Quality management in the operation planning process by means of harmonized modelling

by Prof. Dr.-Ing. Ekkehard Wendler
Aachen University of Technology (RWTH Aachen)

Introduction

The railway system with its manifold interdependences has to be based on an exact network-wide planning of all system aspects as

- resources (infrastructure, rolling stock, staff)
- requirements (customer demands, e.g. demand for a train path)
- quality parameters (punctuality, queueing times, environmental aspects)

The main subject of these planning processes are movements of vehicles (trains, locomotives or shunting movements) on an existing or future infrastructure. In this context, the train movement over space and time is called a train path.

For these planning processes, both the trains and the infrastructure must be represented by means of suitable models. So, operation models are widely used in all basic planning processes in railway management, i.e.

- long-term strategic network design
- design and capacity assessment of infrastructure
- train path allocation
- real-time operation control

Depending on the respective planning aim, different planning horizons and different levels of detail in modelling are necessary for the operation models mentioned above. For example, operation models supporting the network design focus on a planning horizon of several years or decades. The customer demands are to determine only in a rough pattern (e.g. trains per hour and line). On the other hand, operation models supporting the operation control process need a horizon of a few hours only. In the operation control process the requirements of the customers are precisely known.

In Europe the open access to the railway networks is pushed by the European Union. Among others, in Germany open access has quite a good success. In the sense of open access, a train path represents a customer relation between a train operation company (TOC) and an infrastructure manager (IM). Therefore all procedures and tools dealing with train paths have to model this relation. Moreover, in the train path allocation process this relation between TOC and IM has to be even actively supported by the model. In some cases, the TOCs run their own pre-scheduling tools supporting the TOC-internal coordination of the train paths before contacting the IM for the official train path allocation.

Today for most of the models a certain minimal consensus exists regarding the model elements used: The infrastructure is abstracted to (linear) railway lines \( s \) (sequence of track sections and switches). The train movement is described by a function of time \( s = s(t) \) and is represented in a \( s-t \)-diagram (fig. 1).
Fig. 1 Graphical timetable as $s$-$t$-diagram
Problem description

Nowadays, a large variety of computer tools exist for the four basic operation planning levels of the IMs as well as for pre-scheduling by the TOCs. These tools are not compatible between different IMs (a special problem in Europe). Even within a single railway company, the compatibility of the tools is not guaranteed! The tools use to use sometimes the same, sometimes similar, sometimes completely different basic data structures. In some cases the tools are connected by certain interfaces. The architecture of the tools is heterogeneous.

An integration of the various systems by means of strong interlinks is necessary for a continuous data flow and to prevent a loss of quality. Strong interlinks are precluded by the heterogeneity of the functional and the software architecture. The knowledge to interpret data structures is strongly connected to the algorithms to evaluate and manipulate these data. Much effort is necessary to transfer the information between the several tools via the mostly ASCII-based interfaces. This problem occurs not only for data, but also for functional modules. In most cases the functional modules are strongly fixed in their own tool. An use of data and modules in different functional environments seems to be possible from a theoretical point of view, only. In practical cases the transferability fails because the data structures operate sufficiently only in a similar functional environment. A real online interlink between different tools seems to be very difficult.

The lack of possible interlinks seems to be a very serious problem for the use of infrastructure data. The incompatibility leads to several data bases existing in parallel. Even if these parallel databases are conscientiously maintained, the databases will be inconsistent already after a short time. For example in infrastructure planning the description of the current situation might be based one infrastructure data not fitting the real track layout, although the dispatching department (the neighbour department!) has just updated their own infrastructure database. The other way round, one can imagine a case in which the department of strategic network planning suggests to upgrade a railway line and to change a certain detail somewhere in the network. The idea of upgrading the railway line is transferred to the infrastructure planning department, but some of the details of the idea get lost, although the strategic department has described all their ideas in their own planning tool. This may happen due to the incompatibility towards the data structures of the planning tools used by the strategic department and the infrastructure department!

Regardless of the good will of the planning engineers to exchange complete data between the different planning levels, a free data flow is obstructed or may be totally prevented if being confronted to different data structures and data formats. Thus, a consistent planning is impossible. Basic requirements of quality assurance can not be fulfilled. A noncompliance to specifications incurs expenses. But, what happens, if there is no stringent data specification from the start?

Present situation at Deutsche Bahn AG

There is a great number of tools for the different planning tasks also in Germany. At least in the train path allocation, a certain de facto standard of data structures and data format modelling the infrastructure and the train paths has been established.
Installing the RUT system (Rechnerunterstütztes Trassenmanagement - Computer based train path management), the Deutsche Bahn AG has decided to use the SPURPLAN model developed at the Aachen University of Technology to represent the infrastructure data. Within SPURPLAN, the switches, the signals, the halts, the changes in velocity and gradients and further characteristics of infrastructure are modelled by nodes of a directed graph. The edges of the graph are track connections between the several characteristics.

Routes of trains along the graph are represented by sequences of nodes and edges. In principle, the SPURPLAN is able to model a network of any size. At the moment, Deutsche Bahn AG is connecting their existing smaller SPURPLAN databases into a large network-wide database precisely representing all tracks and train routes. In the first step, this database is used only by the train path management with its computer based system RUT. A further use of the SPURPLAN idea by the dispatching centers is to be implemented. It is necessary to use this database also for further business processes inside and outside of the IM.

**Harmonized modelling**

To harmonize data structures between different business processes dealing with infrastructure and train path data some ideas will be sketched in the following.

Usually, railway lines are the basic data structure in present infrastructure databases. In this context, lines consist of sequences of tracks and switches. Usually, a line begins and ends at switches or at buffer-stops. The line-based modelling has the disadvantage that both passing tracks and branching points are only scantily represented or not modelled at all. From a theoretical point of view, for each track not included in a regular line an own auxiliary line must be created to represent passing tracks and branching points correctly. However, this seems to be to onerous. To skip these inconveniences it is suggested to use track-based models instead of line-based models.

Thesis 1: Track-based models should be preferred to line-based models. The track-based model knows two basic elements: the switch and the track section.

However, for a couple of tasks (e.g. printing official timetables) line-based information is necessary. This line-based information must be generated from the track-based model.

Let

\[ S = \{S_i\} \]

be the set of the lines used in the model and

\[ G = \{G_j\} \]

the set of the existing track sections.

Thesis 2: \( S \) and \( G \) form a functional relation \( f: G \rightarrow S \). The inversion is not valid. The line-based model cannot be used as principle of order of the track-based model. The line-based model can be used only as additional description of the track-based model.
Switches and track section form the basic elements of the track-based model. From a theoretical point of view the track-based model can be interpreted as a mathematical graph. The graph is to be supplemented by several attributes, like track lengths or stationings, velocities, gradients etc. There are two basic possibilities to attribute the graph. Either a node-attributed graph or an edge-attributed graph may be used.

There are two main advantages of the node-attributed graph compared to the edge-attributed graph. The first advantage is that most of the attributes (stationing of a signal, of a switch or of a stop etc.) are just represented by a node. Therefore, the graph is to be supplemented by further nodes as signals, approaching signals, train tail detectors etc. Moreover, line-based attributes are easily to be transformed into nodes. Instead of cutting tracks into sections of constant velocity, any change of velocity is described by a node. A tunnel, for instance, is described by a node of beginning and another node of end of the tunnel.

Of course, also punctiform attributes may be transformed to line-based attributes. However, doing so a highly redundant graph is generated because all types of attributes have to be repeated in all edges although an attribute does not change in every edge. A supplement of a new type of attributes (e.g. velocities for tilting trains, stationings of ETCS balises) forces to integrate this new attributes in every edge, regardless of whether that attribute varies. So, the second advantage of a node-attributed graph is its non-redundancy.

Thesis 3: A node-attributed graph should be used in a harmonized infrastructure data model. Edge-attributed graphs lead to highly redundant structures with respective additional effort in data handling and problems in consistence and conversion.

For several tasks there is the necessity to assess capacity parameters and to identify bottlenecks in the railway network. Such tasks occur in the long-term network planning, in the infrastructure planning but recently also in the train path allocation process. The law of the EU dealing with the train path allocation requires that, under certain circumstances, the IM has to identify and declare congested infrastructure. When infrastructure has been declared as congested, the IM shall carry out a capacity analysis. In terms of capacity, there is a large number of different terms, e.g. railway line capacity, junction capacity, network capacity. In part, these terms are not well defined. However, these diffuse terms of capacity must be reduced to basic terms. The next thesis suggests how to decompose a large network into basic terms.

Thesis 4: Single-server queueing systems (SSQS) and set of (parallel) tracks are the smallest elements of a railway network (basic elements) for which the definition of capacity makes sense (fig. 2).
A SSQS is a set (as large as possible) of neighbouring infrastructure elements (switches and track sections) in which only one train movement may take place at a time. In terms of the queueing theory, the SSQS consists of a single server and a fictitious infinite waiting room. A set of (parallel) tracks is modelled as a \( n \)-server queueing system (\( n \) … number of tracks in the set). The capacity of even smaller elements of the network (e.g. for a single switch or a single block section) is as high as the capacity of the SSQS the smaller element belongs to. A capacity assessment for smaller elements than SSQS or set of tracks leads to redundant results. Parts of the networks larger than SSQS or set of tracks have to be assessed by decomposing them into the basic elements (SSQS and set of tracks) and by chaining these basic elements.\(^5\),\(^6\).

Beside these basic requirements according to the modelling of infrastructure there are also requirements to model the train movements.

**Thesis 5:** The occupation of an infrastructure element by a train has to be represented by a blocking time. Beside the physical occupation of an infrastructure element by a train, these blocking times include also switching times, a reaction time and an approaching time (fig. 3).
It is important to model blocking times not only from a safety point of view. This fact shall be explained by the example of a classical block section. There might be at maximum one train at a time in a block section. From a safety point of view, the blocking time consists of the physical occupation time of the first vehicle and the release time of the tail of the train. Afterwards the block section is completely released. Maybe, the times for switching the routes and signals are added.

For operation planning this approach is safe, but insufficient. To the parts of blocking time necessary for safety the reaction time and the approaching time have to be added. Especially, the approaching time is a very important component. The approaching time is the time the train takes to run from the approaching signal to the main signal. Only passing a green approaching signal (next main signal is also on green) ensures the train a ride without any obstruction.

If there is a continuous track to train transmission instead of track-side approaching signals an approaching time must be defined, too. In such a case, the place where the train would have to start braking if the next block section was still occupied is to be identified. The time the train
takes to run without braking (!) from this place to the next block corresponds to the approaching time for continuous track to train transmission.

Ergo, blocking times can be modelled for different signalling and train protection systems based on the same principle. Connecting all blocking times for a train running on a certain route leads to a blocking time step function, in the case of moving block to a blocking time ribbon\(^7\).

Trying to couple both models dealing with operationally defined blocking times and models with only safety based blocking times results in fatal planning mistakes.

In every level of operation planning the question of train succession has to be answered. The blocking time step function plays an important role to model the succession of trains in a correct manner.

Thesis 6: To ensure a train movement without obstruction a second train may follow a first train not earlier than the minimum headway time (fig. 4). The minimum headway time is determined by pushing together the two blocking time functions as dense as possible without any overlapping. To calculate the minimum headway time, it is always necessary to take into account the whole ride of the two concerning trains at least in the jointly used section.

Fig. 4 Blocking time functions and minimum headway time. Screen-shot of a RUT project.

An overlap of two blocking time functions corresponds to a planning conflict. Between two blocking time step functions a buffer time can be placed. Buffer times reduce the propagation of delays in the real time operation.
Especially in capacity assessment sometimes the mistake is made to consider only a track shorter than the jointly used section. This mistake leads to too short minimum headway times and to a too optimistic estimation of capacity. To assess for instance the capacity of a SSQS the minimum headway times must be based on the jointly used section of the two trains and must not be based on the section the SSQS includes only.

Thesis 7: In operation models it is necessary to distinguish between minimum headway times and buffer times.

In certain planning tools especially in train path allocation buffer times and minimum headway times are not distinguished. There is only a very general headway time including both minimum headway time and buffer time. In this case transferring data from the train path allocation process to the operation process leads to a loss of necessary information. In the real time operation just the buffer times are important instruments of the dispatchers. If train path managers aggregate buffer times and minimum headway times, how much buffer the dispatcher is allowed to deal with? That’s why already the train path manager has to make distinction between the buffer times and minimum headway times.

The minimum headway times also depend on the traffic profile of a railway line. Fig. 5 shows a railway line declared to concede priority for fast trains. Nevertheless, a slow freight trains uses the line, because there is capacity available (upper plot). If there is a higher demand for capacity from the fast trains, the slow train has to give way and is passed in a passing station. In this case, the minimum headway time is related to the passing section (section between two passing stations, lower plot)\(^8\).
Fig. 5 Minimum headway time for trains with low and high priority respectively.
Another case is shown in fig. 6. This line includes a special section with a harmonized speed profile for instance in the surrounding of a large junction. If capacity is available, a train can run faster if technically possible (upper plot). However, to calculate the minimum headway time the harmonized speed in the special section is taken into account, because in the case of heavy traffic all trains’ velocities are harmonized here. The train paths of the fast trains are bended (lower plot).

Fig. 6 Harmonized speed influences the minimum headway time.
Thesis 8: Capacity assessment always has to focus on heavy demand. Both categorisation (harmonized speed profiles) and dissociation (special priority rules) as discussed for instance in the German Netz 21 philosophy lead to different rules to calculate the minimum headway times.

**Summary and outlook**

Quality assurance in the railway planning and operation process is based on stringent and harmonized modelling principles. A couple of basic principles were presented above. The use of these principles is recommended for every planning level. There is only a small number of elements which are used in certain levels for special tasks only (tab. 1). For instance, the SSQS is mainly used in the network planning level and in the infrastructure planning level.

<table>
<thead>
<tr>
<th></th>
<th>track-based modelling of the infrastructure</th>
<th>non-redundant node-attributed graph</th>
<th>capacity assessment for SSQS and set of tracks</th>
<th>blocking time includes approaching time</th>
<th>minimum headway time and buffer time distinguished</th>
</tr>
</thead>
<tbody>
<tr>
<td>long-term network design</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>capacity assessment</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>train path allocation</td>
<td>√</td>
<td>√</td>
<td>(✓)</td>
<td>√</td>
<td>✓</td>
</tr>
<tr>
<td>real-time operation control</td>
<td>√</td>
<td>(✓)</td>
<td>√</td>
<td>√</td>
<td>✓</td>
</tr>
</tbody>
</table>

(✓) used only for special tasks

Tab. 1 Recommended use of modelling elements in the different planning levels

Tab. 2 shows the present situation for the DB Netz AG, the IM within the Deutsche Bahn group. As mentioned above, the DB is already using most of the necessary modelling principles in train path allocation. The node-attributed graph is to be inherited to the real-time operation control. Regarding the long-term network design and the capacity assessment the situation is inhomogeneous.
In the next few years some effort is necessary to push forward the harmonization of data structures and modelling principles. Furthermore, the theses stated should give some input for a discussion on this topic especially within Europe.

A step further to integrate the different systems may be a more general data broker architecture as discussed between Deutsche Bahn group and the academia at present. Such a broker system has to handle services, algorithms and data. The present tools are to be modularised. Their complexity can be reduced. New tools are able to use services of already existing tools (e.g. calculation of running time or decomposition of a network into SSQS). The broker is supplemented by an integrated data management system adjusting demands for data quality and topicality.

A main condition to be able to implement such a broker system for operation planning processes is a consensus on basic modelling principles as discussed in this paper.

References


