Describing and Evaluating Train Services on the Swiss Railway Network from a New Perspective

Venkatasubramanian Mahadevan, D-BAUG ETH Zürich

Conference paper STRC 2007

STRC | **7th Swiss Transport Research Conference**
Monte Verità / Ascona, September 12. – 14. 2007

Describing and Evaluating Train Services on the Swiss Railway Network from a New Perspective

Venkatasubramanian Mahadevan
D-BAUG ETH Zürich
Zürich

Phone: +41 41 727 21 35
Fax: +41 41 727 21 39
email: mahadevv@student.ethz.ch

September 2007

Abstract

The work described herein[1] is undertaken in the context of the Swiss Federal Railways (SBB) Project PULS 90, which aims to develop methodologies, decision support, and automated systems for improving the quality of services that are provided to railway customers. To this end, this paper presents a new approach for representing and evaluating the train services that are provided on the Swiss railway network. The core of this approach focuses on the concept of the Global Service Intention (GSI), which consists of all the service offers than SBB has made to its customers. In other words, the GSI is intended to be a representation of train services from which a detailed timetable can be derived from. Our representation of the GSI is based on a spatial-temporal graph structure in which each vertex represents a station and the vertices are subdivided into sets representing the temporal dimension. A train service is represented by an edge (or collection of edges) connecting vertices across temporal boundaries, with weights on each edge corresponding to the weighted importance of that particular service. This form of representation is beneficial for a variety of analyses including, but not limited to, timetable stability, network flow, and commodity flow analyses. The notion of stability in the context of service fulfillment versus simply delay measurement is also discussed and a prototype software demonstrator for constructing test GSIs and evaluating various service fulfillment scenarios (according to certain prescribed criteria) is shown.

Keywords

train service – spatial temporal graph – service fulfillment

[1] This work is part of the PhD dissertation [Mahadevan V. (2006)] ongoing by the author at ETH Zürich.

1. Introduction

Modern railways are confronted with the dual problems of increasing operating costs and competition from other forms of transport such as road and air travel. Thus, in order to gain a competitive advantage and preserve their share of the marketplace, railways must provide customers with a high level of service. Therein lies the fundamental problem – the term "high level of service" is a vague concept at best. Consider the example of connecting trains and minimizing delays to customers as a measurement of service level: does providing a high level of service entail ensuring that connecting trains always wait for late trains? Or does it entail ensuring that connecting trains never wait for late trains? Or perhaps it involves a combination of both strategies depending on the individual circumstances? Further to this example, consider the following scenario:

- Train A from Zürich to Zug carrying 500 passengers is running 10 minutes late i.e. it will reach Zug 10 minutes later than expected. All the passengers on this train require a connection to Arth-Goldau provided by train B at Zug.
- Train B from Zug to Arth-Goldau currently has 200 passengers on it and they are waiting to depart immediately.
- Train C provides the same connection to Arth-Goldau from Zug, but it departs 15 minutes later than train B.

The aforementioned scenario raises some interesting points. A strategy that could be adopted is as follows: ensure that the maximum number of passengers do not have to endure any further delays (Train A is already running 10 minutes late). In this case, train B will wait for train A since only 200 customers will be delayed instead of 500. However, consider the situation where the commercial value of train B leaving on time is far greater than the commercial value lost by further delaying the 500 passengers from train A e.g. if train B provides further critical connections at Arth-Goldau that have no substitute. It is still possible for train A passengers to use train C to reach their final destination with a 5 minute additional delay, but customers waiting for train B at Arth-Goldau could potentially be subject to a greater delay. These are, of course, just 2 possible scenarios out of a multitude of possibilities especially when you consider the complexity of even a small segment of the entire railway network. Table 1 summarizes the options available and the related consequences pertaining to the discussed example (platform change times are not considered).

The example described above serves to illustrate the fact that taking planning and operational decisions aimed at providing customers with a high level of service based on varying criteria is not a straightforward process – there are several variables involved in making an optimal decision such as passenger satisfaction, commercial considerations, and overall stability of the railway network. To this end, several important factors must be considered. Firstly, the importance of particular train services must be identified; secondly the impact on the overall customer service (and inherently the loss of associated value) when particular train services are cancelled must be considered; thirdly, it should be possible to evaluate whether proposed services (and subsequent changes to these services) are realizable with the given resource constraints (rolling stock and topology). Clearly, automated analysis and decision support tools are required for these tasks due to the complexity of a typical railway network.

Table 1 Example of train connections and their impact on passenger delays

Action	Consequence	Number of Passengers Delayed (excluding initial 10 minute delay of Train A)	Number of Minutes Delayed (excluding initial 10 minute delay of Train A)
Train B does not wait for Train A	Train A passengers take Train C	500 from Train A	5 minutes delay for Train A passengers
Train B waits for Train A	Train A passengers take Train B	200 from Train B	10 minutes delay for Train B passengers

The ability for a railway organization to analyze and make intelligent decisions at both the planning and operational levels in order to optimally serve customer needs is therefore an extremely valuable capability. Currently, there is certainly a need for decision support tools and methodologies to aid in this effort at SBB. Hence, the work presented herein has a strong motivation and applicability to real-world scenarios at SBB.

2. The Concept of the Global Service Intention

The notion of the Global Service Intention (GSI) is built upon several other concepts that need to be defined first. These are presented in the next section.

2.1 Basic Definitions of Service Intentions

1. **Train Service Intention**: describes a train service from a start station to an end station and consists of the following:
    - an unique identifier;
    - an importance value (currently, this is just a number that determines the importance of the particular service intention but realistically it should be a combination of various factors such as passenger loads and commercial importance for example);
    - a list of stations served by the service intention;
    - a minimum transition (duration) time for the service intention from start station to end station;
    - a list of connections that the service intention provides and at what stations the connections take place i.e. a list of other service intentions that this service intention provides a connection for and the stations at which these connections take place.
    - the timeframe (period within the hour) that the service intention originates at the start station and the timeframe where the service intention has stops at stations along the way.
2. **Local service intention (LSI)**: the set of service intentions corresponding to a particular station region. For example, the set of service intentions available from the station Luzern.
3. **Global service intention (GSI)**: the collection of all service intentions on the railway network.

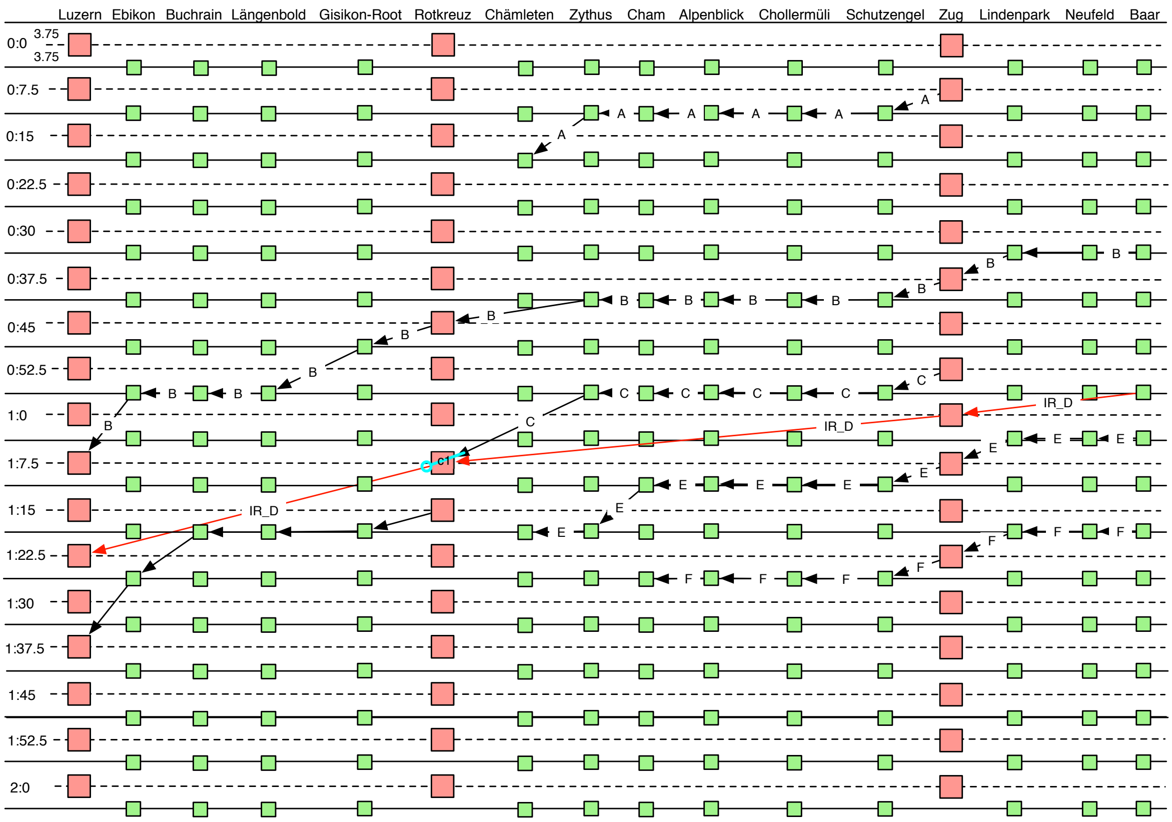
4. **Stability**: a timetable is considered stable when it is possible to absorb a specified set of degraded resources while continuing to fulfill the GSI.

5. **Iterative Control Algorithm (ICA)**: scheduler that can derive a concrete timetable based on the GSI. See [Caimi G. et al (2007)] for further information.

2.2 Global Service Intention in Detail

The data structure (model) used for the GSI is shown as an example in Figure 1.

Figure 1 GSI example



The data structure is a spatial-temporal graph with the following properties:

- The time dimension is represented as a sequence of temporal layers (see Figure 1). For the purposes of our initial study, an hour is broken up into layers of 7.5 minutes each (this is a parameter). It is reported by SBB that 7.5 minutes is the typical threshold at which a customer will tolerate waiting for a train, so it makes a good choice when designing a GSI as deficiencies in service intentions can quickly be exposed.

- Each station is represented as a set of nodes. One node per station exists in each of the temporal layers. Within each node is defined a list of connections of service intentions and a train change time for customers. A connection example is shown as a blue line labeled “c1” in Figure 1 connecting service intentions C and IR_D at station Rotkreuz.

- Stations are classified as either Main stations or Small stations:

  - Main station: a station that lies exactly on the center of the temporal layer (e.g. at 0 mins, 7.5 mins, 15 mins, etc.). These stations are considered to be the main system nodes of the railway network. The meaning of this is that if a service intention originates or stops at one these stations, it has a restriction that the scheduler (ICA) must ensure that in the final timetable the train must be at the station within a time equal to or within a certain threshold (we will call this threshold Δ which is 3.5 minutes in the given example of Figure 1) of the relevant center of the temporal layer. For example, in Figure 1 there is a service intention named “IR_D” going from Baar to Luzern. The service intention is at center 0 mins of the 9th temporal layer at node Zug and ends at node Luzern at center 22.5 mins of the 12th temporal layer. This means that the scheduler is restricted to producing a timetable entry for this service intention that is 0.0 mins +- Δ and 22.5 mins +- Δ at those particular stations. Of course, the transition time for the service intention must also be valid and enforced for this process to work. An example of a final timetable entry derived from this service intention could therefore be: Zug (01:02) → Luzern (01:25) with a transition time of 23 minutes.

  - Small station: a station that lies on the boundaries of the temporal layers. Unlike the main stations, there is no restriction on the times that the scheduler (ICA) must assign to trains starting or stopping at these stations in the timetable.

- Service intentions are represented as directed edges connecting nodes within or across temporal layers in the graph (depending on the transition time length). The edges store all the pertinent information about the service intention (a unique identifier, importance value, transition time) and in conjunction with the implied information obtained from the node structure of the GSI, other information that is defined is: the timeframe of the service intention (i.e. the temporal layers through which the service intention edge passes through) and the list of stations served.

The GSI is therefore a high level representation of a final timetable. This abstract representation can be used to derive a final timetable as well as provide useful information on the effects of disruptions and cancellations of train services on the customer as it essentially provides a complete picture of what services have been promised to the customer and in what timeframe those services must be delivered.

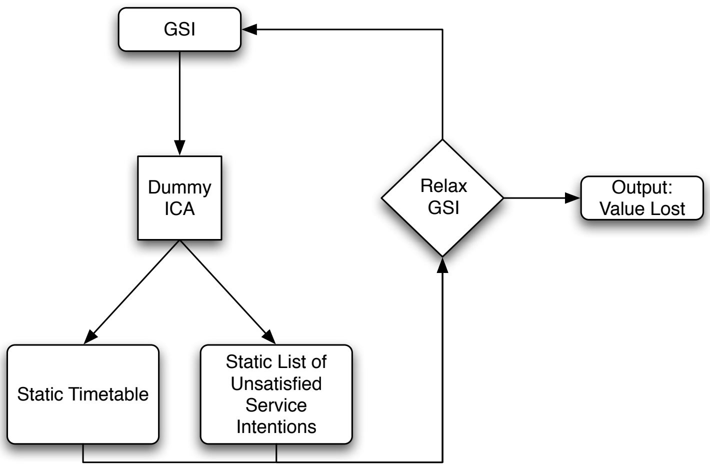
3. Approach

The unique structure and information content of the GSI lends itself to several different types of analyses, ranging from simple value calculations to reachability analysis models. The process loop shown in Figure 2 is used for our experimental approach. Due to the fact that the ICA at this point in time is still not implemented, a dummy ICA that simply outputs static predefined timetables and static predefined lists of unsatisfied service intentions (failure reports) corresponding to an input GSI is used instead. The following are the steps of the process loop:

- Input GSI is fed to the dummy ICA which outputs a static timetable and a static list of unsatisfied service intentions corresponding to the input GSI.
- The GSI is then "relaxed" if there are service intentions that cannot be satisfied by the produced timetable (in the real world this could be due to topology constraints, lack of rolling stock or other infrastructure, etc. but in this case we are just using static predefined timetables and static predefined lists of unsatisfied service intentions). In this case, certain service intentions must be removed from the input GSI in a process called GSI relaxation. An output of the relaxation process is a calculation of the value lost by removing some service intentions in order to obtain a timetable that satisfies all service intentions contained in the input GSI.
- The process is repeated until there are no more unsatisfied service intentions i.e. the GSI is relaxed enough such that a feasible timetable is "derived" with no unsatisfied service intentions.

Even this simple approach can be useful in quickly ascertaining what value is lost when a service intention is removed from the GSI. This validates the GSI concept perfectly since we now have a method of measuring the impact on customers when train services are compromised. For the moment, this is just a numerical value but it could be easily extended to include weight factors such as passenger loads, for example. The key problem, however, is deciding which service intentions must be removed from a GSI when it is relaxed. For this, 2 algorithms are proposed in the next section.

Figure 2 Process loop for experimental approach



3.1 Algorithms for GSI Relaxation

3.1.1 Value Calculations

This is the simplest analysis from which useful results can be derived. Recall that service intentions have an importance value associated with them. Hence an algorithm for GSI relaxation could function as follows within the process loop:

```
1. Read the failure report from the ICA and collect the list of service intentions
   that could not be fulfilled.

2. Remove the service intention with the lowest value from the GSI.

3. Repeat the process loop for generating a timetable with the newly relaxed GSI.

4. If a failure report is again produced by the ICA, go back to the first step and
   continue relaxing the GSI until a complete timetable based on the current GSI
   with no failure report is produced.
```

The above method is simply a greedy algorithm that cuts out the service intention with the lowest importance value from the GSI and repeats the process until a complete timetable based on the relaxed GSI is found. An advantage of this, albeit simplistic, algorithm is its ease of implementation and low computational overhead since the value calculations are very simple.

3.1.2 Reachability Analysis

Beyond simple value calculations, another useful analysis of the GSI is based on the notion of the transitive hull. The transitive hull of a directed graph (in this case the GSI) indicates if it is possible to go from station i to station j if there is a path between i and j. The following data structures are required:

```
m: adjacency matrix

m[i][j]: edge between i and j (1 indicates an edge, 0 indicates no edge) i.e. 1
indicates that there is a standalone service intention between station i and
station j; 0 indicates otherwise.

h: transitive hull matrix

h[i][j]: 1 iff j is reachable from i i.e. it is possible to find a path from
station j to station i. A path in this case could be a standalone service intention
or a collection of service intentions going from j to i.

n: number of rows / columns in m and h
```

The algorithm for computing the transitive hull is given below:

```
Initialize the transitive hull by:

1. setting the elements of the h matrix equal to the elements of the m  matrix i.e.
   initialize the transitive hull with the standalone service intentions.

2. setting the values of h[i][j] equal to 1 where i == j i.e. it is possible to
   reach node i from node j when both nodes are the same obviously.  Compute the
   transitive hull to determine reachability via collections of service intentions
   as follows:


    for (k=0; k < n; k++)

      for (i=0; i < n; i++)

        for (j=0; j < n; j++)

          h[i][j] = h[i][j] OR (h[i][k] AND h[k][j])
```

Using the transitive hull can provide us with a superior method of relaxing the GSI versus using simple value calculations. Recall that service intentions can possibly provide connections to other service intentions. So cutting a service intention from the GSI can potentially affect other service intentions as well. Therefore simply subtracting the value of a standalone service intention is not the most accurate measure of determining the value lost when the GSI is relaxed. The collection of service intentions and their associated values that are affected by a service intention cut should therefore be considered when relaxing the GSI. The transitive hull of the GSI captures these relationships; when a service intention is removed from the GSI, the transitive hull can be recomputed and compared with the previous transitive hull to ascertain the value lost across the collection of service intentions. However, in this case an importance value must be assigned to each connection as well so that it can be determined what value is lost explicitly because of the connection loss in addition to the value loss caused by removing the standalone service intention.

In addition to the 2 computational methods outlined above, it is certainly possible to also specify a static set of rules for GSI compromises based on predefined criteria. For example, one could specify that a certain set of service intentions must never be removed from the GSI in a static rule set or have a linear ordering of service intentions to be relaxed. While these may certainly be useful methods, they are not very interesting and hence will not be discussed further.

3.2 Prototype Implementation

A prototype GSI demonstrator has been developed that illustrates some of the concepts expounded upon in the earlier sections. Namely, the prototype attempts to complete a simplified cycle through the process loop shown earlier based on a simple use-case example.

Creation of the test GSI for use in the prototype is achieved by "reverse-engineering" an existing timetable from the Official Swiss Timetable [Official Swiss Timetable (2006)]. This is necessitated by the fact that the GSI does not yet exist in any useable form at SBB. A test timetable chosen was the Zurich-Zug-Luzern timetable (pages 803 – 829 from [Official Swiss Timetable (2006)]). See Figure 3 for a sample of the timetable structure and content.

The extraction of a GSI from the timetable data proceeds according to the following steps:

- Extraction, parsing of data, and conversion into XML format. The column highlighted in green in Figure 4 represents a service intention in the timetable. A timetable entry therefore consists of:
    - a train id (e.g. S1 22115).
    - a train type (S-Bahn, IR, or EC).
    - tuples of (station, arrival time, departure time) for each train (e.g. (Zug, 05:06, 05:06)).

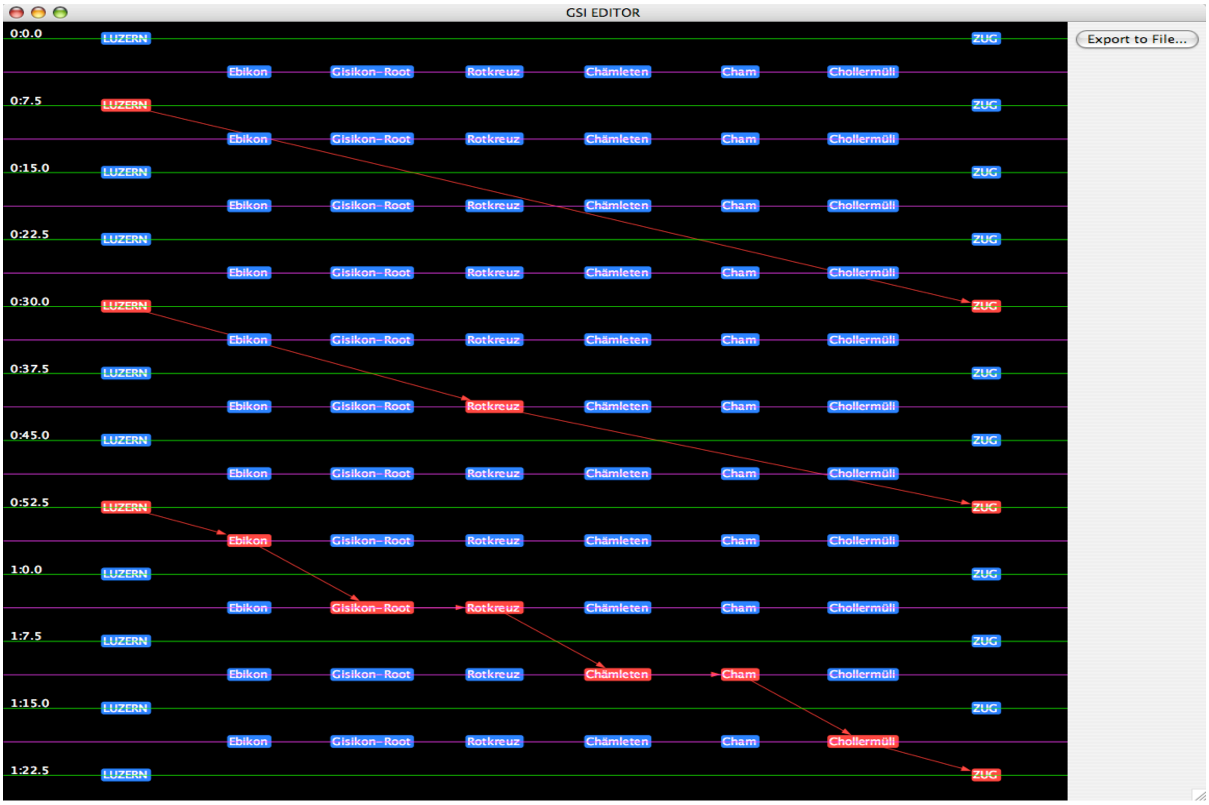
- Abstraction of the XML timetable file into a XML GSI file. Each column represented in the timetable (Figure 3) corresponds to a service intention. The XML timetable file is parsed and abstracted into a GSI as follows:
  - since the train ids from the timetable are unique, they are also used as the service intention id for each service intention.
  - importance values for the service intentions are not found in the timetable data; for testing purposes these are assigned at random.
  - an ordered list of station stops from start to end is created for each service intention.
  - a transition time from start station to end station is computed and assigned to each service intention.
  - a list of service intention connections organized by service intention id and station name. These are tuples of (service intention id that connection is provided for, station where connection is provided).

Figure 3 Timetable Sample Used as GSI Source Data

	S1 22115	S1 21115	S1 21117	IR 2303	S1 21117	S1 22119	S1 22123	IR 2305	S1 21119	S1 21123	IR 2307	S1 21123
Zürich HB 7				**535**				**604**			**635**	
Zürich Enge												
Thalwil				**543**				**613**			**643**	
Chur 900											**513**	
Thalwil											**637**	
Thalwil				**545**				**614**			**645**	
Baar		Ⓐ531		**557**		Ⓐ601	Ⓐ612		631	Ⓐ646	**657**	
Baar Neufeld						602	613			647		
Baar Lindenpark		533				604	615		633	649		
Zug 4		Ⓐ535		**601**		Ⓐ605	616	**628**	635	Ⓐ651	**701**	
Zug	Ⓐ506	536	Ⓐ551	**602**		⚒606	617	**629**	636	⚒651	**702**	
Zug Schutzengel	507	537	552			607	618		637	652		
Zug Chollermüli	509	539	554			609	620		639	654		
Cham Alpenblick	511	540	555			611	622		640	655		
Cham	513	542	558			613	Ⓐ624		642	658		
Cham	513	543	558			615			643	658		
Hünenberg Zythus	514	544	559			616			644	659		
Hünenberg Chämleten	Ⓐ516					⚒618						
Rotkreuz 3		547	Ⓐ604	**609**					647	⚒704	**709**	
Rotkreuz		549	➡	**610**	Ⓐ614				649	➡	**710**	⚒714
Gisikon-Root		551			616				651			716
Längenbold		554			619				654			719
Buchrain		556			621				656			721
Ebikon		557			624				657			724
Luzern 17		608		**625**	Ⓐ635			**649**	708		**725**	⚒735

The end result of the reverse engineering process is a XML GSI file containing the above information. This file corresponds closely with the content of a GSI defined in section 2. The prototype implements the simple process loop of Figure 2 and the simple value calculation algorithm described in section 3. Figure 4 displays a screenshot of the prototype displaying a GSI.

Figure 4 Screen shot of Prototype Displaying a GSI



The GSI file can be visualized using the prototype and a run of the process loop initiated. Dummy static predefined timetables and a static predefined list of unfulfilled service intentions are created in advance prior to running the process loop and output of the GSI relaxation process (value lost) is logged to a file. The user can also click and delete service intention edges manually and the prototype will then calculate the value lost based on the importance of that service intention.

Recall that the definition of stability in the context of GSI fulfillment is “a timetable is considered stable when it is possible to absorb a specified set of degraded resources while continuing to fulfill the GSI". In this case, by following the process of GSI relaxation and looking at the value lost when service intentions have to be removed from the GSI, it is possible to get a feel of timetable stability in the context of service fulfillment. For example, if there is some problem with the track infrastructure and the GSI can still be fulfilled using the same timetable, the timetable can be considered stable. However, if a new timetable needs to be produced and if it does not satisfy the complete GSI, then it's stability can be called into question. The degree of stability is then determined by the number and importance of service intentions that can no longer be fulfilled. Certainly this presents a new way of looking at timetable stability compared to approaches that focus primarily on delay calculations such as in [Herrmann T. (2006)].

4. Conclusion and Future Work

This paper introduced the reader to the concept of the Global Service Intention (GSI) and how it can be utilized to describe train services on the Swiss railway network from the perspective of service fulfillment. A new definition of stability in the context of GSI fulfillment versus simply delay measurement was also introduced and a simple prototype demonstrator for experimenting with the GSI concept was described.

There is still a great deal of work that remains to be done. In the future, more advanced algorithms for GSI relaxation will be implemented that are more “intelligent” when relaxing the GSI in case a timetable cannot be derived from it. Various inputs such as topology and rolling stock resources will be used to aid in the decision of what is realizable and what is not realizable on the current infrastructure. In addition, instead of using static data to simulate the operation of the scheduler (ICA), a real implementation that dynamically generates a concrete timetable from the GSI will be used. Finally, a statistical analysis module that measures stability in the context of GSI fulfillment will be implemented.

5. References

Caimi G., Burkolter D., Herrmann T., Chudak F., Laumanns M. (2007) Design of a new railway scheduling model for dense services, *Proceedings of the 2nd International Seminar on Railway Operations Modelling and Analysis*, Hannover, Germany, March 28-30, 2007.

Herrmann T. (2006) Stability of Timetables and Train Routings through Station Regions, *PhD. dissertation*, The Swiss Federal Institute of Technology Zürich, 2006.

Mahadevan V. (2006) PULS 90 Dissertation Proposal, *PhD. dissertation proposal*, ETH Zürich, 2006.

Official Swiss Timetable (2006), http://www.fahrplanfelder.ch, version: 10 Dec 06 - 08 Dec 07.