JESTECH 16(2), 39-51, (2013)

Engineering Science and Technology, an International Journal
(JESTECH)

journal homepage: jestech.karabuk.edu.tr

Methods of assessing railway infrastructure capacity

Thorsten Büker*

*VIA Consulting & Development GmbH, Aachen, Germany

A R T I C L E I N F O

Article history:
Received 27 November 2012
Accepted 20 April 2013

Keywords:
Railway capacity,
UIC Code 406,
Compression,
Concatenation,
Queueing theory,
Level of Service

A B S T R A C T

The following article starts by reviewing the areas of railway applications in which the concept of capacity has either always played a part or has begun to do so recently. It moves on to introduce the discipline of modelling the capacity consumed by train movements. Building on this, several methods of establishing the capacity of railway lines are explained. Augmenting the article is a summary of which of the methods dealt with can also be adopted for nodal areas. An outlook section sketches out how the various results arrived at could be factored into overarching indicators for network components.

1. Introduction

For the purpose of identifying what action is required, the numbers of trains forecast or targeted for a given planning horizon are compared and contrasted with the capacity of the railway infrastructure under review. Capacity is taken here to be the potential quantity of trains assuming an either explicitly or implicitly defined level of quality. There are a number of factors that exert an influence on capacity, amongst them:

- the mix of trains
- speed profiles
- the signalling system
- the automatic train protection system or automatic train control system
- train priorities (not necessarily the same under pathing management as in actual operation),
- the level of quality targeted (not necessarily the same under pathing management as in actual operation).

If a forecast or targeted quantity of trains cannot be coped with by an infrastructure, then there is a need to consider suitable infrastructural measures (modification of block sectioning, incorporation of additional passing opportunities, use of different designs of switch system, changing to a different signal/automatic train control system etc.) and to perform a fresh calculation of productive efficiency in an iterative planning process. It is possible as well, given the applicable degrees of freedom, to influence productive efficiency by varying the mix of trains, homogenising speeds or adjusting train priorities.

It is not only when railway infrastructure is expanded that capacity calculation plays a part but also when it is downsized. It may prove necessary in the event of demand falling off, traffic flows being transferred to other corridors or the mix of trains being altered to likewise effect adjustments to capacity on economic grounds. It is quite likely, specifically in the context of downsizing railway infrastructure, that the concept of capacity will also have been adopted in domestic legislation, as is the case in Paragraph 11 of Germany's General Railway Act [1].

In the course of liberalising railway networks, finally, the capacity concept was also incorporated into European railway law. Article 23 of Directive 2001/14/EC [4] defines "congested infrastructure" as being a section of infrastructure in which "it is not possible to satisfy requests for infrastructure capacity adequately". Attention is accordingly addressed towards the processes of pathing management and timetabling. Article 22 (3) of 2001/14/EC allows the infrastructure manager (IM) to adopt special priority criteria when awarding paths on a section of line designated as being congested as a means of dealing with the increased level of demand. This option of special priority arrangements has yet to be consistently transposed into domestic legislation however. Conversely, some Member States have

* Corresponding author. *E-mail address: th.bueker@via-con.de* (T. Büker).

made it obligatory in their own legislation for IMs to conduct capacity analyses of congested lines and submit “capacity enhancement plans”.

How capacity is determined also plays a key role for pathing management in a second respect: Article 17 of 2001/14/EC permits the IM and a “railway undertaking” (i.e. a train operating company or TOC) to conclude multi-annual framework contracts governing the use of railway infrastructure. This is designed to lessen the economic risks facing IMs and TOCs due to the high proportion of fixed costs accruing to both by means of longer-term contracts that tie up infrastructural capacity accordingly. One drawback of large-scale long-term tie-ups of this kind, though, is that TOCs who have not signed a long-term utilisation contract may run the risk of being completely barred from using specific rail infrastructure or else they will have difficulty entering the market. To remedy this, Germany’s Railway Infrastructure Usage Regulations (EIBV) [5], for instance, flesh out the provisions of European law by stating that capacity awarded on the basis of framework contracts must not exceed a threshold of 75 % of total capacity.

2. Modelling The Capacity Consumed by a Train Movement

Any meaningful method of assessing capacity is crucially predicated upon there being a suitable capacity consumption model. This model needs to establish a link between infrastructure utilisation and service quality. This first involves modelling capacity consumption for a single train path.

The key question to be answered with regard to capacity consumption concerns describing and quantifying the capacity consumption for a single train move (“path”). Studying the interaction between individual scheduled paths allows the behaviour of the infrastructure elements in place to be computed. In Germany use has for many years been made of the blocking-time model devised by HAPPEL in 1959 at Aachen to model capacity consumption [10]. Since a software program for computer-aided pathing management was introduced in 1998, this model has also been used to compile timetables in Germany. The UIC’s “Capacity” Leaflet 406 [19] also recommends adopting the model for capacity studies furthermore.

The main idea underpinning a “stepped blocking-time series” is that a train movement actually occupies a block section demarcated by two stop signals for longer than the physical process of occupying the track (Figure 1). Consideration needs to be given to route formation time, sighting time and approaching time for the move between the distant and stop signals before the front of the train reaches section signal A. Once section signal B has been passed, it is necessary to add the clearing time (rear of train passes end-of-train detector) and the route release time to the blocking time before the “train out of section” message is issued. The sum of these blocking-time elements is referred to as the blocking time and indicates the capacity consumed by a train movement.

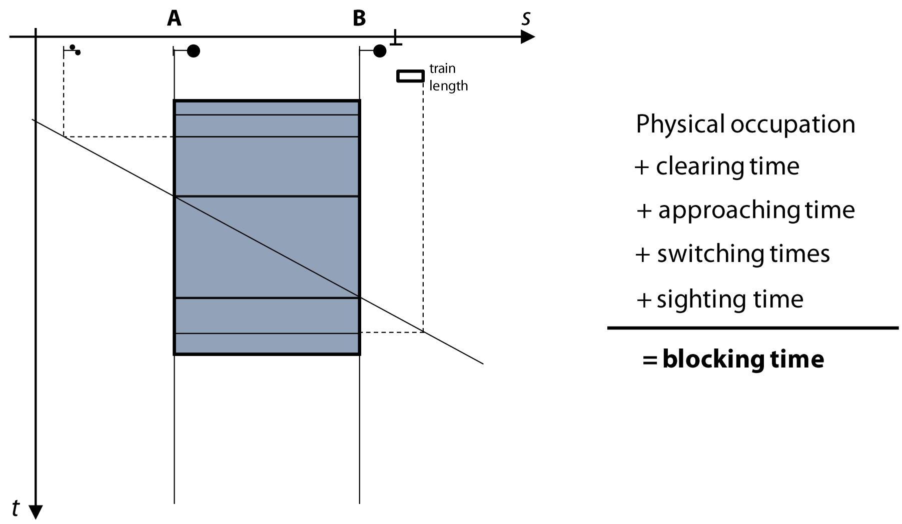

Figure 1. Constituents of the blocking time in a time-distance graph

The point constituting the beginning of the blocking time is that by which the driver must have been notified of a stop signal being pulled off. Under conventional distant/stop signal systems, this is the point in time at which the distant signal is detected, which does not necessarily coincide with the brake initiation point. It is usually arranged to occur well before the brake initiation point. If the driver has not been notified of the prevailing stop signal having changed to a “proceed” aspect by the time this point in time arrives, he or she is required to assume the stop signal is at danger and to start applying the brakes. This generally leads to a deviation from the scheduled (non-conflicting) time-distance curve.

It is always assumed where blocking-time "boxes" or elements are concerned that the time-distance curve underpinning them can be performed without hindrance, since approaching a stop signal at danger is ruled out both in running-slot planning and in productive efficiency calculations. The assumption of a move free of hindrances draws on the fact that receiving restrictive information at the distant signal leads to more capacity being consumed in most cases, since it is necessary to brake and pick up speed again.

An exception concerns blocking-time boxes in the context of a scheduled stop. There may be a case here for the stop to be approached with the signal in advance, frequently the platform starting signal, still at danger. This prevents the following block section from being occupied too early. It may, however, also result in restrictive supervision of train running influencing the speed profile into the stop.

Capacity consumption has only ever been regarded on a solitary basis per track-occupation element (e.g. block section, set of points, stopping point) hitherto. Widening our approach to embrace a sequence of such track-occupation elements yields a segment of infrastructure without any means of altering the order in which trains run, which is known as the passing section. The length of the passing section may vary depending on the service priorities of the trains involved, since - although a change of sequence would be technically feasible - it is not possible to posit a potential passing event for all combinations of trains at every passing station. Pushing the stepped blocking-time series for two trains on the passing section up close to each other allows the minimum headway time to be ascertained at the beginning of the section (Figure 2).

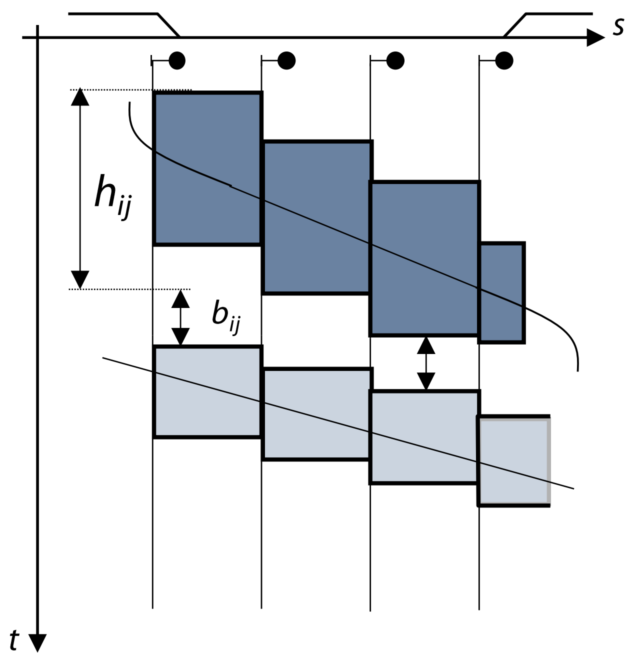

Figure 2. Stepped blocking-time series

The capacity consumed on a line is accordingly always a function of (at least) two trains i and j and is referred to in what follows as minimum headway time h_{ij}. Where a buffer time b_{ij} is inserted between two stepped blocking-time series in the timetable, it has the task of attenuating the delay transmitted from train i to train j should train i be running behind schedule. The ruggedness of this approach becomes clear once it is realised that minimum headway time h_{ij} compactly pools all the relevant properties of the infrastructure, train, automatic train control system and any service priorities (via the passing section).

The blocking-time model, which, as set out above, relates to the conventional distant/stop signal system, can be adapted to any signalling and train control system, notably including the European Train Control System (ETCS) [3]. It is also possible by forsaking the philosophy of hindrance-free running to investigate infill components by assessing knock-on delays [20].

3. Methods of Establishing Line Capacity

Set out below are three of the most common methods of establishing line capacity together with a rundown of their distinguishing characteristics. All of the methods presented involve describing capacity consumption in terms of stepped blocking-time series. An at least mesoscopic representation of the railway infrastructure is required as a means of establishing stepped blocking-time series, though the form adopted should ideally be microscopic. Reference is made to [11] for a discussion of the depth of detail required when modelling infrastructure.

It should also be pointed out at this juncture that neither time/event-synchronous nor asynchronous simulation exercises constitute methods of establishing capacity. Rather, these serve to validate the compatibility of infrastructure and specific timetables. Moreover, the time required to run a simulation series (at the microscopic level) is too high in practical applications to produce meaningful information on capacity.

3.1. Unscheduled waiting times

The standard means of calculating the productive efficiency of railway lines in Germany is the STRELE tool. The algorithm adopted in STRELE is derived from preliminary theoretical studies on measuring buffer times in railway timetables conducted by Schwanhäußer at RWTH Aachen [17], who himself drew on the aforementioned preliminary studies by Happel.

If various train types travel on line *n*, then it is possible to specify n^2 minimum headway times and express them as matrix *H*. If trains travel at differing speeds on a line, the resultant effects are factored into the matrix of minimum headway times. Hence, longer minimum headway times arise for "slow-fast" sequences of trains and shorter minimum headway times for "fast-slow" sequences of trains. Comparatively long minimum headway times arise if a sequence-of-trains scenario involves trains running in opposite directions.

If the quantity of trains of a specified type n_i is known, then it is possible on that basis to specify the probability of a specified train type p_i appearing within the overall train spectrum of n_{Trains} trains within a period under review $t_{Interval}$:

$$p_i = {}^{n_i}/_{n_{Trains}} \quad (1)$$

Where capacity studies are conducted for infrastructure planning purposes, values n, n_i and n_{Trains} are arrived at on the basis of the forecast service schedule. In the absence of detailed knowledge of the timetable for medium to long-term planning horizons, a random mixture of train types is frequently assumed. This is then referred to as being a non-timetable dependent means of establishing a line's productive efficiency.

A period of one day is generally selected for review, in which case $t_{Interval}$ = 1,440 min. If the service schedule has a markedly different pattern at day than at night, then it is possible to user shorter sub-periods as a basis. Periods under review $t_{Interval}$ < 4 h ought to be avoided, however, since it is then no longer possible to arrive at dependable productive efficiency values.

It is possible with the aid of probabilities p_i to compute an average minimum headway time h_{Avg}:

$$h_{Avg} = \sum_{i=1}^{n} \sum_{j=1}^{n} p_i \cdot p_j \cdot h_{ij} \quad (2)$$

This average minimum headway time correlates as follows with the average buffer time b_{Avg} provided for in a timetable configuration:

$$h_{Avg} + b_{Avg} = {}^{t_{Interval}}/_{n_{Trains}} \quad (3)$$

Using the theoretical scenario of a timetable with no buffer times allows the theoretical productive efficiency of line $n_{Trains,max}$ to be specified:

$$n_{Trains,max} = {}^{t_{Interval}}/_{h_{Avg}} \quad (4)$$

This number of trains cannot be run in practice, though, since a timetable of this kind compiled without resort to any buffer times at all would give rise to a theoretically infinitely poor quality of service. Even the slightest disruption would be passed on to the next train and then, like falling dominos, to all remaining trains.

Consideration is instead given to a line's practical capacity $n_{Trains,opt}$, which is also referred to as its nominal productive efficiency. Knowledge of the requisite average buffer time $b_{Avg,req}$, which serves to ensure the timetable's market-responsive stability, is needed to determine this:

$$n_{Trains,opt} = {}^{t_{Interval}}/_{(h_{Avg} + b_{Avg,req})} \quad (5)$$

Specification of practical capacity is thus dependent upon a given "level of service". It is worthwhile as a means of grasping the nature of such a level of service to analyse in greater depth the mechanism of delay transmission on which Schwanhäußer conducted research. As can be seen in Figure 2

Figure 2, train *i* requires a delay v_i greater than the buffer time at the beginning of the passing section before it passes on any delay to train *j*. If v_i is less than the buffer time, no delay is transmitted. Hence, as set out above, buffer time b_{ij} lowers the likelihood of knock-on delays arising. If, by contrast, delay v_i is greater than the sum of buffer time and minimum headway time, then a change of sequence occurs at the commencement of the section of route the two trains share. The equation to establish whether knock-on delay k_j arises for train *j* is thus as follows:

$$b_{ij} < v_i \leq b_{ij} + h_{ij} \tag{6}$$

If, by contrast, the delay is greater than the sum of buffer time and minimum headway time, the knock-on delay is suffered not by train *j* but by train *i*, since a change of sequence occurs. This only causes a knock-on delay k_i, if the delay is smaller than $b_{ij} + h_{ij} + h_{ji}$, however. If the initial delay is even greater, then no knock-on delay occurs under this two-train model. The equations shown here assume the two trains are of equal status. (To simplify matters, no consideration is given to the original delay v_j suffered by train $_j$ in advance.) To summarise, the knock-on delay suffered by the second train *j* is to be determined thus:

$$k_j = \begin{cases} v_i - b_{ij} & ; \text{if } b_{ij} < v_i \leq b_{ij} + h_{ij} \\ 0 & ; \text{else} \end{cases} \tag{7}$$

In addition to the aspects already referred to, it is necessary when computing knock-on delays to have account to the following conditions:

- service rankings of the trains concerned
- distribution of buffer times within the timetable, as this constitutes a random variable
- delay transmission not only to the next applicable train but also to the next train but one, the next train but two and so on ("knock-on delays of a higher scale"),
- delay transmission in lengthy queues, and
- the fact that delays upon entry into a section are not constant values but random variables.

Recapitulating the entire derivation would exceed the scope of the present article and the interested reader is accordingly referred to [17] and [21]. There is also a wide-ranging discussion of the two-train philosophy for adoption in models of delay propagation in [2]. Bearing all the boundary conditions listed above in mind, Schwanhäußer managed to formulate an equation for calculating average knock-on delay k_{Avg} for all trains. (It is not shown in full here.) The equation is frequently referred to as the "STRELE equation" in railway operations science owing to its having been adopted in the aforementioned STRELE software tool for computing the productive efficiency of railway lines.

Multiplying the average value of knock-on delays by the quantity of trains forecast in the period under review yields the sum of knock-on delays, which can be used as an indicator for describing service quality:

$$\sum \mathrm{E}K = n_{Trains} \cdot k_{Avg} \tag{8}$$

The function sequence relative to the quantity of trains is illustrated in Figure 3, in which, notably, a pole is to be discerned at the point of theoretical productive efficiency $n_{Trains,max}$. Two deliberations need to be pursued regarding the extremes in the function sequence:

- On the one hand, it is self-evident that no railway line can be run at this level of theoretical productive efficiency, since the quality of service would theoretically become infinitely poor. There would be extreme modal reactions on the transport market, meaning that customers would switch to other means of transport. TOCs' operating costs would theoretically know no limit, since the extreme instability of the timetable would necessitate keeping huge amounts of rolling stock and staff in reserve.
- On the other hand, it is also obvious that trains can never be run without any knock-on delays whatsoever. The railway system is constantly subject to (external) disruptions that may lead to delays upon entry into the area under review and those originating there. These in turn will, with a certain degree of probability, lead to knock-on delays within the timetable structure.

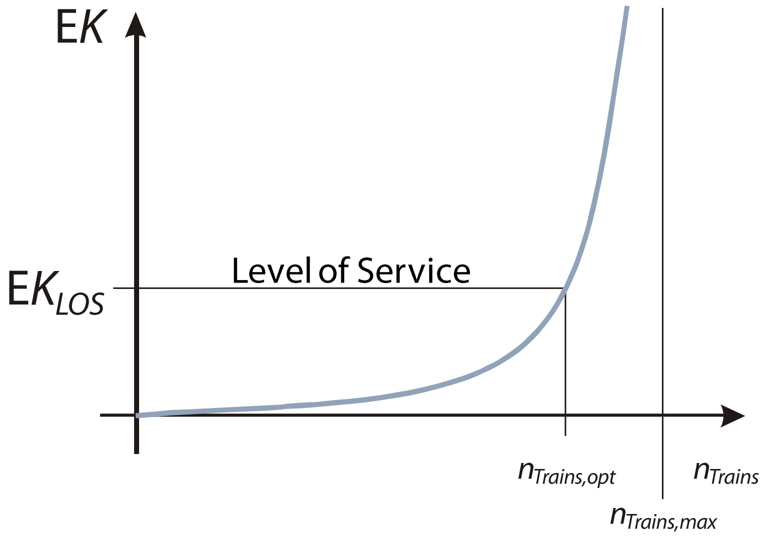

Figure 3. Average knock-on delay depending on the number of trains

As can also be seen in Figure 3, it would only be possible to produce a service with no delays at all if $n_{Trains} = 0$, which would, of course, also mean completely doing without revenues. Costs *C* and revenues *R* on a railway line are now represented in Figure 4 as a function of a number of trains n_{Trains} [18]. It is revealed that there is an optimum point $n_{Trains,opt}$ at which a maximum contribution margin, the differential between revenues and costs, is arrived at. It is possible with the aid of transport-science models to approximatively identify the point. This then serves as the aforementioned level of service, which allows the economically optimum quantity of trains to be established. (And implicitly also ensures a requisite average buffer time $b_{Avg,req}$.) Given that the transport-science models to be adopted are fraught with a degree of uncertainty, work is often carried out on an economically optimum segment in practice. Where the input-intensive process of working with transport-science models is dispensed with, a substitute functional correlation between the admissible sum of knock-on delays per day and the share of passenger trains in the service schedule has been derived from expert interviews amongst signalling staff. This functional correlation output the permissible average knock-on delay EK_{LOS}.

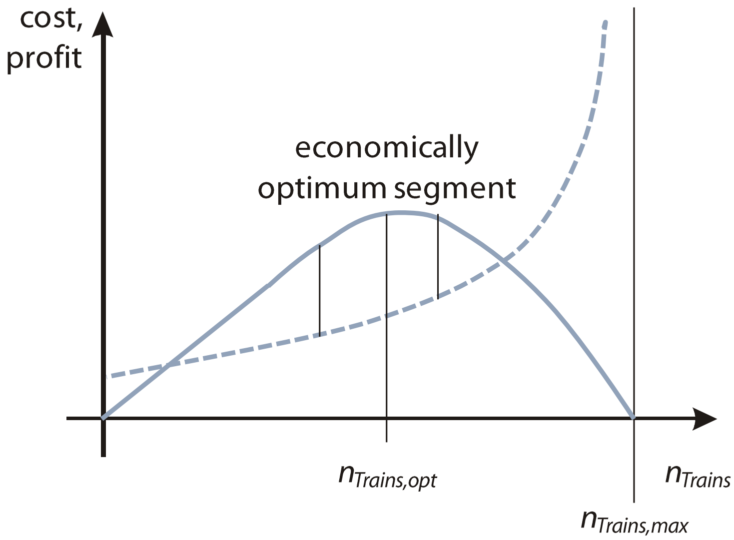

Figure 4. Economically optimum segment

The STRELE equation has now also been integrated into the new LUKS® software tool, whose German initials stand for “Performance Studies on Nodes and Lines” and which allows the productive efficiency of defined line items in a railway network to be computed [12]. Figure 5 depicts the definition of such a line item.

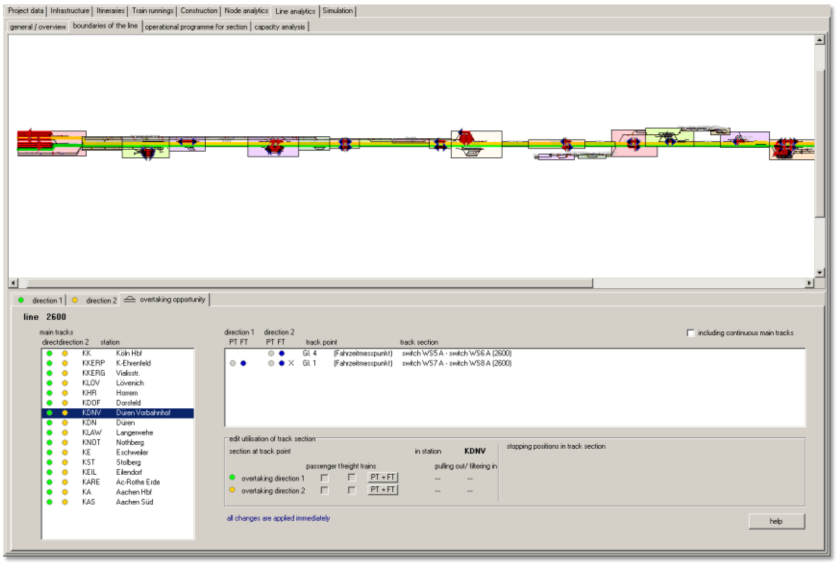

Figure 5. Definition of a line item in LUKS®

Achieving meaningful information on the capacity of a section of line involves making the area under review around this segment so big that it is bounded by stations at which trains of the second-highest service priority make scheduled stops. Otherwise, the minimum headway times for the section of line might be underestimated by assuming a shorter headway for the section of line than is actually feasible.

With a view to guaranteeing consistent infrastructure planning for long and medium-term horizons, the STRELE equation has also been incorporated into strategic network planning. A coarser infrastructure model is used in this context than for modelling in LUKS®. Minimum headway times are not computed with the aid of stepped blocking-time series here but are, instead, directly captured as matrices or else established with the aid of estimators [15].

3.2. Concatenation Pursuant To UIC Code 406 And Additions Thereto

UIC Code 406 ("Capacity") issued by the International Union of Railways in 2004 [19] provides a standardised description of capacity consumption on the basis of stepped blocking-time series as well as containing a method for determining the capacity of railway lines. Line capacity is nevertheless calculated to a simplified procedure. Under this procedure all stepped blocking-time series for a predefined timetable are pushed so chronologically close to each other within the period under review that adjacent stepped blocking-time series are contiguous but never overlap (Figure 6). A concatenation method of this kind was first elucidated in 1967 by Adler 0.

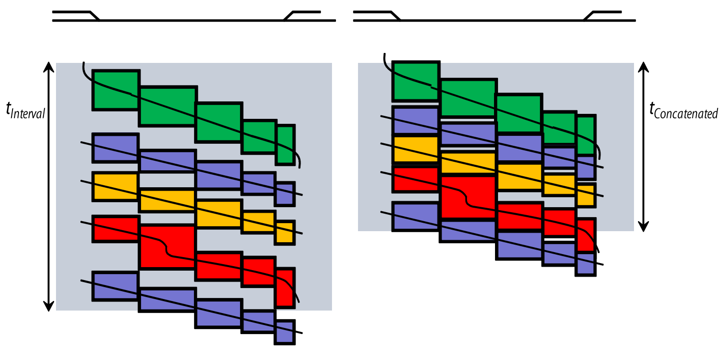

Figure 6. Concatenation of blocking-time series

The concatenated track-occupation factor can be computed from the concatenated timetable:

$$\rho = {t_{Concatenated}}/{t_{Interval}} \tag{9}$$

The differential between period under review $t_{Interval}$ and sum of concatenated blocking times $t_{Concatenated}$ constitutes the sum of buffer times available between train paths in the non-concatenated timetable. No distinction is made as to what use could be made of these buffer times for the purpose of designing further paths within the existing framework and whether they are required to ensure the stability of the timetable and for engineering works. The practical capacity of the railway line $n_{Trains,opt}$ can be calculated using the following equation:

$$n_{Trains,opt} = \rho_{perm} \cdot n_{Trains} \cdot {}^{t_{Interval}}/_{t_{Concatenated}} \tag{10}$$

The upper limits for track-occupation factor ρ_{perm} recommended in UIC Code 406 are shown in Table 1.

Table 1. Recommended upper limits for the concatenated track-occupation factor (as cited in UIC Code 406)

Type of line	**Peak hour**	**Daily period**	**Comment**
Dedicated suburban passenger traffic	85 %	70 %	The possibility to cancel some services in case of delays allows for high levels of capacity utilisation.
Dedicated high-speed line	75 %	60 %	
Mixed-traffic line	75 %	60 %	Can be higher when number of trains is low (smaller than 5 per hour) with strong heterogeneity.

The method of operating with a concatenated track-occupation factor rapidly gained currency amongst European IMs once it had been published in UIC Code 406. This is because it is relatively easy to apply. At the same time, though, it is an approach with one serious drawback: unlike the procedure set out in Subsection 0, UIC Code 406 does not make explicit reference to any notion of quality. The question as to how much buffer time is required in the timetable on grounds of stability is not addressed.

For the purpose of determining the upper limits cited in Table 1, therefore, the UIC arranged for a series of comparative studies to be conducted with various European IMs. Comparisons included those made with the essentially more rugged STRELE-based procedure for selected railway lines in Germany and Austria [22]. Ultimately, then, the upper limits cited in Code 406 have been calibrated against the STRELE procedure, though in the latter they relate to the range of optimum economic effect.

It is also necessary when adopting the concatenation method to exactly specify the boundaries of the infrastructure under review so as to avoid capacity being overestimated. Landex sets this out in detail in [13]. Here we are continuing with the nomenclature selected in [9]. First an area of investigation within which the service schedule does not alter notably needs to be defined. It is for this area of investigation that a capacity pronouncement is made and it is this area of investigation (plus the period of investigation) that defines the volume of trains under review. Positioned "around" this area of investigation is an area under review, which is similarly bounded by operating control points within which changes of sequence amongst the trains under review are conceivable under realistic conditions. Depending on the situation, the boundary of the area under investigation may coincide with the boundary of the area under review. Particularly where successions of junctions in large nodes are involved, however, the area under review will be larger in extent. Figure 7 illustrates how spatial boundaries are drawn.

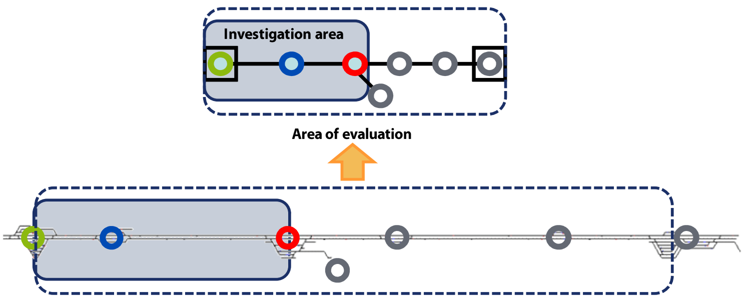

Figure 7. Areas under investigation and review

The concatenation method does not allow account to be taken of any priorities trains may enjoy. These are merely factored in implicitly through their having been borne in mind when the timetable superimposed on the base pattern was compiled.

A high degree of reference to a real timetable is, however, desirable for the purpose of gauging the 75-percent criterion for framework contracts mentioned earlier. It is additionally necessary as a means of verifying that this nominal limit is observed to adopt a uniform procedure of awarding capacity under framework contracts that is sufficiently accurate in traffic terms, is legally watertight, and takes account of capacity-related aspects. Hence there is a need to specify a procedure with the aid of which it is possible to quantify track capacity and the capacity consumed by train movements bound by framework contracts back at the stage when these are actually being drawn up. Following a selection procedure, UIC Code 406 has been incorporated into the pathing management process at DB Netz AG.

It was not, however, possible to adopt the procedure set out in the UIC Code in its original form, since it was necessary to add a range of further aspects to the basic model. In particular, the task of refining the procedure's methodology concerned the crucial distinction between train movements that are bound by framework contracts and those that are not, with their specific respective properties. There was also a need to develop specifications made in the Code regarding which reference timetable is to be used, how the period under review is to be determined and how the boundaries of sections under review are to be defined. It transpires in practical and automated concatenation, furthermore, that the method set out in UIC Code 406 fails to cover several special scenarios. Thus, for instance, there is insufficiently precise specification of how to handle multiple train crossings in the area being investigated (passing to/from the opposite siding track), any reduction in stopping-time recovery margins or trains joining and leaving the line. For a more detailed discussion reference is made to [8] and [9].

3.3. Establishing the Track-Occupation Factor as Per UIC Code 406

Unlike the "STRELE equation", the concatenation method devised by Adler assumes the existence of a real timetable when establishing capacity, since this timetable is compressed. Particularly where lengthy horizons are considered, however, it may well be that there is no such timetable of this kind as yet or else there is no precise knowledge at the beginning of a timetable period of freight paths scheduled ad-hoc. Pronouncements on the productive efficiency of infrastructure directly depend on the sequence of trains, moreover, which may be extremely variable where lengthy horizons are considered.

It is for this reason that UIC Code 406 sanctions proceeding independently of timetables with the aid of minimum headway times as a remedial measure. This approach is comparable to adopting unscheduled waiting times as set out in Subsection 0 except that track-occupation factor ρ is arrived at directly via average minimum headway time h_{Avg} and quantity of trains n_{Trains}:

$$\rho = {n_{Trains} \cdot h_{Avg}}/{t_{Interval}} \tag{11}$$

3.4. Saturation method

A third means of determining line capacity is the saturation method, another timetable-dependent procedure. Instead of compressing the timetable by means of concatenation, however, a check is run of how many additional non-conflicting paths would be feasible in the underlying timetable set-up in order to establish the maximum level of capacity.

The first step involves classifying a path from the timetable deemed representative of the section of line under review. Consideration is, for instance, given to speed profiles, stopping patterns along the section of line and frequency of appearance in the timetable. This "elementary path" is now envisaged as recurring in the train diagram. This yields $n_{Trains,max}$ elementary paths for the time slot under review. The actual timetable embracing n_{Trains} paths is then superimposed upon the base pattern generated (Figure 8).

There remains a total of $n_{Trains,conflict}$ elementary paths that are in conflict with other instances of track occupation under the timetable. (In the example, there are twelve elementary train paths whereof ten are subject to conflicts.) The quotient yields track-occupation factor ρ:

$$\rho = {n_{Trains,conflict}}/{n_{Trains,max}} \tag{12}$$

This track-occupation factor can conceivably be built upon by conducting an implicit assessment of practical capacity by analogy with the procedure for concatenation, i.e. adopting the threshold values from Table 1:

$$n_{Trains,opt} = \rho_{perm} \cdot n_{Trains} \cdot {n_{Trains,max}}/{n_{Trains,conflict}} \tag{13}$$

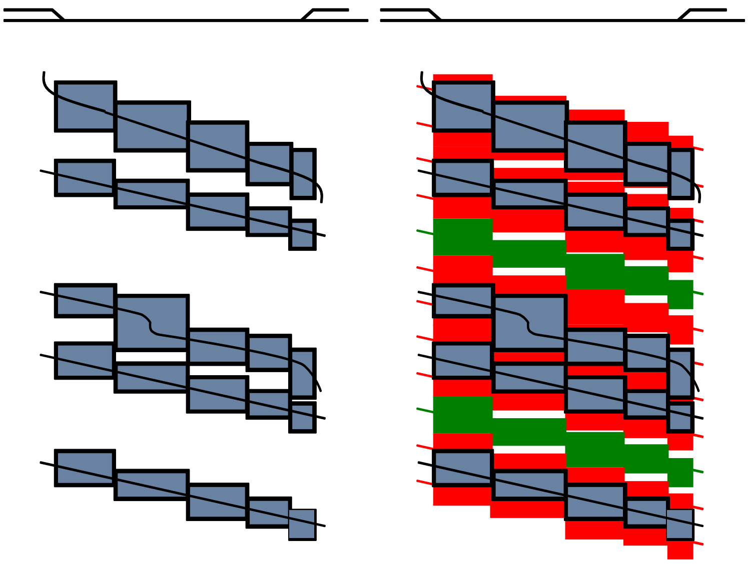

Figure 8. Saturation method

The saturation method involves similar deliberations when it comes to defining the boundaries of the area under investigation as with adoption of the STRELE equation and concatenation. The area of investigation describes the spatial extent to which the timetable is superimposed on the base pattern. It needs to be extensive enough for consideration to be given to conflicts between elementary paths and trains with significantly different speed profiles over several subsections if need be.

The method is susceptible in that it is greatly dependent on how elementary paths are distributed absolutely through time. Moving the base pattern by as little as a minute can yield substantially different results. It could conceivably be refined by defining an elementary train and repeatedly conducting path searches for it in the timetable instead of superimposing the timetable on the base pattern of elementary paths. Applicable path-search algorithms are a constituent in the simulation of scheduling (cf. applicable Section in [12]). Adopting an asynchronous approach for this additionally enables account to be had of the priorities of the trains introduced [7], something that is not possible with the original procedure. It can likewise be contemplated varying the characteristics of the elementary train in this step and hence of mapping the actual service schedule more realistically.

4. Methods of Establishing Node Capacity

Numerous practical examples in recent years have shown that the capacity of nodal areas usually effectively defines overall capacity. Despite lines being efficiently designed, it is frequently not possible to move trains “onto” or “from” them. Thus, for pronouncements to be reliable, it is essential that attention also focus on capacity in the nodal area. It is intended in this Section to provide a supplementary overview of procedures facilitating analysis of traffic in nodes.

Nodes, here, are basically understood to be stations with a set of tracks and a throat comprising one or more switching complexes. The switching complexes may be considered to be route nodes, though a straightforward junction similarly constitutes a route node. In the wider sense, a network section can likewise be interpreted as being a route node. With a route node generally connecting several lines, concatenating a real timetable yields data of impaired meaningfulness due to interactions between incoming and outgoing traffic whilst also rendering it impossible to identify the real weak spots in the nodal area. Similarly, adopting the saturation method is only worthwhile to an extent, since even elementary paths would have to be run through the node point to point and would have to be prevented from conflicting with one another. Neither of the two timetable-dependent approaches are suitable for handling nodes, therefore, and simulation has to be ruled out here too.

Instead, all methods of establishing node capacity involve modelling the service schedule along non timetable-dependent lines as well as decomposing the track topology into queueing systems. Such queueing systems may be single-channel, partially multi-channel or continuously multi-channel. From a queueing-theory point of view, a node constitutes a network section with a multiplicity of queueing points concatenated sequentially and in parallel. This network section has to be divided up into components for its capacity to be established. Single-channel components are referred to as sectional route nodes (SRNs), which represent the largest continuous section of track on which all movements conflict, as portrayed in Figure 9. Division into sectional route nodes is completely independent of route and signal logic, incidentally, being solely based on the topology of the track plus stopping points. The influence of control systems is, instead, wholly taken account of by means of minimum headway times.

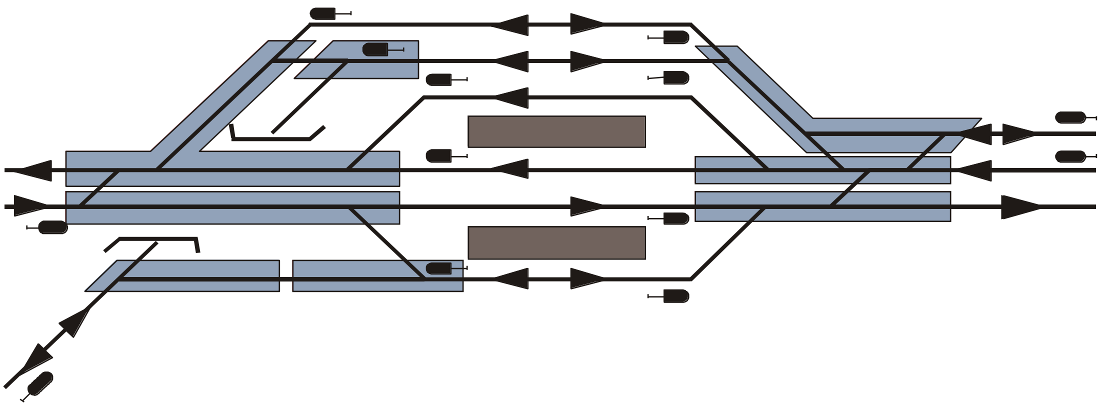

Figure 9. Dividing a station up into sectional route nodes

Dividing a station or a network section up into sectional route nodes has an important practical relevance, since it is of interest for dimensioning issues to be able to identify system bottlenecks. Hence, division into SRNs is a vital means of accurately localising bottlenecks. Items not constituting SRNs are waiting points, i.e. passing tracks and sets of tracks on which trains can wait for a section to become clear.

Station throats generally comprise several SRNs, which can be combined to form a non-sectional route node (RN). The RN represents a queueing system that may be multi-channel at least part of the time where moves in parallel are feasible. Sets of tracks, to conclude, are regarded as multi-channel systems for queueing-theory purposes.

4.1. Unscheduled Waiting Times

A distinction needs to be made when considering unscheduled waiting times as to the queueing system for which they are to be established:

- The procedure elaborated by Schwanhäußer for the line, which is set out in Subsection 0, can be adopted virtually wholesale for the purpose of determining the capacity of each sectional route node (SRN) individually. The unscheduled waiting times arising in the sectional route node (knock-on delays) can be contrasted with a targeted level of service and in this way yield the maximum quantity of trains for any given period of time that can be operated through the sectional route node.
- To compute the waiting times arising in the non-sectional route node (RN), it is possible to reduce the latter to a single-channel pseudo-system as an assisting measure. However, a process known as “specified extrapolation” proves more rugged in practical applications. This involves raising the quantity of moves negotiating the SRNs in an RN until such point as the SRNs concerned attain their permissible capacity. Thereafter, only the numbers of trains on routes leading to SRNs that have yet to be saturated are successively raised. This approach allows the RN’s practical capacity to be established whilst also retaining the routing. If the exercise is limited to raising numbers of trains on individual point-to-point routes and for individual train types, it is, for instance, possible to look into how much more freight traffic a node can accommodate assuming the pattern of passenger services remains constant. This has become a standard approach in Germany and is directly supported in LUKS®. Nießen illustrates in [14], moreover, that it is also possible to establish the theoretical productive efficiency $n_{Trains,max}$ of an RN given variations of routing. This can then be used to arrive at practical productive efficiency $n_{Trains,opt}$.
- A self-contained method of establishing the unscheduled waiting times arising in a set of tracks is currently at the research stage. The particular challenge here involves applying the mutual justifiability of the queueing channels to queueing theory.

It needs to be ensured when defining the area under review “around” the nodal area being investigated that minimum headway times are not underestimated due to the area under review being too small.

4.2. Scheduled Waiting Times

Unscheduled waiting times provide information as to the level of quality with which infrastructure is capable of handling traffic. It is also worthwhile, however, delivering pronouncements on how much scheduled waiting time arises in the course of timetabling if the paths ordered by TOCs cannot initially be implemented in non-conflicting form. Discrepancies can, for instance, arise between paths ordered and the running slots actually provided on account of running-time allowances and extended stopping times. These are duly referred to as scheduled waiting times.

Factors besides the quantity and mix of trains impacting on the scheduled waiting times arising are in particular the minimum headway times, their mean variation, and timetabling priorities. With the mean variation of minimum headway times also exerting an influence on the indicator, it can be elegantly demonstrated that homogenising speed profiles causes capacity to rise (though this shall not be exhaustively demonstrated in this article).

Whereas it is assumed when establishing unscheduled waiting times that the initial timetable was non-conflicting and that, as a consequence, any knock-on (secondary) delays arise out of initial (primary) delays, exercises to establish scheduled waiting times posit random intermediate times of arrival for paths ordered. As with unscheduled waiting times, it is possible to define a targeted level of quality using the measure of permissible scheduled waiting times. Put another way, this states how much average modification of an ordered path is regarded as being market-compliant.

As in Subsection 0, we distinguish by queueing system:

- It is possible to establish the scheduled waiting time arising at an SRN and to compare and contrast this with the permissible level of service. This enables those areas of infrastructure to be determined that pose particular challenges for timetabling.
- Figure 10 illustrates the "timetable quality factor", i.e. the quotient of actual scheduled waiting time and permissible scheduled waiting time per SRN.
- As with unscheduled waiting times, here, too, either an approximatively single-channel pseudo-system may be posited or else there is the option of performing a specified extrapolation in order to establish the productive efficiency of the entire RN.
- Procedures for establishing scheduled waiting times in sets of tracks already exist, which is not the case with unscheduled waiting times. Whilst the procedure devised by Potthoff permits computation of the characteristic values for low track-occupation factors ("light traffic") [16], the Hertel approach is considered to be a "heavy-traffic model" [6].
- It is also possible by adapting the single-channel queueing system for an SRN to apply the STRELE equation approach set out in Subsection 0 to the task of establishing scheduled waiting times along the section of line. Why such an analogy makes sense is adumbrated in the concluding Section 5.

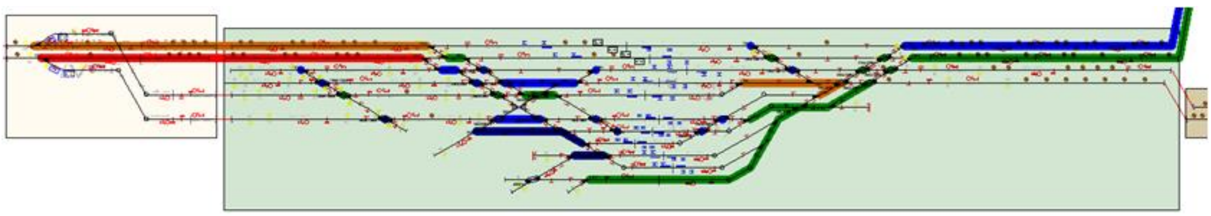

Figure 10. Scheduled waiting times in an illustrative network

The same pronouncements regarding the dimensioning of the area under review as made in Subsection 0 apply to SRNs, RNs and sets of tracks, the same boundary conditions as in Subsection 0 to any consideration of lines.

4.3. Establishing the Track-Occupation Factor

Although UIC Code 406 solely describes the process of establishing line capacity, its non-timetable-dependent approach can also be applied to SRNs (and with limitations to RNs). As set out in Subsection 0, this involves establishing an average minimum headway time h_{Avg} for the SRN and arriving at a market-compliant quantity of trains via track-occupation factor ρ_{perm}. The boundaries of the area under review accordingly need to be drawn as set out in Subsection 0.

5. Concluding Summary and Outlook

There are a series of capacity issues that need to be addressed both in operational infrastructure planning and, increasingly, also in pathing management. Highly-evolved models and software tools are available to this end. It is not possible to approach all the multifarious questions to which answers must be found adopting a single method or tool. Differences arise in respect of the level of detail of the infrastructure model, the necessity for either a non-timetable-dependent or a timetable-dependent procedure respectively, and whether implicit or explicit quality criteria are adopted.

So as to avoid conflicts in the practical handling and interpretation of results, it is advisable for the methods used to be based on a single set of principles where at all possible. This is already the case in a large number of countries in which the blocking time model is now adopted for virtually all scenarios.

By contrast, there continues to be a need for discussion regarding, for example, the method to be selected for the purpose of identifying railway lines set to become congested in the period ahead. It is a question here of reaching a compromise between pronouncements on medium-term trends on the basis of an abstracted service schedule and of specific constellations in the timetable under review.

Interest is also increasingly focusing on the definition of key-performance indicators that allow two types of information to be provided on capacity across an entire network. They serve as a means of describing how many paths it would be possible to make available to the market. Another matter of interest is the degree to which capacity provided is translated into paths actually worked (and how many orders it may not have been possible to fulfil). It is necessary in this context to segregate non timetable-dependent and timetable-dependent approaches in suitable manner so as to avoid information being corrupted. It is additionally important to select an appropriate form of standardisation conducive to a valid aggregation of indicators across networks. Also worth pursuing, to conclude, is the issue of how capacity in nodal areas is to be adequately factored into cross-network KPIs.

References

[1] Adler, G.: "Die Verkettung der Streckenbelegungen und der Belegungsgrad einer Gesamtstrecke". Dissertation, Hochschule für Verkehrswesen Dresden (1967).
[2] Bundesgesetzblatt (Federal Law Gazette): „Allgemeines Eisenbahngesetz (AEG)" (2012) (Germany's General Railway Act)
[3] Büker, T: "Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen". Dissertation, RWTH Aachen (2010).
[4] Büker, T.; Kogel, B.; Nießen, N.: "Influence of the European Train Control System (ETCS) on the capacity of nodes". Research report commissioned by the International Union of Railways (UIC) (2010).
[5] Directive 2001/14/EC of the European Parliament and Council.
[6] Bundesgesetzblatt: "Verordnung über den diskriminierungsfreien Zugang zur Eisenbahninfrastruktur und über die Grundsätze zur Erhebung von Entgelt für die Benutzung der Eisenbahninfrastruktur (EIBV)", (2009) (Germany's Railway Infrastructure Usage Regulations)
[7] Fischer, K.; Hertel, G.: "Bedienungsprozesse im Transportwesen", Transpress Berlin (1990).
[8] Gröger, T: "Timetable Simulation by Sophisticated Conflict Resolution". In: Computer in Railways IX (Comprail 2004), (2004), pp. 573-581.
[9] Gröger, T.; Kuckelberg, A.; Wendler, E.: "A UIC-compliant, practically relevant capacity-consumption evaluation algorithm". In: Proceedings of the 4th International Seminar on Railway Operations Modelling and Analysis, Rome (2011).
[10] Gröger, T.; Nießen, N.; Wendler, E.: "Kapazitätsverbrauch rahmenvertraglich gebundener Zugfahrten". In: Eisenbahntechnische Rundschau (ETR) 61 (2012) 1+2, pp. 15-21.
[11] Happel, O.: "Sperrzeiten als Grundlage für die Fahrplankonstruktion". In: Eisenbahntechnische Rundschau (ETR) 8 (1959) 2, pp. 79-90.
[12] Hansen, I.; Pachl, J.: "Railway Timetable & Traffic: Analysis - Modelling – Simulation", Eurailpress (2008).
[13] Janecek, D.; Weymann, F.; Schaer, T.: "LUKS - integriertes Werkzeug zur Leistungsuntersuchung von Eisenbahnknoten und -strecken". In: Eisenbahntechnische Rundschau (ETR) 59 (2010) 1+2, pp. 25-32.
[14] Landex, A.: "Methods to estimate railway capacity and passenger delays". Dissertation, DTU Transport (2008).
[15] Nießen, N.: "Leistungskenngrößen für Gesamtfahrstraßenknoten". Dissertation, RWTH Aachen (2008).
[16] Oetting, A.; Nießen, N:. "Eisenbahnbetriebswissenschaftliches Werkzeug für die mittel- und langfristige Infrastrukturplanung". In: Tagungsband 19. Verkehrswissenschaftliche Tage, Dresden (2003).
[17] Potthoff, G.: "Verkehrsströmungslehre Bd. 1 (Zugfolge auf Strecken und Bahnhöfen)", Transpress Berlin (1970).
[18] Schwanhäußer, W.: "Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn." Publications by the Transport Science Institute at RWTH Aachen, Vol. 20 (1974).
[19] Schwanhäußer, W.: "Wirtschaftlich und betrieblich optimale Zugzahlen auf Eisenbahnstrecken." In: Eisenbahntechnische Rundschau (ETR) 58 (2009) 9, pp. 488-495.
[20] International Union of Railways (UIC): "UIC Code 406 Capacity" (2004)
[21] Wendler, E.: "Influence of automatic train control components with an infill functionality on the performance behaviour of railway facilities". Research report commissioned by the International Union of Railways (UIC) (2005).
[22] Wendler, E.; Nießen, N:. "Stochastische Modelle in der Eisenbahnbetriebswissenschaft". In: ZEVrail Glasers Annalen 132 (2008) 1-2, pp. 40-48.
[23] Wendler, E. et al: "Empirical analysis of capacity consumption for the new capacity leaflet of UIC". Research report commissioned by the International Union of Railways (UIC) (2001-2002).