The Road to Incremental Allocation & Incremental Planning

Content and Potential

Martin Aronsson, Malin Forsgren, Sara Gestrelius

Swedish Institute of Computer Science

September 2012

SICS Technical Report T2012:09
ISSN 1100-3154
The Road to Incremental Allocation and Incremental planning

Content and potential

Martin Aronsson, Malin Forsgren, Sara Gestrelius
Swedish Institute of Computer Science (SICS)
Email: {martin, malin, sarag@sics.se}

September 2012

Abstract
This document summarizes how SICS envision “The road to Incremental Allocation and Incremental Planning” and is in essence a more detailed description of what SICS formulated as “the road to the dynamic train plan” in the project Den Dynamiska Tågplanen (The Dynamic Train Plan). Moreover, this document summarizes some of the results and tests that have been performed in the projects Tågplan 2015 (Train Plan 2015) and Marackasen (The Maraca). These results support the reasoning behind the first steps of the road to Incremental Allocation and Incremental Planning.

Introduction
At Trafikverket (the Swedish Transport Administration), there are several related, ongoing projects and initiatives collectively referred to as Incremental Allocation, Successiv tilldelning (ST) in Swedish, and Incremental Allocation, Successiv planering (SP). They spring from a series of research projects carried out at SICS (Swedish Institute of Computer Science) from the year 2005 and onwards. The research was initially funded by Banverket (the Swedish Rail Administration) and then by the newly formed government agency Trafikverket which took over the responsibilities of Banverket in 2010.

The purpose of ST and SP at Trafikverket is to develop and implement a process founded on a new methodology for planning railway capacity. This methodology aims at enabling more efficient use of existing capacity and resources.

A corner stone in the new methodology is the distinction between delivery commitments on the one hand, and production on the other hand. The relationship between the two is also central to the methodology: Delivery commitments define the goals for the production. With this in mind, the purpose of a planning activity depends on in what part of the process it is carried out: Long term planning aims at establishing the delivery commitments while production planning aims at fulfilling the commitments.
This document consists of two parts. The first describes a number of concrete steps that we as researchers believe will ensure a smooth transition from the current planning approach towards the planning process that Trafikverket aim for in ST and SP. The second part of the document describes some of the benefits of adopting the new approach using some illustrative examples.

The steps towards the new planning approach

The following steps, or stages, towards the new planning approach are based on the assumption that it is not wise to introduce too many big changes at the same time. Big values are at stake in the train plan process and every change needs to be subject to quality assurance. It is important that everyone involved will have the chance to adapt to each change and thereby feel secure in his/her own new role before the next change is introduced.

The stages below are numbered in the order in which we, as researchers, assume they will be implemented.

Stage 0

*Single out the delivery commitments from the production specifics in the train plan.*

The train plan (tågplanen in Swedish), i.e. the yearly plan that is produced in today's process, is essentially a production plan. This stage involves making all relevant parties aware of the distinction between delivery commitments and production specific details in the train plan. Once this stage has been implemented, the railway undertakings (RU:s) will receive a set of delivery commitments as reply to their application instead of the detailed train paths.

The process of constructing the yearly train plan, the so called train plan process or long term process, is not changed at this initial stage, nor is the ad hoc process that follows the train plan process. This is why we call it Stage 0 instead of Stage 1: concepts are introduced, but no changes to the main processes need to be made.

By talking about delivery commitments instead of the train plan, especially in TRAV (the contract: Trafikeringsavtal), the foundation for the following steps is created. The change introduced at this stage affects the RU:s, and it is very important to gain approval from them for the idea of separating the delivery commitments from the production specifics.

Stage 1

Provided that Stage 0 has been implemented, Stage 1 assumes a process that is essentially identical to the current one, but with one major difference:

*Instead of setting the production specifics in stone at the end of the annual capacity allocation process, settle only the delivery commitments.*

In today’s process, when the annual capacity allocation process is finished in mid-September and the contracts between Trafikverket and the RU:s are being written, the train plan has traditionally been viewed as part of the contract. The train plan is a very detailed document, and it has been interpreted as a production plan by the RU:s and Trafikverket alike.
Refraining from fixating the production specifics at the end of the train plan process enables the production plan to be re-optimized at any time during the ad hoc process.

This results in the following advantages:

a. The yearly train plan can be divided into daily plans which can be optimized separately whenever a need to do so arises.
   - Every change to a particular day should improve the plan for that day, otherwise there is no point in making the change in question.

b. A final optimization of the plan for the upcoming day can be carried out before it is handed over to the dispatchers.
   - The ultimate goal is to hand over the most correct plan possible, where correct in this context means two things:
     1. The plan corresponds to the way in which the trains will in practice be dispatched in the event of no disruptions, and
     2. The plan is optimized with respect to its ability to fulfill the delivery commitments, even for the case when not all things go as planned.

   Note that point number two above in essence says that the plan should be robust.

c. The feedback loop between dispatchers and timetable constructors will be shortened.
   - Trafikverket will have the possibility to modify the production plan immediately after evaluating its actual performance quality during operations, which allows for much better feed-back loops and learning possibilities.

d. Track possessions (downtime required for maintenance) may cause changes to train paths as long as the delivery commitments are respected.
   - The chance of finding longer, continuous periods suitable for maintenance increases.

The train plan must not contain any resource conflicts in the current process, and this requirement does not change with the implementation of this stage. This stage gives Trafikverket the opportunity to optimize the plan, but also poses no requirements on them to do so. If the plan is never optimized, it simply has the same quality as in Stage 0. Stage 1 is thus backward compatible with the process of today in this regard.

Implementing this stage, Trafikverket abandon the rule that says that train paths cannot be modified after the train plan process has ended. During the ad hoc process, the short term planners will now be free to modify train paths as long as they still respect the delivery commitments. The production plan must be conflict free after each change to it, so that everyone in the process is sure that it is still a feasible plan. The difference from the previous approach is that every day can have a different timetable for the same train (as long as the same delivery commitments are respected).
Stage 2

Allow a train to have different train paths for different days already in the long term process, i.e., in the yearly plan

In the previous stage, once the train plan process (long term process) is over and the train plan has been constructed, the short term planners are free to optimize the production plan as long as all the delivery commitments are respected. Before any such optimization takes place, a train will initially, in the train plan, have the same train path every day it runs. This is an artifact of the manual procedures involved in constructing the train plan in the long term process. After Stage 1 has been implemented, it no longer makes sense to require that the long term planners construct a train plan where each train has the exact same train path every day it runs: That property will be violated as soon as different days are optimized separately by the short term planners anyway.

In Stage 2, the long term planners will be encouraged to allow trains to adjust to the different days they run already during the construction phase. In the most extreme case (point 1a, Stage 1), every day can be viewed as if it has its unique variants of all trains. That level of detail is not yet possible to manage in the long term process. The planners are at this stage not expected to treat every single day separately, but keep their eyes and minds open to the possibility of using capacity more efficiently by splitting some trains into two or three variants. When this stage has been implemented, a train will not automatically have the same train path every day it runs when the long term process is finished.

If efficiently handled, variants free up capacity in the train plan. This capacity is a hidden asset today since it is not immediately visible. Some of it is naturally exposed when the individual days in the train plan are singled out, and today capacity revealed in such a way is indeed used by the dispatchers, albeit in an ad hoc fashion. However, by identifying it already in the long term process, it can be used to increase the general overall ability to deliver, pave way for more efficient train paths, enable more efficiently planned track possessions, or to make room for more trains.

The notion of variants is not foreign to the process of today, but the administration around the procedure is rather stilted and every such variant becomes a separate train with a unique train number. Moreover, variant creation today is completely manual: Situations where capacity can be utilized more efficiently if variants are introduced are identified by ocular inspection, and there is also no special support for the actual creating of variants in the IT-tool the planners use (TrainPlan). The goal of Stage 2 is to develop the long term process so that splitting trains into many variants becomes a natural thing to do.

Allowing a train to have different train paths for different days in the train plan adds a challenge to the long term process that is not an issue in the short term process: In the short term process, the delivery commitments are already settled in contracts, whereas the goal of the long term process is to produce a plan from which the exact details of the delivery commitments can be extracted, and upon which contracts can be based. Luckily, the long term planner has the applications where details such as the location for the delivery commitment can be found, often associated with a time window representing the RU’s wish for the train to arrive and/or depart there. Thus, the task of the long term planner is to plan the variants of a train as if they were different trains, except for the requirement that they need to end up with the same delivery commitments.
This stage is backward compatible just like Stage 1 since the long term process still outputs a conflict free train plan and only delivery commitments are established in contracts. Just like before, if the opportunity to create variants is not used, the process is identical to the process of today in this regard, with every train expected to use the same identical train path every day it runs.

Stage 3

For train services that are marred by uncertainty, postpone their precise scheduling until closer to the real-time operations.

The process used to establish the delivery commitments is not changed with the implementation of Stage 3, apart from the fact that not all resource conflicts need to be resolved.

It is clearly not efficient to spend time – and infrastructure capacity – on resolving resource conflicts that will not occur. In some cases it is known already in the long term process that certain trains run the risk of being canceled even before the train plan takes effect. In other cases, the weight and/or length of a train might not be known long in advance, making (among other things) the running time of the train hard to predict. If it is possible to postpone the decisions about exactly where such a train should meet other trains, capacity can be used more efficiently than if a worst-case scenario is used.

The list of possible reasons for uncertainty mentioned in the previous paragraph is not exhaustive. The main contribution of Stage 3 is to give Trafikverket the opportunity to deal with uncertainty by postponing conflict resolution. Trafikverket might for instance wait until 8-12 weeks before these trains are supposed to run before regulating the train meetings in which the trains are involved. Some or all of the uncertainty will be gone by then, and decisions can be based on facts rather than on guesses.

This stage in the implementation of Incremental Allocation and Incremental Planning interacts mainly with the freight train railway undertakings (freight train RU:s) and the planning of their engines and crews. However, also the passenger RU:s experience an increasing need to make changes in their yearly plans to accommodate variations. In the latter case, it is generally sufficient to make major changes every three months or so.

This step involves abandoning the requirement that the result of the long term planning process is a conflict free plan where all the meetings are regulated. One of the implications is that the delivery commitments will be based on a plan that is not conflict free.

This is the first step that is not compatible with the old long term process that has as its main goal to produce a conflict free plan. However, a plan that is conflict free in theory is in fact still not necessarily executable in practice, making the hard requirement of the old process questionable in the first place. The number of changes during the ad hoc process is very large, in particular for some categories of trains. Still, we propose that this step is implemented in a “soft” way, e.g. initially for very few trains and only for a limited number of days.

The long term goal with this stage is to level out the workload and to focus on value adding activities during the planning process. Today, a lot of energy is spent on planning in detail for a production that in the end will not be executed the way it is planned anyway. Also, the level of detail in the plan
makes it very difficult to accommodate for any changes requested by the RU:s during the ad hoc process today. The latter will have been remedied already with the implementation of Stage 1, when the focus changes to preserving delivery commitments rather than keeping the initial train plan with all its details intact. The workload is leveled out by changing the quality of the plan, and thereby also the effort needed to produce it, to match the available data in the applications.

When it is time to implement this stage, we expect that Stage 2 has given the organization sufficient experience to dare start planning without always regulating all meetings: Planning with alternative conflict resolution methods, such as adding extra running time supplements instead of regulating meetings in detail\(^1\), can be viewed as a generalization of planning with variants as illustrated by the example below.

**Example**

Two variants are used during the long term process to explore the limits of the delivery commitments. These two variants do not need to end up explicitly in the train plan. Instead, the finished plan at the end of the long term process can simply rely on the fact that both variants fit, and the explicit variants are replaced by for instance running time supplements that create the needed room for meetings and overtakings. Later, during the short term process, one of the two variants might be used for a particular day if the circumstances are similar to what they were in the long term process. Otherwise, if the situation has changed, the train might very well be planned differently, although of course still respecting the delivery commitments.

Please note that Stages 1-3 have not implied any major news for the RU:s. Just like after Stage 0, the response to their applications will be in the form of delivery commitments that Trafikverket promise to fulfill. Rather, Stages 1-3 affect the internal process at Trafikverket by removing the need to schedule all trains in detail in the long term process. That kind of planning is instead deferred to a later point in time, when the exact details (such as weight and length) for the trains are available. For some RU:s, mainly freight train RU:s, this difference will be noticeable, to others it will not.

This step is not backward compatible in the same way as previous steps, since the delivery commitments are being established without explicitly having scheduled all meetings and overtakings but instead rely on other ways of resolving the conflicts. Thus, there is no guarantee that it is possible to schedule all meetings and overtakings such that all delivery commitments are respected unless this was in fact ensured by using variants while planning, as described in Stage 2.

**Stage 4**

**Introduce a new planning method in the long term planning process.**

The aim is to introduce a planning method with which it will be possible to establish the delivery commitments and ensure that they can be fulfilled without having to plan the production at the same time.

\(^1\)Research shows that adding running time supplements to one or both trains involved in a conflict could be an alternative method for conflict resolution if it enables regulation of the meeting at some later stage in the planning process. More research is required to establish the size of the time supplements and the way in which time supplements work together if more than one train is involved in a particular resource conflict.
The type of planning method we propose is in the spirit of the ones typically used in the sales departments of various industries to establish when a product can be delivered. In industry at large, that sort of planning is based on various key performance indicators and the currently known capacity utilization of the “production plant” such as the general quantity of a product that can normally be produced per day at a plant, or an equivalent facility, e.g. the number of cinnamon rolls a baker can make per day, etc. Establishing the date and time of delivery to a customer in this way requires a lot less effort than making a full production plan, but is often accurate enough. The details are only worked out once the actual production is imminent, and/or the contract has been signed. The production plan itself is subject to continuous revision.

The main reason for introducing a new planning process is to avoid having to redo planning steps. Planning should be based on known data. Any efforts put into a planning step that later turns out to be useless (because circumstances changed) is wasted energy, and such activities are non-value adding. If less effort is required in the long term planning process, fewer people will have to be involved, which effectively also reduces the problem of trains being handled by many different planners during the process. More effort might be required later, closer to the production phase, but then the planning activities will at least be value adding to a greater extent since the data used for planning is less uncertain at the later stages.

Note that the production plan may not be public. It is a tool for the producers (Trafikverket) used to ensure good utilization of their resources while making sure that they can deliver on time to their customers. Further, planning and execution of maintenance of the plant and necessary equipment should be a natural part of the process.

The most effective planning method for this step is a topic for further research. It could be a combination of methods regulating meetings and adding time supplements, or some other not yet “discovered” method. Given that nearly all other industries use some sort of high level delivery planning without having to work out all the exact details of production, it is highly unlikely that no such method should be appropriate for Trafikverket (and other Infrastructure Managers).

**Step 5**

*Adjust the application process to the market and allow applications with different time frames: services announced long in advance alongside “on demand” services.*

The right processes should be in place by now and applications with different time frames can be taken care of in an efficient manner both during the planning phase and during the real-time operations. By “right processes in place” we mean that the right decision can be made by the right person at the right time.

This is the last stage, i.e., Trafikverket now have Incremental Allocation and Incremental Planning implemented in their organization. The process involves only value adding activities, meaning:

- Any work done is justified by someone else’s needs
- Work is evenly distributed over the work force and over the year
- Everyone in the information chain delivers information of sufficiently good quality to the next one in the chain, on time (this includes actors outside of Trafikverket, such as RU:s and
Since less time passes between planning and decisions made by e.g. the RU:s about what trains to run, the possibility to give feedback and react to feedback is increased. The process is based on selected parts of the Toyota Production System and thus bears many similarities to lean production.

Illustrating the potential
The following sections describe some of the tests that have been carried out to test and demonstrate the potential of the ideas behind Incremental Allocation and Incremental Planning. A lot of technical details have deliberately been left out in this document, and we refer the interested reader to one of our papers, “The Maraca – a tool for minimizing resource conflicts in a non-periodic railway timetable” (RailRome, 2011), for a more detailed account of the tool we have been using in the test, the Maraca (Marackasen in Swedish).

Before ST and SP were initiated at Trafikverket, i.e., the initiatives at Trafikverket that will implement the ideas outlined in this document, we had performed basic tests that tried the ideas of what is now called Stage 2 and 3. Once Trafikverket initiated their own, internal projects to implement our ideas on a broad scale, our focus turned towards the stages that we call Stage 0 and Stage 1 in this document to show how they might be implemented and thereby also paving way for the implementation of the later stages. Our recent research efforts have for this reason mainly been concerned with exploring the effects of these two stages, with the emphasis on Stage 1.

Stage 0
This stage effectively shifts the focus from following a plan to achieving a desired result: Planning is a means to an end, nothing more and nothing less. The goal for the planning activities should be to ensure the fulfillment of the contracts. The contracts mainly concern departures and arrivals with commercial significance to the RU, but may also include other quantifiable characteristics such as keeping running times within certain limits.

An important advantage of this shift of focus is that the discussions between Trafikverket and the RU:s will focus on the product (the train services), and not on the way in which the product will be produced.

The figures below illustrate the shift in focus, as described above. Event times marked with green circles in Figure 2 are the only ones of interest to the end customers: passengers, goods transport buyers, and railway contractors performing maintenance to the tracks. As long as these times are adhered to during the real-time operations, the details of the production plan is of minor importance to the RU:s.
Figure 1: The yearly plan

Figure 2: Commercial activities (delivery commitments) in the daily graph
Stage 1

Already at this stage, there will be benefits of adopting the new approach. The following sections summarize the motivation for the potential advantages of implementing Stage 1 that have been identified in the research projects at SICS.

Production optimization

TrainPlan, which is provided by Funkwerk IT, is the timetabling tool used at Trafikverket. The train plan, i.e. the yearly plan constructed by the long term planners, is thus represented in TrainPlan and can be visualized as time-distance graphs.

At Trafikverket they use the so called the daily graph ("daglig graf") closer to the real-time operations. The daily graph is a graph representing the plan for the individual day. It is based on the yearly plan in TrainPlan, i.e. the train plan, but unlike the yearly plan, the daily graph is up-to-date with respect to cancelled trains and newly added train paths. We have evaluated the possibility to optimize the daily graph and we present the results here.

The test was performed on all traffic north of the station Bräcke up to Narvik (not including the tracks towards Östersund). Since we were unable to obtain data from the short-term process, we used individual days from the yearly train plan instead. We believe that optimizing the daily graph on real data from the short-term process would potentially have given us even better results since cancelled trains would have created “holes” that the optimization could have found use for.

We forced our optimization prototype tool Marackasen to respect the event times for all arrivals and departures involving commercial activities. By respecting the event time for an arrival, we mean that

---

2 This would especially be true for newly canceled trains, since freed-up capacity from early cancellations would likely already have been used for ad hoc applications where train paths are added on a first come, first served basis.
the train must arrive no later than a given time. In other words, arriving early is okay. Likewise, respecting the event time for a departure means not departing earlier than a given time, whereas departing later is fine.

Marackasen was also restricted to keep the train sequences on the links unchanged, meaning that no train meetings would be allowed to take place at any other location than the location in the original plan. Further, Marackasen keeps track of the number of trains simultaneously residing at a location, preventing locations from being utilized by too many trains at the same time (according to the given capacity, which is specified as a number). Under these circumstances, the running times for all trains were minimized for three random days, and also for the yearly plan as a comparison. Note also that both the input and the output were conflict free timetables, and the aim of the experiment was to explore the potential in optimizing every day separately.

In Figure 4 below, the effects of using Marackasen are displayed. Train meetings involving a train that does not run this particular day are removed, and the time from those removed meetings is “collected”. This time is either moved to the next meeting or it is removed altogether and the train arrives early to its final destination. In principle, the optimization straightens the train paths for trains that have some meetings other days but not this day.

![Figure 4: Straightening out train paths](image)

The table below gives the amount of technical time (in seconds) in the timetable after optimizing on a single day, day 200, and the whole train plan of 2010 (T10).
**Definition**

We define technical time as the time the train stands still and waits solely due to meetings and overtakings.

**Table 1: Technical time in the timetable before and after optimization for a random day and the yearly plan**

<table>
<thead>
<tr>
<th></th>
<th>Day 200</th>
<th>The yearly plan (T10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>280 420</td>
<td>89 218 787</td>
</tr>
<tr>
<td>After optimization</td>
<td>246 004</td>
<td>85 347 726</td>
</tr>
<tr>
<td>Difference</td>
<td>34 416</td>
<td>3 871 061</td>
</tr>
<tr>
<td>Difference, incl. fewer accelerations and decelerations</td>
<td>38 251</td>
<td>5 146 350</td>
</tr>
<tr>
<td>CPU time (laptop)</td>
<td>1.65 seconds</td>
<td>3.71 seconds</td>
</tr>
</tbody>
</table>

Just counting the technical time excludes a substantial gain in running times since this does not take into account the time that is no longer needed for acceleration and deceleration in relation to the removed stops. For this reason, we have added a row providing also this data.

There are probably many reasons as to why there is 4.3% to be gained in total running time by optimizing the yearly plan. First of all, the timetable planners are most likely not striving only to minimize the running time, but have other goals as well. Second, the infrastructure is divided into several areas with one timetable planner for each area, which means that the solution for the whole infrastructure will be suboptimal even if every separate area could be planned in an optimal way.

Luckily, we don’t need to know why the running times of the yearly plan is 4.3% from optimum to be able to say that the difference in percentage between the individual day and the whole year, approximately 12-4=8%, can be attributed entirely to the fact that the days in the yearly plan are different. Day 200 is of course only one of those days, but at least we can say that if we are allowed to optimize that day, the resulting plan is significantly more efficient in terms of running times than if we kept the initial plan, i.e., simply extracted day 200 from the yearly plan as it is.

Remember that we preserved the train sequences on the links for this optimization. As long as the locations for the meetings are the same as in the train plan, we can assume that a timetable constructor has deemed the train meetings possible on the given infrastructure locations. If we let Marackasen move meetings, we can no longer automatically assume that the solution represents a feasible timetable with respect to meeting locations since we have not given Marackasen enough data about e.g. track lengths at such locations.

Tests indicate that we would save approximately twice as much if Marackasen was allowed to change the meeting locations. While Marackasen can be configured to allow meetings to be moved, there are currently at least two reasons not to. In addition to no longer being able to guarantee that the choice of meeting locations is feasible the computations would also take significantly more time since we would be removing a restriction that cuts the solution space dramatically.
One way of tackling the feasibility problem is to let the system keep track of which locations allow for which combinations of trains to meet, and which locations that are not suited for trains with certain characteristics. Modeling this in an appropriate way so that a computer can take it into account is indeed possible, but requires a lot of effort and access to expert knowledge.

Note that the ultimate goal of optimizing every day separately might not be to minimize the running times of the trains. By straightening out the train paths, we obtain a plan that much more resembles the way the trains will actually be dispatched (unless there are delays, in which case the dispatchers will have to improvise anyway). By moving meetings to other locations, we may additionally increase the robustness to maximize the chances of a punctual performance.

**More efficient long-term planning**
In the summer of 2010, we performed a test of Marackasen to investigate how it could be used to facilitate the long-term planning process. During May and June, we received ten data snapshots from the ongoing timetabling process, containing all trains that belonged to one of the timetable constructor areas for that year. Every such snapshot constituted an example of the data the constructor was working with at that particular point in time.

Every constructor documents what trains he/she is currently working with, and what trains he/she has passed on to the next constructor. Thanks to this documentation, we could conclude which trains in the area were currently actively being scheduled by the constructor. The goal was to use Marackasen to prune the working material from resource conflicts and in this way help the constructor in his work.

We tried two essentially different parameter settings for Marackasen. In the first test, a similar restriction to that of the production plan optimization test was used (see the previous section). This means that Marackasen was not allowed to change the order of the trains on the links. Note, however, that two trains only have a defined train sequence order, or precedence, on a track link (or at a station) if they are not involved in a resource conflict there. For the second test, Marackasen was allowed to alter the established train sequences, thereby getting better opportunities to find alternative meeting locations.

Of course, Marackasen could remove more conflicts in the second test, but to a price: the CPU times for the optimization increased most significantly. Also, we could in the second test not be sure that Marackasen found feasible solutions, since the constructors rely on expert knowledge accumulated over many years when they plan meetings and overtakings. Marackasen cannot (yet) compete with that kind of expertise.

The conclusion was that Marackasen is useful and has the potential of making the long-term planning more efficient, but using Marackasen as a tool would most likely change the way the constructor works. Marackasen is quicker than the constructor when it comes to “cleaning up” the timetable from minor resource conflicts, but the constructor on the other hand is a better strategist. Exploiting this, the constructor should be the one to select good meeting locations and to place the trains in approximate time slots. Then he/she could use Marackasen to clean his/her “draft solution sketch” after which a new iteration based on the same principle could take place.
Figure 5 shows the gains from using Marackasen in this test, in terms of so called conflict seconds. In every resource conflict in the draft solution, the involved trains are in conflict with each other for a specific number of seconds, i.e., the (least) number of extra seconds that would be needed to resolve the conflict. Marackasen minimizes the sum of these conflict seconds.

![Figure 5](image-url)

**Figure 5**: The result of minimizing the number of conflict seconds in ten test cases

The “gain” in the figure is the number of conflict seconds that Marackasen could remove for the ten respective test cases. The number of conflict seconds still left after the minimization is visualized by the lower part of the vertical bars.

The CPU time, in seconds, as a function of the number of trains in each snapshot is displayed in Figure 6 below. Note that the execution times for data set 2 and 3 were 0.45 and 0.47 respectively, and the number of trains 485 and 496, making these two points in Figure 6 coincide.

![Figure 6](image-url)

**Figure 6**: The CPU time for the ten minimization problem (the values for data set 2 and 3 were almost identical).
From Figure 5 we can tell that something must have happened between snapshot 4 and 5, making the number of conflicts rise suddenly, after which the number decreases again. The reason was that the constructor did not do anything between the dates for these two snapshots (he was on vacation). During this time, the number of trains “arriving” to his area (passed on to him by a constructor working in a neighboring region) increased significantly, naturally increasing the complexity and the number of conflicts dramatically.

We can also see that Marackasen removed fewer conflict seconds in snapshot 5 and 6 than in later snapshots. Our explanation for this is that Marackasen lacked the “tools” to do much before the constructor had made sufficiently many strategic moves to once again have a feasible strategy for the basic structure of the plan.

Note that the number of trains in consecutive snapshots never decreases, and that the largest number of conflict seconds could be removed from the last data snapshot (number 10). This too supports our earlier assumption that Marackasen works best when the appropriate strategic moves have been made by the constructor. The last snapshot was very close to the final timetable, and approximately two thirds of the conflict seconds could be removed.

The results of the second test, i.e. when Marackasen was allowed to move meeting locations, are given in Figure 7:

Figure 7: The number of conflict seconds after minimization with variable meeting locations

Figure 7 shows that Marackasen could remove more conflict seconds when allowed to move meeting locations. For all ten snapshots, the gain in percent was larger than in the first test. The CPU times for finding the optimal solution (i.e., minimizing the number of conflict seconds) with Marackasen increased noticeably, and they were much less predictable. See Figure 8 below, where the execution time and number of trains were once again more or less identical for snapshot 2 and 3.
In the first test, the CPU time was more or less a function in the number of trains, whereas in the second test, the CPU time tended to depend more on how “hard” the problem was (which in turn depends on the number of conflicts and the level of congestion), which is something we cannot know in advance.

Last but not least, note that Marackasen was able to do a better job in test 2 when allowed to move meeting locations. This was especially noticeable in snapshot five and onwards.

**More efficient ad hoc train paths**

The process will be better equipped for making room for ad hoc train paths once the distinction between what (the delivery commitments) and how (the production specifics) has been fully acknowledged at Trafikverket.

The following example is based on an application in the ad hoc process from an RU in 2009. Due to technical reasons we had to modify it somewhat since no straightforward way of exporting data from TrainPlan existed in 2009. We used data from the train plan without any of the changes (additions and cancellations) that might have been made since it was published up till the date for this particular ad hoc train path application. Also due to the lack of data export possibilities, no data about track possessions were available to us. However, the example still illustrates an important point.

During the ad hoc process of 2009, SJ (the largest passenger RU in Sweden) submitted an application for three trains for the period from August 24 to December 12 of 2009. Train 7021 was meant to run between the cities of Boden and Umeå, Monday through Friday, departing at 6:14 a.m. from Boden and arriving in Umeå at 9:40. Trains 7022 and 7023 would depart from Umeå back to Boden in the afternoon.

We focused on 7021 but we refer to it as train 100021 to emphasize that the example is based on a real ad hoc application, but that the setting is not completely authentic due to the problems of missing data as discussed above. Once the data for the train was created, we used Marackasen to find the optimal placement in the train plan.

The line pointed to by the arrow in Figure 9 shows the train path without any conflict resolution at all, with small triangles showing where commercial stops were meant to take place. If the train was
not in conflict with any other trains, its arrival time could be as early as right after 9 a.m., as shown in
the distance-time graph of Figure 9.

Figure 9: The minimal running time for train 7021, ignoring conflicts with other trains

The prevalent process at the time stipulated that all conflicts must be resolved, and that ad hoc
trains only may use remaining capacity (no existing train path may be changed in any way). The result
after conflict resolution is shown in Figure 10.

Figure 10: The train path for 7021 when observing all current rules for conflict resolution
To fit in with the rest of the traffic on every day of the year, the train would be arriving at 10:20 a.m., an hour later than if it would have been allowed to run undisturbed.

Introducing variants of train 100021 provided us with a slightly better solution. After establishing what other trains 100021 interfered with, two variants were introduced: one for Monday through Thursday, and one for Friday. We required both variants to have as short running times as possible, but we also wanted them to have as similar arrival times as possible in Umeå. Figure 11 shows the result of optimizing the problem with Marackasen.

Figure 11: Two variants of 7021, still observing all current rules of conflict resolution

As Figure 11 shows, The Monday-Thursday variant of 100021 would arrive at 10:05. On Friday, that variant would arrive at 10:20, just like before the train was split into two variants. If the RU (SJ) had been presented with this solution and had accepted it, they would probably have chosen to announce this as two different trains to the customer.

The optimization just mentioned abided by the current rules of ad hoc train path planning. When we instead planned as if Stage 1 of Incremental Allocation were already implemented, i.e., we planned as if we were free to move trains in time at all locations except for the ones explicitly regulated in the contracts, we got a significantly better result. Figure 12 shows the result after using Marackasen to find a train path for 100021 under the circumstances just mentioned.
The train path for 7021 when only agreements stipulated in contracts need to be respected

100021 would arrive in Umeå at around 9:30 a.m., 50 minutes earlier than when we plan according to the rules of the current process. The difference is that other trains have moved slightly to make room for 100021, but without arriving too late at (or departing too early from) the locations where arrivals and departures are important (those locations are marked with triangles in the time-distance graphs).

During real-time operations, the dispatchers try to maximize the chance of trains being on time to the locations where they have commercial activities. Several interviews have made clear to us that even if the current planning process is not yet focused on delivery commitments, the dispatchers still value locations with commercial activities higher than other locations. E.g., a train that meets another train at a specific location on most days, but not on this particular day, is naturally not asked to wait for the non-existing train for a meeting that will not take place. Instead, the dispatcher tells the driver to drive on. The result is that the train in question will arrive early to the next location.

To avoid obvious mismatches between the plan and the way the plan will be carried out by the dispatchers, a stop that facilitates a meeting should not be in the plan that the planners give to the dispatchers on the day on which only one of the trains in question runs. The planners know what train meetings will take place on a given day and this is an example of one of the things that we will require of the daily production plan, once the new process is in place.

All in all, this example suggests that there are great benefits of Incremental Allocation in the ad hoc process. Also, splitting trains into variants can improve the situation already in today’s process. Using optimizing techniques to optimize every day separately, as in the previous examples, is one way of exploiting the possibilities of having different variants of the same train, with one variant per traffic day.
With Stage 1 of Incremental Allocation implemented, the total running time of the train would have been around 3 hours and 15 minutes. In today’s ad hoc process the same train would have got a running time of 4 hours and 5 minutes. The difference is 50 minutes, and since the train was scheduled to run on 80 days, the total saving in running time with the new approach would have been 4000 minutes.

The real train 7021 ended up with a timetable that had not been completely conflict regulated for the period between October and December 13, after which the train had a properly planned train path (according to the current rules) since a new yearly timetable was introduced at that time. The follow-up showed that the punctuality was about the same before and after the timetable change. Keeping in mind that there were track possessions during the period, and also winter which usually makes the punctuality go down somewhat, this still seems to suggest that it might indeed be possible to run occasional trains that have not been conflict regulated in the same way as is required today, without an automatic punctuality drop. This is however a very small example and it is too early to draw any general conclusions from it.

**Concluding remarks**

This document provides a detailed description of how to implement the concepts of Incremental Allocation and Incremental Planning at Trafikverket, the Infrastructure Manager in Sweden.

The implementation is divided into five well-defined stages. The first stages are designed to add new features in such a way that it is possible for Trafikverket to adapt to them gradually, at their own pace, before moving on to the next stage. The idea is to change the mindset of the organization before any major changes to the processes are introduced. In that way, the organization is steered towards a full implementation in a controlled way.

Every change that is introduced is a step in the direction towards a full implementation of the concepts, and every change has both its challenges and its benefits. The second part of the document describes tests that highlight what Trafikverket can expect to gain by adopting the changes that are introduced already in the earliest stages, providing Trafikverket with extra motivation to start the transition.

A full implementation of Incremental Allocation and Incremental Planning will have to take time, but the sooner the organization is ready, the sooner they will enjoy the full benefits from it.