Monetary Evaluation of Dispatching Decisions in Consideration of Mode Choice Models

Marcel Schneider, Nils Nießen

Abstract—Microscopic simulation tool kits allow for consideration of the two processes of railway operations and the previous timetable production. Block occupation conflicts on both process levels are often solved by using defined train priorities. These conflict resolutions (dispatching decisions) generate reactionary delays to the involved trains. The sum of reactionary delays is commonly used to evaluate the quality of railway operations, which describes the timetable robustness. It is either compared to an acceptable train performance or the delays are appraised economically by linear monetary functions. It is impossible to adequately evaluate dispatching decisions without a well-founded objective function. This paper presents a new approach for the evaluation of dispatching decisions. The approach uses mode choice models and considers the behaviour of the end-customers. These models evaluate the reactionary delays in more detail and consider other competing modes of transport. The new approach pursues the coupling of a microscopic model of railway operations with the macroscopic choice mode model. At first, it will be implemented for railway operations process but it can also be used for timetable production. The evaluation considers the possibility for the customer to interchange to other transport modes. The new approach starts to look at rail and road, but it can also be extended to air travel. The result of mode choice models is the modal split. The reactions the end-customers have an impact on the revenue of the train operating companies. Different purposes of travel have different payment reserves and tolerances towards late running. Aside from changes to revenues, longer journey times can also generate additional costs. The costs are either time- or track-specific and arise from required changes to rolling stock or train crew cycles. Only the variable values are summarised in the contribution margin, which is the base for the monetary evaluation of delays. The contribution margin is calculated for different possible solutions to the same conflict. The conflict resolution is optimised until the monetary loss becomes minimal. The iterative process therefore determines an optimum conflict resolution by monitoring the change to the contribution margin. Furthermore, a monetary value of each dispatching decision can also be derived.

Keywords—Choice of mode, monetary evaluation, railway operations, reactionary delays.

I. INTRODUCTION

Analyses of railway infrastructure and operations are conducted using methods of railway operations research. A major purpose of railway operations research is the dimensioning of railway systems based on a functional relation between capacity utilisation and an utilisation based quality parameter.

There are two levels of processes in railway operations research. The first one is production of a zero-conflict timetable. The train operating companies request specific train paths. The dispatcher tries to write a timetable that accommodates all train path requests, in line with existing access rights. Some requests may compete for the same available capacity. In that case, there are different measures to find a solution to the conflicts for the relating companies. In the end there will be a zero-conflict timetable. The second process starts on that basis. In railway operations, different types of delays can occur due to perturbations caused by infrastructure, staff or rolling stock. A dispatcher solves these conflicts in a way that operations recover back to normal as soon as possible.

In both processes occur waiting times for the involved train operating companies. In the timetable production process, scheduled waiting times are considered whereas unscheduled or reactionary waiting times are considered in railway operations [1]. This paper focuses on railway operations only. In this process level a functional relationship is established based on the accumulation of all reactionary delays [2]. Current approaches for a transport economic quality perception are based on a limitation of the sum of all reactionary delays or on heuristic linear perception functions of delays.

There is no transport economic measure for the customer’s perception of individual dispatching decisions in railway operations so far, without well-founded heuristic functions. This paper represents an alternative approach based on a combination of methods from railway operations research and mode choice models. This provides a more detailed perception of reactionary delays by taking into account effects on the end-customer.

A coupling of the methods of transport economics and railway operations research has so far only been used in the context of strategic infrastructure planning and long planning horizons. This kind of infrastructure planning requires a high degree of abstraction. The modelling involves macroscopic network graphs and a grouping of train services with similar characteristics. In this context elasticity based demand models are widely used. By contrast, in railway operations research, the modelling is based on a microscopic infrastructure input. This allows dispatching decisions for a short term planning horizon. For the perception of these decisions the strategic network planning approach cannot be used since its degree of abstraction is too high.

This new approach wants to link microscopic operations simulations with mode choice models. These models also rely...
on a macroscopic description, yet the degree of abstraction is considerably lower compared to strategic network planning. The usage of mode choice models allows accounting for the modal split of end-customers and its increase or decline. This has an impact on the revenues of affected train operating companies. The revenues and variable costs are used to calculate the contribution margin, which depends on the dispatcher’s decision. By comparing the contribution margin in the presence of conflicts to the one of the zero-conflict timetable a monetary evaluation of reactionary delays can be obtained.

Fig. 1 Coupling of railway operations and mode choice models

II. METHOD

In railway operations simulations occurring occupation conflicts are solved based on the given priorities of trains. The new approach aims to represent monetary effects of delays in microscopic simulation programs. This is achieved by an iterative re-evaluation of the given train priorities which in turn influences the conflict management decisions. The overall model can be divided into the three submodules “simulation of railway operations”, “mode choice model” and “change of contribution margin”, which are coupled according to the graphical representation in Fig. 1.

The basis of the railway operations simulation is a microscopic model which requires a detailed input of infrastructure data and train service specification. Moreover, it also contains a zero-conflict timetable with train-specific running times and service frequencies. This data forms the input to the macroscopic mode choice model. The passengers’ choice amongst competing modes of transport on a given journey is determined using this model (cf. chapter IV) so that the modal split can be derived for each journey.

Based on the determined modal split and train-specific parameters such as the maximum number of seats, average degree of usage (based on all serviced routes) a train operating company’s train specific revenues can be derived. In addition, the costs for running the train as a function of running time and distance are determined. For each train and each route costs and revenues are summarised in the contribution margin so that they can be compared to the zero-conflict and hence zero-delay case.

The zero-conflict timetable is the foundation of all railway operations and dispatching simulations. By entering randomised model-entry lateness this schedule is perturbed leading to the formation of infeasibilities such as occupation conflicts in the model area. These conflicts are solved according to the trains’ priorities. If conflicting trains are forced to wait or slow down before entering a track segment the conflict management induces reactionary delays “\( t_d \)” of trains. The delays and the resulting increase of travel times are transferred to the macroscopic modal choice. In this module the customers’ perception and assessment of train services are modelled affecting their choice of transport mode.

Based on the new input parameters and the updated customer decisions a new modal split of rail bound transport is calculated. For the train operating company these migration processes result in a decrease of revenues. Moreover, the compensation of service irregularities results in an increase of costs. By updating the contribution margin a monetary evaluation of reactionary delays of train rides can be performed. For the same delay scenario and a given occupation conflict this procedure is repeated with different combinations of train priorities. In general, this leads to different conflict management decisions and consequently
results in a change of reactionary delays of conflicting train rides. These reactionary delays can again be monetarily assessed according to the procedure in Fig. 1.

By repeating this procedure the conflict management decision with minimal monetary loss is determined. Hence, the proposed approach allows determining an economically optimal conflict resolution decision by monetarily re-evaluating prescribed priorities of conflicting train rides and calculating the corresponding changes to the contribution margin.

III. CONFLICT RESOLUTIONS IN RAILWAY OPERATIONS

Railway operations simulations can be used both to investigate current operational processes and to improve future schedules. On both process levels occupation conflicts are solved based on prescribed train priorities. The operations simulation of the paper focuses on aim at modelling anticipatory actions in the dispatching process. For each type of train the corresponding input parameters are determined from historic delay data. They include information about the probability and the expected value of delays to trains [3].

Based on this data, randomly initiated delays are imposed onto the unperturbed system. They consist of model-entry lateness of trains entering the model area and initial delays occurring within the model area. Initial delays can be caused by dwell times overruns or speed reductions resulting from incidents such as signal failures, asset failures, rolling stock or train crew availability or other unforeseeable effects. Delays resulting from interaction between different train runs explicitly do not fall into this category. Model-entry lateness results from delays having occurred outside the model area irrespective of their origin i.e. it is not distinguished between them being caused by initial delays outside the model area and them being from reactionary delays due to conflicts before entering the model area. The occurrence of initial delays, which are also called first order delays since they do not involve the interference between different train rides, can lead to occupation conflicts in the model area. In Fig. 2, the occupation conflict between two train runs “i” and “j” is depicted based on a blocking time staircase representation of train runs.

For the resolution of conflicts the dispatcher disposes of a number of measures which are subjected to certain conditions. For railway network in Germany, there is a Deutsche Bahn (DB) guideline 420 [4]. This guideline contains rules for a transparent communication and cooperation between the largest infrastructure company in Germany, DB Netz AG, and the train operating companies, who want to run trains on their infrastructure. Module 420.0201 sets out the following equally important dispatching objectives:

1) Quickest possible restoration of normal operations.
2) Ensuring traffic fluidity.
3) Improvement of the overall punctuality of all trains.
4) Optimise capacity utilisation of both railway lines and stations.

In order to regulate the dispatching process trains are attributed a numerical rank. This index determines the priority a certain train is given in case of track occupation conflicts. High ranks correspond to high priorities whereas low ranks correspond to low priorities. If a track occupation conflict occurs the lower-ranking train has to yield the right of way to the higher-ranking train. That means, it has to slow down or wait until the higher-ranking train has cleared the block section. Hence, the distribution of ranks is fundamental for the dispatcher’s choice of conflict resolution since a difference of rank gives priority to the higher-ranking train as compared to the lower-ranking one. Depending on the ranking of trains the dispatcher has to adhere to the following order of regulations in line with the above mentioned objectives of the dispatching process.

1) Urgent trains have to be given priority over all other trains.
2) High speed passenger services are given priority over all other trains (expect those in 1)).
3) High speed freight services are given priority over all other trains (expect those in 1) and 2)).
4) Trains which are not specifically mentioned in 1) to 3) are treated equally.
5) If two trains have the same rank the faster trains are given priority over slower trains.
6) Trains on special railway infrastructure have priority on these tracks if their operation is aimed at the line of the services the tracks are designed for.

Additional dispatching measures can be introduced in conflict management in railway operations simulations and analysed with respect to the objectives of the dispatching process. Those measures can comprise spatial or temporal changes to train paths as well as a different routing on railway lines or in stations. Four widely used conflict resolutions...
measures are bending of train paths by reducing the speed of the following train (Fig. 3), parallel temporal shift of train paths (Fig. 4), increase of dwelling times (Fig. 5) or the overtaking of slower trains in stations (Fig. 6). These corresponding blocking staircases of these conflict resolution measures are outlined in Figs. 3-6 [5].

The resolution of conflicts results in, aside from the lateness causing the conflict, a reactionary delay for at least one other train. The predefined priorities of trains mean that mostly trains with lower priorities get delayed. In the example of bending (Fig. 3), train “j” runs at a reduced line speed so that the upcoming conflict can be avoided. Train “j” now does not arrive according to its scheduled time at the end of the section but with an imposed delay as a result of the conflict resolution (reactionary delay).

The approach presented in this paper continues to assume priorities assigned to trains. The results are the same for trains of lower priorities in form of reactionary delays. However, priorities are varied so that potentially different conflict resolutions for the same perturbation scenarios and trains can be found. Going forward, the exact location of the reactionary
delays to trains “i” and “j” are also of importance. This is possible to be both on plain lines (as a result of solving the conflict by bending) as well as stations and junctions (as a result of solving conflicts by dwell time extensions, shifting or overtaking). The resulting reactionary delays affect the end-
customer in different ways (cf. chapter V).

The modelling tool LUKS® (Leistungsfähigkeitsuntersuchungen von Knoten und Strecken = capacity analysis of nodes and lines) is being used as part of this work. LUKS® is being used for instance by DB Netz AG, the largest railway infrastructure company in Germany, as well as the German railway regulator, the Eisenbahnbundesamt, as the standard tool for any questions relating to capacity. These can be approached with simulative methods (LUKS-A) [6].

IV. MODE CHOICE IN LONG DISTANCE PASSENGER TRAVEL

Conclusions about the split of passengers or goods between different modes can be drawn from mode choice models. Different approaches exist to determine this modal split. Specific mode choice models have to be applied due to the different parameters and available modes of transport to long-distance, regional and inter-urban passenger as well as freight traffic.

As part of this paper, the mode choice shall be presented using the example of passenger long-distance travel. The applied model [7] is based on generalised costs. This way, the multi-layered service criteria can be characterised. These are correlated with each other for the different modes by means of a route-dependent modal split, and then compared.

The resistance r, which is the product of measurable service criteria and the time perception “TPF (t)” by the customer, forms the basis. The unit of the result of (1) is the resistance unit “RU”.

\[ r = t \cdot TPF(t) [RU] \] (1)

r: resistance [RU]; t: time [min]; TPF (t): time perception function (depends on t) [-].

The entire passenger’s itinerary is being described as a chain of resistances including access to and from origin and destination as well as the main journey made on rail. This enables separate consideration of service criteria and boundary conditions.

Changes in the service offer have a direct impact on the associated resistance. Based on this, the sensitivity of each of the resistances towards the result (modal split) can be analysed. The complete chain of resistances for a long-distance journey with the individual components is shown in Fig. 8.

Some objectively measurable service criteria for rail can be taken from the train service specification. For example, the journey time, the service frequency and the potential waiting time at the platform are of relevance. The associated time perception can be described as a time-dependent function. In principal, this function (2) follows an exponential curve.

\[ TPF(t) = a \cdot e^{b \cdot t} \] (2)

a, b: coefficients; t: time [min].

Each time-related service criterion is being multiplied with the associated weighting coefficient. Thus, the value for the resistance is a multiple of the measurable time. That’s because the end-to-end journey time is subjectively being perceived to take longer than it actually does. Each resistance has a different time perception function. This way it can be represented that for instance the waiting time is perceived more negatively than travel or access time. As train services cater for different markets, the long-distance model differentiates between:

- business journeys
- commuter and education journeys
- leisure journeys (≤ 4 days)
- holiday journeys (> 4 days).

They are different for multiple reasons. For instance, the need to be on time for a meeting at the end of journey or the frequency of journeys plays an important role. These different factors influence how positively or negatively late running is being viewed. The perceived value of time curve for the example of a waiting time on the platform is shown in Fig. 8 depending on the purpose of a journey.

It can be seen that longer waiting times are perceived more negatively than shorter waiting times. The time perception differs depending on the purpose of the journey.

The curve to the left side creates higher time penalties for longer waiting times than the ones to the right side. This means that business travellers are more sensitive towards lateness than holiday travellers. This way, the lower tolerance limit of business travellers towards delays than holiday travellers are considered in the differing functions.
A clear definition of delays resulting from conflict resolution as described in Chapter III is crucial when coupling both simulation and mode choice model. From the point of railway operations, the relevant delays can be described as model-entry lateness, initial, and reactionary delays. The localisation of the customer’s delays is a critical factor for the transfer into the choice mode models. If the passenger has to wait for the train on the platform, this is a delay at the beginning of the journey. If they are already on the train when the train is being delayed as a result of a conflict resolution, this is a delay at the end of the journey. In the latter case, their train potentially departed on time but arrives late at its destination.

In addition, the waiting times identified through simulation have to be assigned to the correct individual resistance. The delay at the beginning of the journey is an element of the interchange resistance at the origin station. In contrast, the lateness at the end of the journey is an element of the running resistance. Delays which are occurring while waiting at a station are perceived more negatively than delays which are occurring during the train journey itself.

Aside from the time components which directly occur during the journey, there are further resistances which relate to events before or after the journey. These are, for long-distance travel, the additional time to allow for travel information or, potentially, a required overnight stay. The mode choice model allows for a comparison of time and cost aspects. The cost resistance takes the fare into account that is due for a specific journey. In addition, the willingness to pay the fare is represented. This means that the fare is also being valued by the customer. The willingness to pay a certain fare is dependent on the type of journey. The variety of fares, such as for single or multiple journeys as well as season tickets, has to be regarded differently. The comparability of time and cost resistances is achieved by an equivalence factor. This also takes into account the competition between different modes of transport by means of subjectively perceived differences in quality.

All resistances of a train journey are added to the total resistance. This is also undertaken for the alternative modes of transport. In this project, car traffic is being investigated as the only competing mode. For this, specific resistances also exist for access to and from the journey as well as the journey itself. However, there are no resistances for transfer or system change. On the other hand, the cost resistance includes a number of components compared to a train journey. The running costs for the car as well as petrol and parking are taken into account. Finally, the sum of all individual resistances “rsum” is being taken in (3) to identify the split for each mode “k”.

$$MS_k = \frac{1}{\sum r_{sum}}$$  \hspace{1cm} (3)

MS: Modal-Split [-]; k: mode of transport; rsum: sum of all individual resistance of each mode of transport [RU].

If individual resistances change due to variances in the service criterial, this has a direct impact on the modal split. Each supply parameter has an own individual resistance. This way, it is possible to quantify the change of demand due to changes in the service offer. The impact of delays due to conflict resolutions from the operations simulations can therefore be quantified through the modal split in the same way. Occurring delays can increase the respective resistance and decrease the modal split of rail. This means that some of the customers will be lost on the observed route, which ultimately leads to a loss in revenue train operating companies.

V. PERCEPTION OF DISPATCHING DECISIONS

The reactionary delays listed under chapter III are used in railway operations research as the key parameter for determining the quality of operations [1]. In order to evaluate this quality in a transport economical way, there is an acceptable limit for the sum of unscheduled waiting times “t unsuc” on a section [3]. The function of the acceptable waiting time refer to (4) depends on the ratio of passenger trains “pptr” that runs on this section within a specific investigation period “t_i”.

$$adm \sum t_{unsuc} = q_o \cdot 0,260 \cdot e^{-1,3 \cdot pptr} \cdot t_i$$  \hspace{1cm} (4)

t unsuc: unscheduled waiting times [min]; qo: quality factor [-]; pptr: ratio of numbers of passenger trains of all trains [-]; t_i : investigation period [min].

An increasing number of trains also mean an increasing level of interaction between trains. Reactionary delays are calculated using the STRELE-equation of Schwanhäußer [2]. They then can be compared with the acceptable sum for the associated area. The ratio of actual and acceptable waiting times is called the quality factor “qo”. DB Netz AG has specified four categories to describe the quality of railway operations in DB guideline 405.0104 [8]. The quality factor is shown for each category in Table I. 

![Fig. 8 Time perception curve for average waiting time [7]](image)
It is also possible to identify the economically optimal number of trains with aid of the acceptable level of reactionary delays. This information can be used to inform the required scale of rail infrastructure. The acceptable level is based on expert surveys of signallers and dispatching staff. The assessment of the optimal number of trains based on the robust operations (quality) on different routes forms the basis. The effects of reactionary delays on the end-customer are not considered. They can switch to competing transport modes if they are not satisfied with the levels of punctuality which can lead to a loss in revenue for train operating companies.

A consideration of end customers is possible by using delay perception functions “DPF” [9]. Again surveys form the basis of this, but they are tailored to the customer and not to signallers and dispatching staff. The differing perceptions of the four different purposes of travel (cf. Chapter IV) are considered. In addition, there is a distinction between two groups depending on the journey distance. The surveys are explicitly aimed at the subjective customer’s perception of delays and journey times. As a next step, the ratio of lateness and journey time evaluation is calculated which results in the delay perception factor. In any case a value of greater than one represents a much more negative perception of delays than the time on the train.

In a further step the “DPF” are used to calculate the modal split of rail. This way, there is not only a comparison of delays to other elements of the rail journey time, but also a comparison with competing transport modes. The resulting decline in demand due to modal shift away from rail will be used to describe the impact on revenue of train operating companies [9], [10]. It is based on the resulting interaction between different trains (STRELE-equation) and a mode choice model based on resistances. The reactionary delays are measured in consideration of “DPF” in an additional delay resistance. There are two different kinds of delay resistances. The localisation of customer’s delays is a critical factor. There is a different perception of delays that occur for the customer on the platform or inside the train.

For each number of trains on the line section a sum of reactionary delays across all trains can be identified. Based on this a modal split for the transport mode rail is determined in competition with other modes of transport. In combination with other train related input parameters such as the number of seats and load factor, the revenue of train operating companies across all trains on this line section are calculated. There are also time- and distance-related costs. This way, capacity utilisation can be economically assessed. It aims at identifying the optimal number of trains that maximises the profit for the train operating companies. In addition, the revenue (track access grant) and cost of the railway infrastructure companies may also be considered, so that an evaluation of the overall rail system can be made.

The new approach does no longer prioritise the economical optimum number of trains on a regarded line section. Instead the main focus is the search for an optimal dispatching decision from an economical viewpoint. The aim is to find a conflict resolution with a minimised monetary loss for all involved trains from the perspective of the train operating companies. The maximum allowable sum of reactionary delays across all involved trains is no longer relevant. Instead the individual delays imposed to the customers receive the main focus. An accurate knowledge of time and location of the incurred delays is crucial. This is being achieved through microscopic simulations of railway operations. It is possible to take the obtained reactionary delays as an input parameter into the selected mode choice model. This will still happen on the basis of a delay resistance, although it is no longer dominated by a “DPF”. It will be managed by using the time perception functions that depended on the purpose of travel (cf. chapter IV). It also depends on the type of resistance. The separate consideration of System change or travel resistance is crucial. Delays have to assign to the correct type of resistance with the correct perception through the customer.

Therefore the new approach is pursuing a coupling of the microscopic simulation model with the macroscopic mode choice model. However, an adaptation of the reactionary delays is necessary. The macroscopic model calculates the modal split based on average input parameters. The microscopically determined delays are only single events and have only very little influence on the modal split. This needs to be quantified in more detail in further studies.

The monetary evaluation of dispatching decisions should nevertheless consider the resulting reactionary delays. This is done based on the change to revenue and cost, which are summarised in the contribution margin “CM”. Additional costs to train operating companies are generated due to the impact on rolling stock and train crew cycles, as well as energy consumption. The revenue “R” stream is solely generated through fares. Reduction in ticket sales and increase in cost “C” due to conflict resolutions “cr” can be considered in comparison to the zero-conflict and zero-delay timetable “tt”. This is applied on all trains “j” that are involved in the conflict.

$$\Delta CM_j = \Delta R_j - \Delta C_j = (R_{tt,j} - R_{cr,j}) - (C_{tt,j} - C_{cr,j})$$  (5)

CM: contribution margin; R: revenues [EUR]; C: costs [EUR]; j: involved train; tt: timetable; cr: conflict resolution.

The determination of various conflict resolutions is controlled by adjusting the allocated priorities of the involved trains. Therefore, the simulation of railway operations is carried out for the same model-entry lateness and initial delays but for other priority constellations. Each dispatching decision generates different reactionary delays, revenue, cost, and thus a different contribution margin for each train. In order to
obtain the monetary value of each individual conflict resolution, any contribution margin is compared with that of the zero-conflict timetable (5). Each measure leads to, in general, a higher number reactionary delays, a loss in revenue and an increase in cost in comparison to the zero-conflict timetable. Therefore the aim is, to find a measure that has the maximum change of contribution margin “ACM” based on the timetable.

Alternatively, the step of priority allocation is repeated until the monetary loss is minimised. Thus, an optimal conflict resolution is determined with regard to the change of the contribution margin. Each dispatching decision has a directly assigned monetary value.

VI. CONCLUSIONS

This paper represents an opportunity to evaluate dispatching decisions in railway operations in an economical way. Reactionary delays are the basis for this and are already used in studies of capacity issues in railway operations research. The same conflicts between affected trains are solved for different priority constellations by using the software tool LUKS® for the railway operations simulations. The determined reactionary delays are the input variable for the macroscopic mode choice model. Thus, there is a coupling of the railway operations and transport economic research methods. An adaptation of output and input parameters is necessary to link the microscopic and macroscopic approach. An accurate quantification of running times and delays is also provided. The generic approach is presented in this paper on the basis of a long-distance passenger model. It uses travel resistances and assigns a subjective perception to each time component.

The subjective perception of reactionary delays of long-distance passengers depends on the purpose of travel. An accurate knowledge of time and location of the incurred delays is crucial. The microscopic simulation of railway operations can guarantee this. An assignment to the correct resistance and thus subjective perception can be ensured. The choice of mode is displayed on the modal split to competing modes of transport. The revenue is determined by the modal split, number of seats and load factor of each train and route. The contribution margin is formed in conjunction of the revenue and variable cost for each train. A transfer of the approach to local passenger and freight traffic will also be carried out with the help of models of choice of transport (WE 3655/1-1).

ACKNOWLEDGMENT

The Deutsche Forschungsgemeinschaft (DFG) supported the results of research within the project “Consideration of monetary effects with the conflict resolution in microscopic, hierarchical simulation approaches with the help of models of choice of transport” (WE 3655/1-1).

REFERENCES