluate a given timetable, then it is still hard to find the optimal timetable. What makes the scheduling of train paths a problem at all is that the railway network is partly congested.

The problem of finding an optimal timetable belongs to the class of highly constrained resource scheduling problems which is inherently NP-complete. Moreover, what might make the problem even much harder in practise is that the train paths to be scheduled are not independent and a lot of the important requests on the timetable concerns connections between trains. Requests on connections between trains may, for instance, originate from passenger transfers and engine circuits. Requests on connections may also encode solutions to more complex constraints such as that there must be some maintenance of an engines after at most $k$ kilometres. Since a lot of the true requirements concerns connections, it is in general difficult to foresee the consequences of one change in the timetable, e.g. delaying a departure of one train, since this may affect a lot of other trains.

5.7 The network statement, slots and train paths

Description of that what basically is applied for in applications for capacity

Preceded by the capacity allocation process is the task to determine the current standard of the railway network. This task is completed by the publication of the $^{10}$Network Statement. The Network Statement is currently the responsibility of Banverket Marknad. The task of updating the TrainPlan data is currently commissioned to Green Cargo.

Let us from now on use the concept of train to denote a rolling stock associated with a transportation task. Even though this concept is not all that precise, the operators applications concerns trains. Trains have fixed departure and arrival locations and fixed paths. The points in time for departure and arrival are sometimes more and sometimes less absolute.

$^{10}$The official name of the network statement is "järnvägsnätsbeskrivning", although often verbally referred to as "infran."
By the network statement the railway network is divided into the elementary parts subject to the allocation of capacity. Such a part of the railway network together with a time interval is a slot. Based on predictions of customer demands and contracts agreed on the operators applies for slot sequences, each slot sequence being the time and tracks occupied by a train. Such a slot sequence is a train path.

An application for capacity is roughly a set of train paths together with additional demands and information, e.g. properties of the trains such as weight, speed and associated transportation tasks.

6 Issues of concern for the parties involved

This section summarise interviews with some people involved in the train path application process: Thomas Franzén (Banverket Trafik), Bengt-Erik Thörnström (Banverket Trafik), Johan Nilsson (Green Cargo), Jürgen Waldhecker (Green Cargo), Bertil Hellgren (SJ). Another source has been meetings with the project group of “Principer för tilldelning av tåglägen”[Fra03]. When we refer to the companies here, we actually refer to answers from these people and/or meetings with the PTAT group. We would like to point out that some opinions may be the persons own, and not the companies official point of view. Nevertheless, we think that it is important to cite the opinions people working inside the timetabling process put forth.

6.1 Issues of concern for traffic operators

For the traffic operators one of the key issues is to be profitable by supplying their costumers with transportation services. Of course, there are many similarities as well as differences between the two classes of operators regarding what they find important in the timetabling process.

6.1.1 Some general remarks

From the traffic operator’s point of view it is important that all timetable designers follow the same principles, and that these principles are known to the traffic operators. There are several reasons for this, e.g. the decisions must be assessable afterwards so that the operators understand why one of them got prioritised over another and the traffic operators must know in advance what the rules are so they can judge the possibility of getting a certain train path. Since the prioritisation principles should be based on “the common good” (i.e. maximise the society’s profit, in some sense, see section 3), it is important that the basis for these principles are known for the traffic operators.

Both SJ and Green Cargo point out problems with the current scheme where the allocation body partition the total timetabling task into timetable regions, where they then make the timetable fairly independently of each other. Green Cargo gives examples of
Figure 4: Accumulated delay and evenly distributed delay. Two alternatives, the red (left) representing the accumulated delay approach and the blue (right) represent the evenly distributed delay approach (the dotted arrows represent the fastest possible traversal). The effects of a disturbance is plotted by the thin black arrows, we can see that the left approach is able to arrive at the destination in time while the right one is late.

trains that travel through one region without large disturbances, but comes to a long halt at the first meeting station when entering a new region (see figure 3), and SJ disagree with Banverket about if the train should get a delay time equally distributed in each timetable region or get the delay at the end of the journey (“it was much better when the trains could have the extra time in the end of the trip, since then the possibilities to close in on the arrival time was much better in case of disturbances”, see figure 4 for a graphical presentation of SJ’s view). The two approaches have their pros and cons, and what is best is being debated, the accumulated approach is better to ensure correct arrival time at destination, while the other may be better for single track lines with many meetings and overtakes (i.e. great risk for small disturbances).

Green Cargo emphasise the many hours that they are standing still waiting for others to pass. They estimate that they spend 40 000 hours waiting for others to pass, something that not only costs in terms of driver time but also in terms of engines that are tied up and not available elsewhere in the net (often at times when they are most needed). This problem is related to the interfacing problems between the different timetabling regions.

Both SJ and Green Cargo advocate a process where the traffic operators first negotiate themselves, to find solutions for conflicts which may not be local to two trains, e.g. they could in negotiations give up some desirable train paths if they in exchange get other, for them, more crucial train paths. In this setting, Banverket should only be brought into the discussion when the traffic operators cannot solve the problem themselves. This negotiation is in opposition to “järnvägsutredning” [När02] (and the EU regulation), where it is stated that the allocation body should do all the timetable design, and the
traffic operators should not even be allowed to negotiate between each other, at least not in this fashion.

Nevertheless, the operators do have informal discussions and inform each other during their initial phases of their planning processes, even though these negotiations rarely have contractual status. The purpose is to inform each other about seasonal differences and to get the cooperative traffic (i.e. connections) to work.

Both Green Cargo and SJ points out that it is a large problem that the timetables (e.g. the first proposals) are incomplete. All conflicts are not visible in the first proposals since not all trains are there, and thus the conflicts become apparent too late in the process. The traffic operators both speculate over what causes these problems. Is it the recent introduction of the TrainPlan system and (lack of) new methodologies connected with this, or is it because of the complexity inherent in the problem itself? It is not clear how this could be avoided and at the same time reduce the lead times in the process.

Both traffic operators also point out that it may be feasible for them to have conflicts in the timetable, but clear prioritisation rules that unites the conflict if it appears in reality (the cargo trains very often often use different but similar train paths than those applied for\textsuperscript{11}). To handle such unregulated conflicts left in the timetable, one would be useful to able to prove the existence efficient solutions for a number of possible scenarios, so that the unsolved problem is not pushed to the dispatching control. Nevertheless, the operators would rather see the conflicts in the proposals than be unaware of that they exists.

The trend for freight traffic currently is larger and larger trains, while for passenger trains they are shorter but departs more frequently. A consequence of this is that the traffic gets more inhomogeneous.

6.1.2 Passenger traffic

For a passenger traffic company, the competitive strength comes from a number of factors, and also differs between interregional trains and regional trains. The objective for the passenger train operator is an efficient implementation of the traffic that realises the predicted passenger flows and meets the demands.

An important issue is the speed of the train. Passengers valuate travel time as perhaps one key value of the transport service. Generally it is preferable to move a complete train path rather than introduce delays (i.e. keep travel time low and adjust both departure and arrival times), or, put in another way, “High speed trains are high speed trains since they travel at high speed”. This does not mean that the duration is all that matters, in particular departure and arrival times are subject to psychological factors, such as “arriving before 22”, and similar informal and psychological “requirements”.

\textsuperscript{11}It could be worthwhile to investigate if the requirement of cargo operators should take the form of flow capacity as a function of time.
The passenger operator attempts to supply a passenger demand while maintaining reasonable costs for e.g. vehicles and personnel which depends strongly on the complex properties of the mix of train paths (“the network”) used. Since passenger services is regarded as representing high overall good, the requirements of the passenger traffic operators have traditionally been given the highest priority. Focus within the organisations traffic operators have recently begin to shift toward giving priority to interchanges and connections rather than specific train paths which should ensure a the existence of a mix of train paths efficiently implement the total passenger flows. An important issue is therefore not only to take care of the trains themselves but to be able to estimate and compare the priorities of traditional train path requirements with those of interchange requirements. The interchange requirements often take the form of long and sometimes closed chains of interdependencies (one train indirectly affects many others).

Passenger trains are often classified into the following categories:

**Regional trains**  Trains that travel within 100 km, often with many stops. Some, as the shuttle Arlanda - Stockholm, travel fast, others, such as commuter trains, travel slower. For these, it is common to argue for regular departures, i.e. a timetable with repeating patterns (“*stiv tidtabell*”).

**Interregional trains**  Trains that travel more than 100 km. These are further partitioned into fast interregional trains (X2000) and slower (engine pulled) trains.

The importance of requirements on basic interval timetable is being debated. There seems to be agreement on that it is more important for regional trains than for interregional trains. A common view is also that the importance of each individual train being part of the basic interval depends on the number of passenger the train is expected to have, or, in other words, for a train with few expected passengers there may be deviations from the pattern. As noted in [Lin02] there is currently no valuation of a basic interval timetable. It is not part of the current models, and this in turn is most certainly because these are targeted at infrastructure assessments. The tentative lower limit for when one should consider e.g. a stop on a station (and thus increasing the travel time) could be about 10 passengers.

SJ does not advocate a “train path market”, i.e. creating an “market” on of train paths on which the train operators could place bids as in a auction. The reason for this is that the market is not well balanced and not well behaved (e.g. the presence of traffic that is subsidised will make the market unbalanced since the cost for that traffic is high, and the cost for the slots relatively low, therefore the cost of getting a high priority slot is relatively cheap for this traffic).

Crew rosters are made late in the process, and is currently not used to pose restrictions in the train path application process.

The following list summarise some of the most important issues for the passenger traffic operator.

---

12i.e. regular departures on identical paths, in Swedish "*stiv tidtabell*"
**Speed/travel time** The time is a key factor for choosing to travel by train rather than by road or air. The passengers want as few train changes as possible, and if a train change is needed, then it should be efficient.

**Tight net** The passenger exchanges should be tight

**Timeliness** The train should leave on time, and arrive on time. Psychology is important, 2:05 hours in travel time can be judged by the customers as much more than 1:58. Another example is that it feels too late to arrive at 22:04 (i.e. after 22), but 21:57 may be judged as good.

**Conservative changes** Changes in the time table should preferably be conservative, and not be issued unpredictably (i.e. the customer wants a published time table, and wants to be able to rely on it).

**The passenger operator’s internal planning process** The work with producing the train path application starts about one year before the actual train path application is made, and normally the newly published timetable serves as the starting point for process of the coming year. The first draft is supplied to Banverket in November. Work on new contracts begin in June, and are mostly finished one year before the traffic starts, and thus fits fairly well with the planning process. For the planning process to work smoothly, the network statement (or at least a draft on how the network statement is going to be) should be published earlier.

Timetable scheduling for interregional trains is made centrally, but many regional solutions are planned locally. This is natural since each business case take care of their own costs, and consequently they form individual production “divisions”. SJ has three regional divisions, Svealand, Götaland, Fjärrtrafik, each having a number of business cases, and a fourth, Production (and additional three which are not relevant here).

Car circuits and engine circuits are introduced early in the planning process. Engines circuits are mostly a function of the car flows (they follow the cars). Some of the vehicles moves around, often over large areas, but a lot of them are bound to one line or one business case (e.g. X2000 between Stockholm and Sundsvall that uses 5 cars, and different sets of “Reginatåg” which are often bound to the region in which the business case is). International traffic has their own vehicles, as they have special configurations.

Some of the circuits are made available to Banverket, e.g. for the Stockholm-Uppsala commuter train, Regina-systems, X2 circuits, but circuits for RC-engines are not normally part of the train path application, since it is very difficult to make the circuits without detailed information about how the large and complex stations are scheduled and this information is not available yet. The minimum time taken to perform an engine turn can vary from up to 2 hours on a large station for an RC, down to 20-25 minutes for an X2 (in emergency cases 8 minutes suffices).

A comment here is that it is unclear who has the responsibility to judge what is a fair valuation of the turn time, is it the traffic operator himself, or is it perhaps Banverket.
The operator wants to keep down the turn time to save resources, whereas the train dispatcher wants the traffic to work smoothly. This issue becomes clear when the operator that has the tight turn times also expects to be prioritised when there are disturbances.

For SJ’s internal process, it is important that their timetable system is integrated with a lot of post-processing systems such as production the published timetable. They therefore still use TIPS system, but plan to convert to TrainPlan (which Banverket and Green Cargo use).

Negotiations with regional operators are made early in the process to take care of the main passenger flows. SJ has the opinion that it is a clear advantage if the traffic operators (mainly passenger traffic but also cargo traffic operators) first make such bilateral conflict negotiations before the applications are submitted to Banverket.

There are also a lot of contracts that regulates the “traffic cooperation” (samtrafiken) and hard constraints on train interchanges that makes it advantageous to cooperate between the passenger traffic operators. Because of this, the synchronised flow of traffic trough the net built up by the passenger trains is very large and complex. A small change somewhere may lead to very large consequences elsewhere. Most of these interchanges are today implicit in the application which make modifications introduced by the Banverket very controversial. In [kE99] we find a list of such contracts.

6.1.3 Freight traffic

For freight traffic, the issues of concern regarding train paths and the allocation of them stems both from business from known customers and predicted business from future customers. Freight traffic demand has a very dynamic character, and e.g. Green Cargo would rather see more frequent timetable shifts than now (once a year). Also, almost all contracts are renewed in January or February (the most common length of a contract is one year), which means that they are renewed after the actual new timetable is released, which means that many trains that are applied for are not based on fixed contracts but more on estimates and the state of the current contract negotiations. Since e.g. departures, volumes and other parameters important in the train path application are being negotiated with the customers, it is almost certain that not everything will be as predicted in the late autumn. The volumes also varies over the year, depending on both the customer’s variations and seasonal variations. Planners at Green Cargo appear to regard the quality of the network statement as acceptable, but points out that it is published to late. Green Cargo would like to see a much more iterative process, with a preliminary infrastructure presented around the 1:st of October.

It is the cargo flows that is important to satisfy. Every day each cargo flow must be satisfied in terms of trains and shunting, but often there exist more than one implementation in terms of trains and shuntings for a car to get from an origin $A$ to a destination $B$. An efficient implementation of this dynamic behaviour is crucial for the success for cargo companies.

Green Cargo is willing to submit a lot of data in the application, such as e.g. track assignment on shunting stations, shunting relations between trains, engine circuit rela-
tions, etc. They also do that today, around 500 associations with the application. The application is made by submitting a TrainPlan file to Banverket Trafik. There is however an administrative problem with this: Since Banverket Trafik does not actually use the TrainPlan file that Green Cargo has submitted, but rather copy every single train and insert it into their working timetable material, the information that Green Cargo submitted which relates the train to each other is lost (see figure 5). It is not only that Banverket loses the information in their working material, but Green Cargo also lose information when they get the published timetable back as a TrainPlan file, since it is not possible to map the timetable produced by Banverket back onto their own submitted material. For this process to work smoothly between the different parties, this issue must be solved, as well as similar issues related to information passing and coordination.

Freight trains can be classified into the following categories.

**Carriage cargo** Single consignments: each car may come from a different customer (Green Cargo does not sell anything less than a car), and the trains are configured dynamically (i.e. the freight company applies for a train path, but the complete picture of what cargo that goes into the train is not known beforehand). These trains are shunted together, and which train that should be applied for is largely determined on predicted flows and the implementation of these flows in terms of trains between shunting yards.

**System trains** A complete train, or set of trains, is hired by the customer. The customer requirements determines the application for train paths. The train is often seen as part of the company’s internal logistic; it leaves from the departure station and should arrive at the arrival station just in time for the components/material to be used in the company’s manufacturing (examples are Volvo and SSAB). The transportations are completely demand driven, and large variations are common (one example is wood transports; when it is hard to get the timber out of the forest due to spring flood, the demand for wood transports goes down, another example is “ice winters” when ships have difficulties to reach the ports in north-

---

![Diagram](image-url)

**Figure 5**: Data loss in allocation process

Freight trains can be classified into the following categories.

**Carriage cargo** Single consignments: each car may come from a different customer (Green Cargo does not sell anything less than a car), and the trains are configured dynamically (i.e. the freight company applies for a train path, but the complete picture of what cargo that goes into the train is not known beforehand). These trains are shunted together, and which train that should be applied for is largely determined on predicted flows and the implementation of these flows in terms of trains between shunting yards.

**System trains** A complete train, or set of trains, is hired by the customer. The customer requirements determines the application for train paths. The train is often seen as part of the company’s internal logistic; it leaves from the departure station and should arrive at the arrival station just in time for the components/material to be used in the company’s manufacturing (examples are Volvo and SSAB). The transportations are completely demand driven, and large variations are common (one example is wood transports; when it is hard to get the timber out of the forest due to spring flood, the demand for wood transports goes down, another example is “ice winters” when ships have difficulties to reach the ports in north-
ern Sweden and more transportation is performed by train, and the number of cargo trains increases).

**Combi trains** Container train, i.e. they are not shunted but travel directly from one origin container terminal to another destination terminal. The cargo content may differ with the containers and customers.

**Terminal trains** One train a day, that goes from some smaller yard or loading place to a large shunting station.

**Postal trains** Fast cargo trains (currently mail), very sensitive for delays.

Since demand varies over time, freight companies need to adjust the traffic to the demand, not only by making the trains shorter than applied for but also by cancelling trains (which is quite common both for “carriage cargo” and for “system trains”), merging trains or splitting trains. This can be compared to road traffic, where the shipping agent can adjust the number of lorries to the demand. To make the adjustment better, freight companies applies for “Extra paths” (E-paths), which may be used when demand changes, e.g. 2 ordinary trains (applied paths) may be replaced by one train running in an E-path when demand is low. Some of the E-paths and applied trains are mutually exclusive, i.e. they will never be utilised at the same time (day). The actual need for the freight company is closer time table shifts, and the E-paths should be looked upon as a way for the freight companies to have possibilities for demand adjustments built into the timetable.

The most important requirements are those on:

**Arrival times** The customers are mostly interested in that the cargo arrives in time and that the departure time is sufficient for them to be able to deliver the cargo to the train (some companies look upon the transportation as a “rolling store” and part of its internal logistics). The most common goal is to manage over-night transport, e.g. Stockholm-Luleå 16:30 - 9:30 and from mid-Sweden (Sundsvall-Östersund) to southern Sweden, but for long distances, the goal is 2 nights. Green Cargo puts forth demands to increase its traffic during daytime.

**Shunting relations:** A large portion of the cargo cars do not travel directly from departure location to the destination station, instead they are shunted at one or more stations to form new trains. Typically they are being pulled from a small station by a terminal train to one of the large “train building stations” where interregional cargo trains are formed. Then they are being pulled to another shunting station where the cars again are shunted. At the end, they either continue abroad, or end up at the destination, often with a terminal train again. The large shunting yards (Hallsberg, Borlänge, Malmö, Sävenäs in Göteborg, and also Maschen in Germany) need some time to be able to handle the shunting. The general requirements is 2 hours and 20 minutes from the last train in to the first new train out.

---

15 Swedish: E-lägen
16 in Swedish: tågbildningsort
There are also restrictions on how many trains that may arrive or depart within a time window, the preferred limit is 4 trains in and 4 trains out in a floating window of one hour, but it is possible to stress that up to 10 trains (but not in the same direction). Green Cargo schedule the use of the large shunting yards themselves.

International cooperation In order to get smooth international traffic, Green Cargo has a cooperation with Railion (and do not utilise “freight freeways”). Green Cargo feels that their cooperation with Railion works well, and that they get rather good paths for their international traffic to e.g. Maschen.

The dynamic behaviour when making the day-by-day planning may change the planned cargo flows. For carriage cargo, Green Cargo applies for a train with certain properties at the train path application time, but they do not in detail know what the train will contain. And it is only some days before departure that the details are known about how each cargo flow is to be implemented through trains and shunting; there are often several possibilities to implement a flow, as discussed above. Green Cargo has planners working in shifts to determine and plan the flows. The goal is to have at least 80% of the train’s capacity loaded.

Green Cargo has yet not managed to deliver a conflict-free timetable application themselves, “caused by lack of time and/or resources”. Green Cargo delivers with the train path application priorities of the trains, so that Banverket Trafik is aware of which trains that should be prioritised in case they have to put one aside in the timetable design when the conflict is Green Cargo’s own internal.

As already pointed out in section 6.1.1 Green Cargo has the opinion that they get much more waiting time when trains must meet or overtake. In [HC03] a comparison between SJ and Green Cargo is made, SJ and Green Cargo has roughly the same ideal running time, but whereas SJ gets an additional 3760 hours for meetings and overtakings (for a year), Green Cargo gets 37721 hours. Since there is no principles agreed on (yet) about priority, Green Cargo argues that this is unfair. This shows at least that there is a need for prioritisation principles, so that the differences can be motivated by the allocation body.

6.2 Issues of concern for the allocation body

This section is based on interviews and other personal communication with sources from Banverket. The material affects the timetable process in one way or another.

6.2.1 Train path application format

The way and format different operators make their application can sometimes pose problems for the timetable designers. Some, particularly smaller, operators may be
very late with their applications, and the applications are sometimes incomplete. According to some sources, up to 25% of the time is devoted to communication with smaller operators. The format also varies a lot, from different kinds of spreadsheets to TrainPlan files or TIPS files. The larger operators give more details in their applications than smaller ones.

6.2.2 Transportation tasks

The rule is that a train path is assigned to a transportation task. This rule is hard to follow since it is not always easy to know what the transportation task is, and even if the task is known (e.g. a “homogeneous” cargo train between two cargo stations) it is not clear e.g. whether two trains belonging to two different operators applies for the same transportation task (same customer). This poses some problems for the designers since they must look “behind” the application to get the correct prioritisation and to know e.g. if there will only be one train out of two when two operators compete for a transportation task. And when they know, it is not clear how they should act with the competing train paths, should they use both (one will disappear) and thus block the tracks, and if not, which one should they favour. There is however more and more customary that larger industries applies for train paths themselves and hires an operator afterwards to carry out the transport.

6.2.3 Rules of thumb

Good rules of thumb will be of help for the timetable designers, since today they act much on what they feel is the correct division amongst the trains and operators. The regulating documents are very unclear today. The PTAT project (see [Fra03]) addresses this issue. The rules are that traffic that are subsidised should be prioritised (which is debatable), that suburban trains have higher priority than interregional trains, and that passenger trains have higher priority than cargo trains.

6.2.4 Interpretation of the train path application and dependencies between trains

Today the designers interprets the applied train paths for cargo trains as possible to move about 15 minutes before or after the applied time. Within these limits it is judged that the transportation task is fulfilled without any violation of the original requirements. For passenger trains, and in particular for SJ, the timetable designers have difficulties to foresee the consequences when they adjust a train. This stands in contrast to the overall feeling that the designers know the dependencies between the passenger trains, but they lack support (systems) to use the information. As a whole, when adding the short deadlines, this means that there is resistance to move passenger trains and that it is easier to move cargo trains.

For passenger trains it is quite easy to see where there are dependencies between trains, since the application is quite detailed and the stops quite short. For cargo trains the
timetable designers have difficulties, it is not clear which trains that have connections to other trains at shunting stations, (this also varies with time, since Green Cargo operationally reschedules the way cars are routed when the demand changes). The designers want more information about dependencies at the shunting stations.

6.2.5 Heuristics in timetabling process

The time table planning process is often guided from parts of the net where it is felt that there are most problems to get a feasible solution. E.g. “the flying kilometre” south of Mälaren (Svealandsbanan) guides the rest of the timetable process in the region.

The timetable designers feel that the lead times have been larger since the deregulation, but all strives to shorten these again.

The capacity issues are important. Infrastructure should be built with many different traffic patterns in mind, and not be optimised toward a single traffic pattern. If it is optimised, then there is little room for changes and adjustments in the timetable process, something that sooner or later happens (not only changes in the new infrastructure, but the problems can be moved into other tracks, which is not necessary if different traffic patterns had been accounted for).

6.2.6 The TrainPlan planning tool

This section summarise some comments and viewpoints.

The timetable tool (TrainPlan) should be extended to give warnings when associations (or other requirements) are broken, so that the designer can be given appropriate help. It is, however, important that this help does not circumscribe the designer in his task (e.g. it should be possible to override warnings). Furthermore, the tool is not aware of the different velocity limits for e.g. points, which means that the designer must add knowledge themselves about these issues and add traversal time when it is needed. This information should be added. It also judged that the tool has too few measure points to be exact enough. Our own general comment on this issue is that it is often the case that plans (timetables in this case) are made too exact, i.e. that planners often want to plan in more and more detail, regardless if that is warranted by the problem. Care should be taken here so that the “scale” of the data model is on the right level of abstraction, and many planners in industry and elsewhere often tend to strive for a scale 1 to 1, which almost certainly is wrong and not possible to achieve.

The tool needs information about which train routes\textsuperscript{19} that are simultaneous and which are not. Without this information, it is hard for the timetable designer to be sure that the plans are executable. Some kind of simple capacity allocation tool for stations would also be desirable.

\textsuperscript{19}In Swedish: Tägvägar
7 Requirements

7.1 Identified requirements

The requirements and limiting factors on the train path allocation process arise from 4 main sources:

1. Customer requirements
2. Operators requirements
3. Vehicle and cargo constraints
4. Inherent and temporary capacity limits on infrastructure resources

In identifying and analysing a requirement we have tried to point out its original source and its most general form. E.g. most of the customer requirements are seen by the allocation body only through the operators applications which currently need not identify the transportation task served by a train requiring a particular path. We firmly believe that this situation has to change and that many customer requirements has to be made explicit in the allocation process.

The exact form of the requirements listed in this section is preliminary and not yet fully analysed. The mathematical form of the requirements outlined in sections 7.2 and 7.3 is also currently tentative. In particular certain requirements (e.g. vehicle and personnel rotations) have nonlinear form that can make conflict detection and resolution very difficult. This is certainly also the case for the scheduling constraints arising from the limited infrastructure capacity. However these are not part of the application format but can be handled by the allocation body relatively independently from the the operator and customer requirements. In the case of the rotations it may be possible to characterise the requirement not with regard to a specific rotation but rather as some overall property of the timetable that tend to allow for efficient rotations. A similar approach was taken for personnel planning in [MA02] but to apply it to this domain clearly needs more work.

7.1.1 Customer requirements (transportation task)

Most customer requirements involve departures and arrivals at given locations in the rail network and path traversal durations for point to point transports. They generally has one of the forms:

1. Some event must happen before, after or at some particular moment in time
2. Some event must take place during a particular time interval
3. The temporal distance between two events must be of at least, at most or exactly a particular duration (e.g. rigid intervals between events of similar type)
4. The occurrence of several events of similar type must occur at fixed intervals (basic interval timetable)

5. Punctuality (i.e., requirements on the robustness of the timetable under indeterministic perturbations)

7.1.2 The operators requirements

The following resource requirements involve on the resources managed by the traffic operator, mainly vehicles and personnel.

In the infrastructure capacity allocation process used today most of these requirements are implicit and generally not made available to the allocation body. Rather the traffic operator attempts to fulfill a requirement of this type by producing a request for infrastructure capacity that is so rigid that the resource requirement can be reliably estimated or limited. In almost every case the request takes the form of an assignment of fixed moments in time to a large number of timetable events; in practice a finished timetable for the traffic of one operator.

We will argue that making these requirements explicit will make it possible to use less rigid forms for the train path application for infrastructure. We will also argue that the use of less rigid forms of the application will simplify the allocation process and benefit all parties in the process.

Vehicle requirements The most important vehicle requirements arise as a consequence of the operators management of their vehicles into a type of plan usually referred to as vehicle rotations. These consists of sequences of movements which are repeated after a characteristic interval. Each sequence of task is a called a circuit since it generally describes a closed circuit of track link traversals performed by a set of vehicles. The properties of the vehicle rotations strongly influence the number of vehicles and the amount of service transports required by the plan. To produce plans that use the vehicle resources efficiently is one of the most decisive factors in the process of producing cost efficient, attractive and competitive rail transports.

Engine circuits The engines (locomotives) are the most costly of the vehicle resource used by the operator. The cost of the use of the engines stems from a variety of sinks. The most important are

1. Capital or leasing costs
2. Maintenance costs
3. Driver salaries

20In Swedish: styv tidtabell
4. Infrastructure usage

The elementary operation connecting the planned individual movements of two trains using the same engine is called a (engine) turn. The local properties of the turns pose both hard and soft constraints on the arrival and departure events in the timetable. E.g. the arrival of the vehicle to a particular location has to precede its departure from the same location by a duration typical to the location and the involved vehicles. However, if the planned movements involve a service transport of the engine from one location to another, the duration required for the turn can be thought of as including adequate time for the service transport. This type of requirement can be modelled as hard limits on the duration between the arrival and departure events and some type of cost function of the distance from some ideal duration.

However more important than the duration of individual turns are overall (global) properties of a set of sequences of movements that characterise the repeatable patterns of the vehicle flow. The sum of the total duration of each circuit determines the number of vehicles necessary to serve the circuit and structure of the circuit, the amount of passive transport necessary to supply active transports in the circuit.

It is possible to characterise the constraint on the timetable posed by a given circuit (see e.g. [KAL01]) but in reality there may be many alternative sets of circuits that gives approximately the same cost so the requirements from the operator should ideally not consist of a fixed set of circuits but a limit on the cost of the best circuit for any given timetable.

Car flows  For many types of transports the cost of car resource usage is depends heavily of the engine circuits since the the cars follow the engine over large distances. On the other hand service transports of cars constitutes a significant fraction of the planned movements of the engines. This means that the car rotations can be seen an input to the timetabling process. However if these requirements are not made explicit in the process, the exact rotations cannot be determined until the timetable has been fixed.

The requirement on the timetabling process posed by the management of cars takes the form of flow values for different types of cars on the links on the rail network. These flow values should be used to determine which service trains we need to run in order to supply the need for cars at particular times at various point in the network. Exactly how this process is managed today remains to be investigated.

Once the service trains has been determined the car flow requirements are implicit in the requirements on the engine rotations provided that constraints on arrival and departure events of the engines also enforce the connections required by the cars.

21The current policy in Sweden is to bill the operators for infrastructure usage based on number of (actively used) vehicles and total distance travelled. Electricity is billed by type of train, distance and weight, as well as track fees, accident fee, diesel fee and some other fees.
**Personnel requirements**  Specific requirements on the timetable posed by the rotation of personnel are hardly used today. The personnel rotations and parings are invariably produced after the fact, i.e. for a given timetable. The duration of turns may, however, be influenced by generic characteristics of the personnel rotations such that a particular location is or is not suitable for particular type of break in the work schedule of e.g. a driver.

The problem to generate personnel rotations and pairings for a given timetable on railways is considered difficult compared with e.g. air transports. And the total cost of personnel is perhaps even higher than that for vehicles. It would, therefore, be desirable to capture and enforce properties of timetables that allow for creation of efficient personnel rotations. Other work on this problem by the authors of this report include [AK01, MA02].

**Prioritisation**  Prioritisation between different parts of the traffic of a single operator has traditionally not been part of the train path application since this has already been handled in the timetable proposed by operator. When generalising the application format to allow for more flexibility explicit prioritisation become a very important part of the requirements of the operator. To handle priorities in general can be quite complicated and the proposed mechanism (in section 8) is still very tentative.

**Variable capacity and alternate train paths**  The cargo operators plan for traffic which should be able to accommodate varying customer demand which is largely unknown at the time of capacity allocation. To ascertain that they can supply this demand they traditionally apply for sufficient capacity to accommodate a maximum estimated demand, i.e. a capacity allocation that they in practise seldom use. A particularly intriguing example of this is the case where the operator applies for three train paths for a particular route but expect to simultaneously use at most two of them.

This is highly undesirable from the capacity allocators point of view and ongoing discussion indicate that this procedure will be depreciated in the near future. Nevertheless the need is certainly real and should be handled in the way that best serves all involved parties. We have not yet attempted to formalise this type of requirement but acknowledge the need to do this in an efficient fashion.

**Conservative changes**  The passenger traffic operators often advocate conservative changes between timetabling periods. It is debatable how real this demand is from the operators customers. Nevertheless there is certain support for this type of requirement (grandfathers right) in the regulatory documents. See however section 4.6 for a discussion of this and related issues. Even if it is decided that the capacity allocation process should take this type of demands into consideration it is so far unclear how they should formalised and compared to other and perhaps conflicting requirements.

---

22It is not yet clear if and how the the pairing process should be allowed to influence the timetabling process.
7.1.3 Connections and exchanges

These requirements occur both as customer requirements and from how the operators organise the flow of goods through the rail network. We have chosen here to treat them separately from the customer and vehicle requirements since they generally have a very specific form. The following main types have been identified:

- Connections to other transport systems, e.g. at harbours, airports and bus terminals
- Exchange of travellers or goods between transports organised by
  - the same operator
  - another operator
- Attachment or detachment of cars arising from traveller or goods flows

These requirements often take the form of either

1. Limits on the duration between arrival and departure events
2. A minimum overlap between two durations, e.g. stop times of two trains at a particular location
3. The existence of departure for a particular destination within a particular duration from an arrival at a particular location
4. That the occurrence of an event take place at any of a set of alternative intervals, e.g. connection with regular departures at a harbour

7.1.4 Weight, length, and cargo requirements

For public rail transports the routes traversed by transports generally depend directly on customer requirements. That is not necessarily the case for goods transports whose route though the network may vary more widely but may instead be constrained by e.g.:

- its weight
- length
- security considerations for dangerous goods.
Speed requirements

- Fast passenger trains will in some cases require particular high speed corridors.
- Since it is also for capacity reasons a bad idea to mix slow and fast traffic slow goods traffic may be required to take alternate routes. If so, it must be clearly stated in the network statement which routes the slow trains should preferably take.

7.1.5 Infrastructure constraints

For the applicant, the requirements on infrastructure is a function of his demands. These requirements have traditionally been on a low level, i.e. in many cases on signal block level. But since the requirements on infrastructure stems from more high-level requirements such as interchanges and departure/arrival times at stations/shunting yards, it may be the case that the application is made on a more abstract level regarding the actual requirements on the infrastructure. This needs further investigation.

Track capacity  The infrastructure requirements for tracks is simply that track capacity is never exceeded. What the capacity of a particular stretch of track is, however, is far from trivial to determine. See also [KCSÅ01].

Single track (bidirectional traffic)  Track capacity for single track lines is generally given on signalling block level and the capacity is always at most one, i.e. a maximum of one train at a time on one particular block. The capacity may be further reduced in cases of very short blocks or due to limits in the signalling system so that a train traversing a particular block may in fact use capacity of several adjacent blocks.

The capacity of longer stretches of single track depend to a very large extent on the rhythm of traffic direction changes and speeds of the involved trains. It is in most cases pointless to estimate the capacity of longer stretches of single track without also considering a particular pattern of traffic.

Double tracks (unidirectional traffic)  The low level requirement on double tracks is the same as for single track but since traffic is unidirectional the capacity of longer stretches of track may be approximated using a temporal headway that depends only on block traversal times and on the relative speed of the involved trains which is generally quite low.

Station capacity  There are requirements on the stations, e.g. some platforms are dedicated to some operators, some platforms are for traffic going through the station, etc. However, most of the requirements on station capacity are solved once the timetable is made, when the actual arrival and departure times are known.
Meeting stations are handled in the TrainPlan process, but there are seldom explicit requirements on these. There may however be implicit requirements, such as train lengths and weight in the sense that heavy trains may have difficulties to start uphill.

In general it would be desirable to let some type of uniform measure on the capacity of stations to influence the time tabling process.

**Routes** The applicant often has requirements on the route the train should take. The route can be described on several abstraction levels, from the most concrete level which is every signal block and every platform track, to a much more abstract level where the necessary “via locations” are given (i.e. when “something happens” that the operator is aware of, e.g. passenger interchanges, train driver changes, etc.

### 7.2 Mathematical formalisation of requirement primitives

#### 7.2.1 Elementary concepts

The basis of the form we have chosen to express the requirements use a number of elementary concepts of which the most basic is an event in the timetable. The most common events are departures and arrivals of travellers, cargo, vehicles and personnel at locations. Events are momentary and lack duration. Durations are instead modelled as the temporal distances between events.

Events may be required to occur at, before and/or after particular fixed moments times. We regard these as properties of events. The occurrences of events may also be limited by the occurrences of other events. The simplest such case is a binary relation between events requiring e.g. an event to take place only after an other event. In general relations may be non-binary and arbitrarily complex to verify and enforce.

Since the requirements of the operators are typically expressed on events on repetitive cyclic schedules (most often one week) the linear inequalities (≤ and ≥) used below will be used to define higher level concepts more suitable for the requirements engineering in section 7.3 and appendix 3. In some cases, where we need to handle longer cycles (e.g. engine circuits), we will make the cycle length and modulo operations used to define the higher level concepts explicit.

#### 7.2.2 Basic mathematical properties and relations

We conjecture that most of the above mentioned requirements can be expressed or approximated by using only the basic properties and relations listed below where \(x, y\) and \(y_0, \ldots, y_n\) are variables denoting the time point of the occurrence of an event in the timetable and \(c_0, \ldots, c_n\) are fixed time points and \(d_0, \ldots, d_n\) fixed durations. These relations form the basis, while we expect the actual train path application to be stated in a syntactically richer set of language, which facilitates reading and understanding.
Occurrence of event at the earliest and/or latest at fixed times

\[ x \leq c_0 \]
\[ c_0 \leq x \]
\[ c_0 \leq x \leq c_1 \]

Upper and/or lower bounds on duration between two events

\[ x - y \leq d_0 \]
\[ x - y > d_0 \]
\[ d_0 \leq x - y \leq d_1 \]

Occurrence of an event within interval defined by two other events

\[ y_0 + c_0 \leq x \leq y_1 + c_1 \]

The occurrence of a specified number of events before and/or after time points defined by a given event  
Let \( x \) be the occurrence of a particular event (e.g. an arrival at a particular location \( l \)) and let \( y_1, \ldots, y_m \) be a number of events that are candidates to enter into a linear relation (e.g. \( \leq \)) with \( x \) (e.g. departures from \( l \)). The following condition enforce that a minimum number \( k \) of the \( n \) candidates stands in a relation (parametrised by a set of durations \( d_1, \ldots, d_n \)) to \( x \).

\[
(\exists y_{i_1} \cdots \exists y_{i_k}) (x + d_{i_1} \leq y_{i_1}) \land \cdots \land (x + d_{i_k} \leq y_{i_k})
\]

This relation may be useful to giving abstract requirements on e.g. vehicle circuits. A condition of this type could be used to state that there must exist at least one arrival from a particular location that can supply a vehicle to be used by a planned departure. It is not yet clear how difficult such a condition would be to maintain in the time tabling process and if it would suit the operators to state their requirements in this form.

Scheduling constraints on track and station capacity  
Hard limits on infrastructure capacity is usually expressed as disjunctive constraints on the infrastructure resources. These so called scheduling constraints are in general very hard to satisfy and are perhaps the main reason why timetabling is inherently difficult. These constraints are not strictly part of the operator/user requirements but rather limits that the allocation authority must handle in its internal process. They do complicate the process of capacity allocation since it is difficult to determine if any of the user requirements are in conflict with the infrastructure resource limits but do not directly enter into the negotiation between operator and capacity allocator. For a technical introduction to models used for infrastructure resource scheduling in the train domain, see [KCO+97].
7.2.3 Weights on properties and relations

All relations above may be extended to handle costs, to be able to express the cost when breaking them. The general form is $R_i \rightarrow C_i = f(\text{vars}(R_i))$ where $f$ is a linear function over the variables in $R_i$ and $i$ is an index over the relations. $\sum_{i=1}^{n} C_i$ expresses how good a particular solution is.

7.3 Mapping the requirements to mathematical primitives

This section attempts to bridge the gap between the general mathematical primitives in the previous section and the requirements discussed in earlier sections. Note that this material only concerns the requirements in the train path application. There is also a need to combine the train path applications, stated in the expressions below, with the infrastructure description to get the material that is actually going to be used in the timetabling design. Also note that the syntax used here is a mathematical one and the actual concrete syntax to be used in an actual train path application will certainly be different.

7.3.1 Notation

In the following we will use the following general notation:

$p$ denotes a location, i.e. everything that can be used as a “way point” in the railway network (stations, important points, etc) i.e everything that has a signature name

$d$ denotes a departure, $d_i$ denotes the departure of train $i$ from its origin location, $d_i^p$ denotes the departure of train $i$ at location (station/signature) $p$

$a$ denotes an arrival, $a_i$ denotes the arrival of train $i$ at its destination location, $a_i^p$ denotes the arrival of train $i$ at location (station/signature) $p$

$e$ denotes an arbitrary event, most often either an arrival or a departure

7.3.2 Trains

Trains should have the data given in table 1 (note that we only include the minimum requirements material passed from the operator, that is relevant from a timetable planning point of view. Thus we omit e.g. the basic time table, “bastidtabell”, which is a function from originating departure, the train type and the given restrictions and may be calculated e.g. by the software GTP available at Banverket). The list is partly based on material from the project “principer för tilldelning av täglägen” [Fra03].

If $R$ is empty the standard values stemming from $GP$ is used. Requirements $R$ in table 1 take the following general form:
<table>
<thead>
<tr>
<th>Field</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>Train number</td>
<td>Unique number for the whole transportation task</td>
</tr>
<tr>
<td>$D$</td>
<td>Dates</td>
<td>The days of the week that the train should leave (this should preferably be a part of $S$, since there are so many similarities. The only difference is that $D$ has the same $i$ over the dates, while $i$ differs in $S$).</td>
</tr>
<tr>
<td>$f$</td>
<td>Train type</td>
<td>Train type according to a common database (which describes the data of the train, such as speed limit, brake values, etc)</td>
</tr>
<tr>
<td>$T$</td>
<td>Train data</td>
<td>Set of different train data relevant for timetable planning (train length, weight, special characteristics, etc, i.e. everything not covered elsewhere).</td>
</tr>
<tr>
<td>$GP$</td>
<td>Global priority</td>
<td>Priority class according to common definition (see [Fra03])</td>
</tr>
<tr>
<td>$IP$</td>
<td>Internal priority</td>
<td>The operators internal priority (any integer, 1 is the most prioritised train)</td>
</tr>
<tr>
<td>$S$</td>
<td>Periodic timetable</td>
<td>Set of pairs, $\langle j, o \rangle$ where $i$ is a train task number and $o$ is an offset in minutes from the current train. This ought to be expressed as a relation between trains, to get a clear distinction between trains and relations between trains.</td>
</tr>
<tr>
<td>$P$</td>
<td>Route</td>
<td>A sequence of “via locations” (station signatures) which the train should pass</td>
</tr>
<tr>
<td>$R$</td>
<td>Requirements</td>
<td>A set of requirements on departure and arrival times, etc</td>
</tr>
</tbody>
</table>

Table 1: Requirements on trains
<table>
<thead>
<tr>
<th>Name</th>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max shift departure</td>
<td>$c_{min} \leq d^P \leq c_{max}$</td>
<td>Maximum shift of departure, for the transport to be of commercial value. Limits may be omitted.</td>
</tr>
<tr>
<td>Max shift arrival</td>
<td>$c_{min} \leq a^P \leq c_{max}$</td>
<td>Maximum shift of arrival, for the transport to be of commercial value. Limits may be omitted.</td>
</tr>
<tr>
<td>Preferred departure</td>
<td>$c_{min} \leq d^P \leq c_{min}$</td>
<td>Preferred departure window, in which all time points are equally good</td>
</tr>
<tr>
<td>Preferred arrival</td>
<td>$c_{min} \leq a^P \leq c_{max}$</td>
<td>Preferred arrival, in which all time points are equally good</td>
</tr>
<tr>
<td>Accumulated traversal time</td>
<td>$c_{min}^{H} \leq \sum_{t=1}^{n} t_t \leq c_{max}^{H}$</td>
<td>Restrictions on the accumulated traversal time. Exactly what should be included in “traversal time” should be clarified.</td>
</tr>
</tbody>
</table>

Table 2: Sample event properties

$(id, r)$ where $id$ is a unique identification number for the requirement\(^ {23}\), and $r$ is a relation that can be translated into one of the basic relations described in 7.2.2. Table 2 gives some sample relations, of which many have been proposed or discussed in the project “principer för tilldelning av tåglägen”[Fra03]. Observe that this is not the proposed syntax, but a mathematical description of possible relations. They should be given meaningful names and clear notation.

By the first four relations it is possible to state both shifts and delays of departure and arrival times. The fifth relation is an example of a relation that is more expressive.

### 7.3.3 Relations

Relations between trains contains the data given in table 3. Note that requirements on a train and some external (not track-bound) transports should be given as a train relation. One could think of a possibility to state relations between a train and e.g, a bus, where the timetable for the bus and train should be made at the same time (as for a train having a relation to another train), but we omit that possibility in this work, to make the material presented here restricted to the train planning process.

$R$ should not be empty. Each $R$ should take the form $(id, r)$ where $r$ takes one of the forms described below. The mathematical formulation is given in table 4.

It is important to note that most of the requirements addressing transportation flows (passenger or cargo) the requirements are stated inside the planning period (typically a

---

\(^{23}\)This identification is useful for later reference in the planning process.
Table 3: Requirements on relations between trains

<table>
<thead>
<tr>
<th>Field</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Identification</td>
<td>Unique identification number for referencing this requirement</td>
</tr>
<tr>
<td>T</td>
<td>Type</td>
<td>Association type (e.g. driver connection, passenger flow, etc).</td>
</tr>
<tr>
<td>C</td>
<td>Priority class</td>
<td>Priority class, based on how valuable the association is (e.g. amount and value of goods, number of passengers, etc).</td>
</tr>
<tr>
<td>L</td>
<td>List of trains</td>
<td>Vector of trains, i.e. time variables are indexed by the index of the train in the vector. The first index is numbered “1”.</td>
</tr>
<tr>
<td>R</td>
<td>Constraint</td>
<td>Set of relations that should be fulfilled</td>
</tr>
</tbody>
</table>

week), while some of the requirements addressing resource requirements (circuit and sequence) the requirements is actually posed over several consecutive instances of the planning period. In appendix ?? we outline a formalisation of how the mathematical statement is to be interpreted to bridge the two views on time, particularly how the terms “before” and “after” used in natural language is to be interpreted.

In the following, $t_i$ denotes a train (number).

**Overlap**, $o(t_1, \ldots, t_n, c_{min}, c_{max}, p)$ Overlap between trains $t_i, \ldots, t_n$, the overlap should be at least $c_{min}$ and at most $c_{max}$, at place $p$.

**Exists overlap**, $v(\langle t_1, \ldots, t_k \rangle, \langle t_{k+1}, \ldots, t_n \rangle, c_{min}, c_{max}, p)$ States that at least 1 train in the first set and 1 train in the second set must have an overlap (see [RAZ00]).

**Before**, $b(e_{t_0}, \langle e_{t_1}, \ldots, e_{t_n} \rangle, c_{min}, c_{max}, p)$ Train $t_0$ should arrive (depart) before all trains $t_1, \ldots, t_n$ with at least $c_{min}$ minutes and at most $c_{max}$ minutes, at place $p$. This may be used to state shunting relations or one way passenger flows

**After**, $a(e_{t_0}, \langle e_{t_1}, \ldots, e_{t_n} \rangle, c_{min}, c_{max}, p)$ Train $t_0$ should depart (arrive) after train $t_1, \ldots, t_n$ with at least $c_{min}$ minutes and at most $c_{max}$ minutes, at place $p$. This may be used to state shunting relations or one way passenger flows.

---

24 It is unclear how often it is useful to state this requirement between more than two trains. If this need is minimal, then it should be simplified to two trains.
<table>
<thead>
<tr>
<th>Name</th>
<th>Notation</th>
<th>Mapping to the primitive constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>overlap</td>
<td>$o(\langle t_1, \ldots, t_n \rangle, c_{\min}, c_{\max}, p)$</td>
<td>$c_{\min} \leq d_{t_j}^p - a_{t_j}^p \leq c_{\max}$ for all $1 \leq i, j \leq n$. Note that this also covers the minimal stop duration for each train at $p$.</td>
</tr>
<tr>
<td>exists overlap</td>
<td>$v(\langle t_1, \ldots, t_k \rangle, \langle t_{k+1}, \ldots, t_n \rangle, c_{\min}, c_{\max}, p)$</td>
<td>$\exists j (1 \leq i \leq k \land k + 1 \leq j \leq n \land o(\langle t_i, t_j \rangle, c_{\min}, c_{\max}, p))$</td>
</tr>
<tr>
<td>before</td>
<td>$b(e_0, \langle e_1, \ldots, e_n \rangle, c_{\min}, c_{\max}, p)$</td>
<td>$e_i^p \leq e_j^p$ for all $1 &lt; i \leq n$</td>
</tr>
<tr>
<td>after</td>
<td>$a(e_0, \langle e_1, \ldots, e_n \rangle, c_{\min}, c_{\max}, p)$</td>
<td>$e_i^p \geq e_j^p$ for all $1 &lt; i \leq n$</td>
</tr>
<tr>
<td>period span</td>
<td>$l(e_1, e_2, k)$</td>
<td>$e_1 - e_2 \leq kZ$, where $Z$ is the length of the planning period (10080 minutes for a week).</td>
</tr>
<tr>
<td>connections</td>
<td>$s(\langle t_1, \ldots, t_n \rangle, \langle c_1, \ldots, c_{n-1} \rangle, \langle p_1, \ldots, p_{n-1} \rangle, k)$</td>
<td>$d_{t_{i+1}}^p \geq a_{t_i}^p + c_i \land t(d_{t_{i+1}}^p, a_{t_i}^p, k)$ for all $1 \leq i &lt; n$</td>
</tr>
<tr>
<td>circuit</td>
<td>$c(\langle t_1, \ldots, t_n \rangle, \langle c_1, \ldots, c_n \rangle, \langle p_1, \ldots, p_n \rangle, k)$</td>
<td>$s(\langle t_1, \ldots, t_n, t_{n+1} \rangle, \langle c_1, \ldots, c_n \rangle, \langle p_1, \ldots, p_n \rangle, k)$ where $t_{n+1}$ is the same train as $t_1$ (same day of week and same train number), but $k$ weeks later than $t_1$.</td>
</tr>
<tr>
<td>train range</td>
<td>$r(\langle t_1, \ldots, t_n \rangle, c_{\min}, c_{\max}, d, P)$</td>
<td>TO BE DETERMINED</td>
</tr>
</tbody>
</table>
The relations in table 4 are illustrated by the figures 6 and 7.

When implementing a support system that can take advantage of these relations, they may be differently used depending on if an operation research approach is taken or a constraint programming approach. In case of a constraint programming approach, these relations are ideal for forming so called global constraints, i.e. encapsulated

25This relation can be implemented as a series of before/after relations, but by giving this high level relation the implementor (timetable constructor) sees the general picture rather than having to construct it.

26This relation is not clear how to mathematically state, nevertheless it is important (“E-paths” is one application, “seasons” is another).
specialised algorithms that take advantage of the explicit structure between the used parameters and variables. In an operations research approach, they will be mapped onto linear equations, and since the basic equations (except 7.2.2) such a mapping exists. All relations in table 4 can be mapped as described in the fourth column, except perhaps the relation *exists overlap* which includes an existence statement. However, the paper [RAZ00] describes how this has been done for the CADANS software used in the Netherlands, which may be applicable here also.

If an operations research approach is taken, these relations, despite being mapped on primitive linear relations, works as a communication media between the operator and the allocation body. The relations expresses in a more comprehensive way what the operator wants to achieve.

### 8 Outline of proposed approach for improvement of allocation process

Please note that this proposal is tentative and has to be verified both with respect to its technical feasibility and acceptability to involved parties!

We propose the use of a format for the application of train paths from the operator which includes:

1. A set of *formal requirements* that encodes the customer and resource requirements. I.e. at least the following:
(a) Earliest and/or latest departure and/or arrival times for complete and/or partial train movements
(b) Connections between pairs of train movements arising from expected traveller or goods flows
(c) Requirements that express conditions that guaranties or characterise timetables that allow for efficient vehicle and personnel rotations

Each timetable that fulfil these requirements should be acceptable to the operator.

2. A fixed timetable that fulfils all these requirements

(a) This timetable should represent the ideal departures and arrivals for each train from the point of view of the operator and/or his customers
(b) It is not generally required to be conflict free with respect to infrastructure resource limitations
(c) It should, however, respect the minimum traversal times (including margins) for each type of train on its path and possibly other limits posed on the requirements set by the allocation authority
(d) For trains that do not, the formal requirements with respect to that train will be ignored in the forthcoming train path allocation process

3. An estimate of the relative cost of various changes in the proposed timetable introduced by the allocation body to solve conflicts arising from to infrastructure resource limitations

(a) This cost estimate should be based on a classification of different type of trains, transports, lines and requirements published as part of the network statement
(b) The cost for changes that do not break any of the formal requirements should depend linearly on the distance from the ideal point in time for each event in the timetable and be relatively low
(c) The cost of changes that do break formal requirements should depend on the type of the requirement, type of involved trains, type of transport, scarceness of involved resources etc as well as on a priority given by the operator

- The last two contributions to the cost should reflect the priorities of the operator but must be weighted to be comparable to similar priorities from other operators

---

27 Note that the exact form of the requirements on rotations is not yet determined
28 C.f. [Fra03]
29 These could include maximum stop times, minimum turn times for engines and other relations between trains.
30 To handle these type of operator requirements efficiently we would propose a decision support system which will be specified in the continuation of this project.
The allocation body must collect the applications from all the operators and identify and resolve conflicts between them. In general this is, of course, nontrivial and to facilitate this process the allocation body must ensure that the requirements are given on a structurally simple form (e.g. linear relations) and probably use to computerised conflict detection system and possibly tools that can suggest and evaluate suggestions for conflict resolution.

We can classify two main types of conflicts:

1. Conflicts between proposed timetables that can be resolved without violating the formal requirements from any operator. We note that:

   (a) The resolution between conflicts of this type should be done subject to minimisation of the cost incurred to the operators
   (b) It is important that the costs are given in a way so that they are comparable between the operators
   (c) The cost estimate given by the operators must be weighted by the allocation body so as to take into account priorities based on estimates of contributions to overall common good for different types of transports
   (d) We argue that this type of conflict is very common and much easier to resolve than the “actual” conflicts described below
   (e) We note that a change initiated to resolve a local conflict can introduce secondary changes, the cost of which has also to be taken into account

2. Conflicts that cannot be resolved without violating formal requirements from at least one operator. We note that:

   (a) These conflicts are actual, that is, expressing conflicting interests between the operators or unresolved conflicts in the requirements of a single operator
   (b) Such conflicts should be resolved taking into account both the its type, the priorities based on estimates of contributions to overall common good for different types of transports and possibly also the priority given to it by the operator
   (c) Possibly one should consider compensating the operator disadvantages by the decision of this type to give a reasonable spread of penalties (overall common good v.s. local prioritisation)
9 Conclusions, results and future work

In this section we attempt to summarise the main conclusions and results of the first year of the project. We also give an indication of the most important aspects of the future directions implied by these results. Note that when it comes to the conclusions some of them are tentative and need further verification, something that we plan to achieve in the continuation of the project.

The foundations for the capacity allocation process is changing rapidly. Laws and regulations that influence the process are being reviewed and revised, both nationally and internationally. The radical deregulation of the rail transport market introduce completely new problems into the capacity allocation process. These problems need to be understood thoroughly and solved efficiently.

There is a strong need for clear and predictable principles for the capacity allocation process. The operators must be able to predict the effect of these principles on their own operations and plan the use of their own resources using these predictions. The timetable designers also need clear and comprehensible procedures to follow. The result of the process should not depend on the division of responsibilities between regions or lines within the allocation body. Nor should it depend on the choice of methods and tools used in the design process.

These principles must also be fair and operator neutral and shall implement the prioritisation given to different types of traffic made by political decisions. The notion of a measurable overall good produced by the transport system is confused. The measure used for infrastructure investment is clearly unsuitable and there are fundamental issues around the connection between the local “rules of thumb” used to distinguish between conflicting train path applications and the global overall good of produced by the traffic as a whole. This issue needs to be thoroughly investigated.

We have also come across several (sometimes conflicting) opinions, e.g. that the process is too slow and/or deadlines too tight, that there is too little time for the timetable designers to produce good results and that the results themselves are unsatisfactory, that together give strong indication that efficient decision support tools and improved methodologies are very much needed in the capacity allocation process. Further analysis is should be able to identify the exact nature of the support tools and type of methods that would improve the process most.

When setting out to define this project, one of main general ideas was that it should be possible improve the capacity allocation process and its result by stating the requirements of the operators and their customers in a more general way than as a fixed timetable. We have come to the conclusion that this approach is still viable even if it needs to be elaborated further to be of practical use. This conclusion is strengthened by the results [Fra03] of the PTAT project (run in parallel with this one) at Banverket Trafik and with a very similar approach. The reactions to the conclusions of that project from the parties concerned has in the main been positive. We have also found that almost all the requirements on the timetable can in fact be stated as properties and relations of events in the timetable which is very encouraging.
The main results of the project so far are:

1. In the accompanying state-of-the-art report [AEK03] we give a summary and analysis of the regulations and other documents and instructions that currently influence the capacity allocation process and some of the methods and technologies available from the scientific community to support the process.

2. In sections 3, 4 and 5 we give an overview and analysis of the current state-of-the-art for capacity allocation in Sweden.

3. In sections 6.1 and 6.2 we have listed and analysed the most important issues raised by some of the larger traffic operators and the allocation body respectively.

4. In section 7.1 we have presented and analysed a list of requirements that we believe captures all important demands on the train paths.

5. In section 7.2 we have presented a list of basic primitives that we have good reasons to believe can be efficiently handled by software tools.

6. In section 7.3 we have presented a first attempt to map the list of requirements onto an abstract mathematical formalism that in turn is mapped onto the basic primitives. This is not complete and needs to be further addressed in the next phase of the project.

7. In section 8, finally, we have outlined a first approach of a revised procedure and protocol for the application process and some of the more important general requirements on the train path application itself.

A number of issues and caveats remain, however:

1. The interpretation of requirements stated on a cyclic schedule (given in Appendix A) needs to be further analysed from a computational point of view and integrated into the main formalisation.

2. We have not yet seriously addressed the question how implement prioritisation between different types of requirements but plan to use the results of [Fra03] as starting point to improve the currently tentative approach presented in section 8.

3. The approach presented in section 8 also give some general ideas on how to ensure that the requirements in the train path application corresponds to the real requirements of the transportation task and does not contain e.g. hidden margins introduced by the operators, but this issue clearly needs more work.

4. Some of the high level requirements that have not yet been fully formalised may turn out to be more difficult to handle from a computational point of view than currently expected.
5. Conflicts between otherwise straightforward requirements may be difficult to detect and resolve since they interact with resource limitations in the infrastructure which are themselves complex to enforce.

6. The number of requirements can become large which may make it difficult to pinpoint and resolve the actual source of conflicts.

Despite these remaining issues we strongly advocate a continuation of the work initiated in the project and argue that these problems can and must be overcome to implement a fair, efficient and maintainable process of capacity allocation.

**Future directions** The next phase of the project will be concerned with resolving the above issues but also and mainly with the actual format of the train path application and the computational aspects of using various possible constructs in that format. The main requirements on the format will be:

1. It should be stated on the appropriate level of abstraction, so that it works as a good communication link between the planners at the traffic operators and the timetable designers.
2. The mathematical interpretation of the constructs used in the format should have acceptable computational properties.
3. It should be extensible to allow for future revisions of procedures and protocols.
4. It should not be possible, or at least very difficult, for traffic operators to misuse the format to e.g. gain a competitive advantage.
5. The information contained in the application should (as far as is feasible) be possible to use also for prioritisation in the dispatching process, e.g. in the case of disruptions of the planned traffic.

**References**


A Requirements on a periodic timetable

Elsewhere in this document are described requirements as if a non-periodic timetable is concerned. That is, these, let us say non-periodic requirements, are requirements on events to be scheduled, each of which takes place only once, when put in operation. The capacity allocation process of today basically considers a periodically repeated timetable, where the period is one week long. That is, the process considers, let us say, periodic requirements, on events to be scheduled, each of which takes place once in each repetition of the timetable period, when the timetable is put in operation.

The aim of this appendix is to show how requirements, given in a non-periodic style, can be interpreted as periodic requirements. It is not the aim of this appendix to give requirements a form which allow them to be efficiently handled by support tools.

A.1 Non-periodic requirements

We begin by considering the non-periodic overlap and connections requirements. The requirements have been presented elsewhere in this document although perhaps not with exactly the same formal definition.

Let \( t = \langle t_1, \ldots, t_n \rangle \) be a sequence of trains and let \( a_i^p \) and \( d_i^p \) be the points in time the when train \( t_i \) arrives and departures, respectively, from place \( p \). The non-periodic overlap requirement \( np\text{-overlap}(t, c_1, c_2, p) \) is that, for any two trains \( t_i \) and \( t_j \) in \( t \) the duration of the time interval both trains stand still at place \( p \) is bound by lower and upper bounds \( c_1 \) and \( c_2 \) respectively. The requirement is given by

\[
np\text{-overlap}(t, c_1, c_2, p) = c_1 \leq d_i^p - a_j^p \leq c_2, \text{ for all } 1 \leq i, j \leq n
\]

Let \( e = \langle e_1, \ldots, e_{n-1} \rangle \) and \( p = \langle p_1, \ldots, p_{n-1} \rangle \) be sequences of \( n-1 \) durations and places respectively. The non-periodic connection requirement is concerned with the possibility of connecting every two successive trains in \( t \) to each other, for example letting the sequence \( t \) constitute a train path.

\[
np\text{-connections}(t, e, p) = d_i^{p_i} - d_{i+1}^{p_{i+1}} \geq e_i, \text{ for all } 1 \leq i < n
\]

This requirement has a stronger form that also gives an upper bound \( C \) of the total duration of the sequence of connections.

\[
np\text{-connections}(t, e, p, C) = d_i^{p_i} - d_{i+1}^{p_{i+1}} \geq e_i \& d_{n-1}^{p_{n-1}} - d_1^{p_1} < C, \text{ for all } 1 \leq i < n
\]
A.2  A close point interpretation

Assume in the following that the requirements concerns a certain periodic time-table. Let $T$ be the timetable period. If standard weeks (Swedish: typveckor) are considered, then $T$ is a week (or the number of seconds or something in a week). In the following we only consider points in time in the periodic timetable and we simply use points to refer to them. For all points $x$ we have $0 \leq x < T$. Given any point $x$ we obtain a set of instances of $x$, when the timetable is unfolded. We use $x(1), x(2), \ldots$ if explicit reference to those instances is needed. In the following an instance of a point refers to a point in time in the unfolded timetable. For an instance $x(k)$ of the point $x$ we have $(k-1) \times T \leq x(k) < k \times T$.

Requirements most often consider instances of points that in time are close to each other compared to $\frac{T}{4}$. The close point interpretation on a requirement on two points, $x$ and $y$ say, is simply to assume that the closest instances of $x$ and $y$ is concerned, if $x$ and $y$ are close. For instance, let the timetable period be $T = 1000$ and let two points be $x = 1$ and $y = 999$ and consider the requirement $1 < \frac{|x-y|}{T} < 5$, where the component of the requirement in non-periodic style is surrounded by square brackets. The two closest instances is obtained by, for example, considering $x(3)$ and $y(2)$. That is $x$- and $y$-instances are considered such that the $x$-instance period succeeds the $y$-instance period. Hence, $|x-y|$ is $x-y+T$, which equals 2 and hence the requirement holds.

A.3  Close points durations

For some of the requirements we need only to interpret the duration in between two arbitrarily given close points and in this section we deal with this interpretation. Let us say that two points is a straight pair if there are instances of them that are close and belongs to the same timetable period. Let $\text{straight}(x, y)$ denote that the points $x$ and $y$ is a straight pair, that is

$$\text{straight}(x, y) = |x-y| < \frac{T}{4}$$

Let us say that two points is a bordering pair if there are instances of them that are close and belongs to two consecutive timetable periods. Let $\text{border}(x, y)$ denote that the points $x$ and $y$ is a bordering pair, that is

$$\text{border}(x, y) = |x-y| > \frac{3T}{4}$$

Notice that a pair of points can never be both a straight and a bordering pair. Let $\text{close}(x, y)$ denote that $x$ and $y$ are close points, that is

$$\text{close}(x, y) = \text{straight}(x, y) \text{ or border}(x, y)$$

We obtain $|x - y|$, the close point interpretation of the duration from $x$ to $y$, by
A.4 Close points sequence duration

In the following let us just for simplicity, use $d_i$ to denote $d_i^P_i$. Using this notation a component of the stronger form of the non-periodic connections requirement is $d_{n-1} - d_1 < C$. Now, let us consider $d_1, \ldots, d_n$ as points (in time in the periodic timetable) and assume that the points $d_1$ and $d_{n-1}$ are not close. Then, how shall we interpret the requirement $[d_{n-1} - d_1] < C$? We may use the intermediate points, that is $d_2, \ldots, d_{n-2}$, in between $d_1$ and $d_{n-1}$ in order to do this, under the assumption that for all $i, 1 \leq i < n, d_i$ and $d_{i+1}$ are close. Let $d = \langle d_1, \ldots, d_{n-1} \rangle$ and let $L(d)$ be

$$L(d) = \sum_{i=1}^{n-2} [d_{i+1} - d_i]$$

That is, $L(d)$ is the close point interpretation of the duration from $d_1$ to $d_{n-1}$, given by a sequence of consecutive close pairs of intermediate points. Hence,

$$[d_{n-1} - d_1] < C = L(d) < C$$

A.5 Close points overlap and connections

By the close points interpretation of duration we obtain the following close point interpretations of the requirements overlap and connections

$$\text{overlap}(t, c_1, c_2, p) = c_1 \leq [d_i^p - a_j^p] \leq c_2, \text{ for all } 1 \leq i, j \leq n$$

$$\text{connections}(t, c, p) = [d_i^p - a_{i+1}^p] \geq c_i, \text{ for all } 1 \leq i < n$$

$$\text{connections}(t, c, p, C) = [d_i^p - a_{i+1}^p] \geq c_i \& L(d) < C, \text{ for all } 1 \leq i < n$$

\[50\]
A.6 Close points before and after

Also other requirements such as the point \( x \) is before \( y \) may be given close point interpretations similar to the durations interpretation. The interpretations of before and after are the following:

\[
[x < y] = \begin{cases} 
  x < y, & \text{if straight}(x, y) \\
  x > y, & \text{if border}(x, y) \\
  \text{undefined otherwise}
\end{cases}
\]

\[
[x > y] = \begin{cases} 
  x > y, & \text{if straight}(x, y) \\
  x < y, & \text{if border}(x, y) \\
  \text{undefined otherwise}
\end{cases}
\]