Optimising train priorities to support the regulation of train services with the assistance of active and deductive databases

C. Sakowitz & E. Wendler
Verkehrswissenschaftliches Institut, RWTH Aachen University, Germany

Abstract

Maximisation of revenues is a fundamental goal of any business-driven railway infrastructure company. In order to achieve this target in the context of traffic regulation, it must try to avoid delays and ensure scheduled connections. However, nominally equal delays to two different trains are not equal in value from an economic point of view in most cases. Moreover, some connections between trains might be more important in this sense than others. There are complex interdependencies and reciprocal effects in railway traffic. Considering these effects, a dispatcher must evaluate possible forms of conflict resolution and the waiting times these give rise to and select the best solution possible. This is not achievable where a time-critical conflict arises at short notice. Even closed mathematical optimisation algorithms encounter limits in the case of larger railway networks due to the enormous number of constraints to be considered. This paper will therefore propose that the optimisation process be separated from the train regulation process. Instead, economically evaluated train priorities for conflict situations are to be determined with the help of active, deductive and normative rules. Existing concepts of "smart" database management systems (DBMS) with integrated active and deductive database functionalities can be used for this application. An active DBMS allows the definition of reactions to be automatically initiated by the DBMS in response to the detection of given database-related events. A deductive DBMS allows new, deducible facts to be specified, administered and specially derived from explicitly introduced facts. Train priorities are generated and assigned in detachment from day-to-day operations for lightly and heavily disrupted railway traffic respectively. Long-term optimisation of these priorities is effected by evaluating past operational data.
Keywords: traffic regulation, real-time rescheduling, decision support system, train priorities, active databases, deductive databases.

WIT Transactions on The Built Environment, Vol 88, © 2006 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/CR060491

1 Introduction

Given the possibility of sudden disruption or failure in railway operation, it is not always possible to run trains as foreseen in the timetable. The task for train dispatchers, therefore, is to promptly adopt suitable measures in the event of any disruption with a view to returning to normal service as soon as possible. The main objectives involved are twofold, namely to a) minimise delay minutes for trains and b) reduce passenger delay minutes over the entire journey chain. Failures in rolling stock, lineside equipment or the track itself lead to specific conflicts between trains which the dispatcher is required to resolve by adopting suitable measures. Priorities play a key role in how such measures are defined.

The conflicts concerned can be divided into two main categories, cf. [1]. Firstly, there are track-occupation conflicts between at least two trains. Owing to the guided motion principle underpinning the railway and to the system of technical protection employed, block sections with fixed boundaries can only be occupied by one train at most at any one time. Where several trains wish to have recourse to the same track-occupation entity simultaneously, a conflict arises that has to be resolved by prioritising one or other of the trains concerned. Lower-priority trains will then suffer an increase in running time as a result that is generally of the order of several minutes. Possible means of resolving the conflict at the expense of the lower-ranking train include switching it to an alternative route, extending a scheduled stop, relocating passing stops, adding an additional stop for operational requirements or increasing its running time, depending on what the situation calls for. A specific variant of these conflicts are track closure conflicts. This is a variation on the conflicting-route scenario in which a track-occupation entity cannot be requested for a specified period due to "external causes" (e.g. servicing work). The necessary response to such a conflict is as for a track-occupation conflict, with the train having by definition to be regarded as "lower in rank" than the external event. Secondly, at larger interchange stations in particular connectional conflicts can arise between trains. Connections serve the purpose of transferring passengers or wagons from a feeder train to one or more connecting services. If the feeder train is delayed by a few minutes, it may no longer be possible to effect this interchange without putting back the time of departure (and hence increasing the running time) of the connecting service. It is then necessary to accord priority either to the passengers or cargo on the feeder train (hold the connection) or to the punctuality of the connecting service (abandon the connection). If the decision goes against the connecting service, the latter's running time may be increased by several minutes. If the decision goes against the feeder train, passengers wishing to change trains may experience delays of up to an hour or more over the entire journey chain. Where freight operations are concerned, there may be delays of several hours in transfers of wagons at marshalling yards. A specific connectional conflict scenario takes the form of the circulation conflict. Here it is a case of the train connecting "with itself", as happens with a push/pull train at its terminus. What obviously distinguishes such connections is that breaking them is not feasible for technical reasons. One means of resolving conflicts may be to relocate the turnround to a previous station by omitting stops so as to allow the connecting service to resume running under the original schedule. Here, too, a decision has to be made on whether to accord priority to the punctual running of the connecting service or to all scheduled stops being called at. Longer journeys for one of two carriage parties are at stake here again.

To be able to optimise priority-based conflict resolution procedures of this sort, a suitable rating criterion is required. Worthy of consideration is a performance regime of the type already adopted by one or two European railways and whose Europe-wide introduction is currently being discussed by the International Union of Railways (UIC). A regime of this kind can help quantify the weighting of punctuality and connectional certainty for a train in terms of transport economics. It is possible in this way to establish a punctuality rating for any scheduled stop by a train as a function of delay minutes. It is likewise possible to determine the penalties for connections not kept. Using the criterion of said performance regime as a basis, the decisions regarding priorities ought now to be specified in such a way that the resultant conflict resolution leads to ratings that are as low as possible.

Although there appears to be a case for resolving the problem adumbrated by adopting mathematical optimisation procedures and approaches of this kind are often proposed in literature in the field (closed linear optimisation, e.g. [2] and overview in [3], enumeration methods, e.g. [4]), their practical applicability is limited since, owing to the large number of constraints to be taken into account, real-time decisions can only be made in respect of comparatively lightly loaded networks with relatively straightforward topologies due to the exponentially rising computing input involved. Other approaches propose simplifying the constraints (e.g. doing away with difficult-to-manage constraints by occupying sets of points at interchange stations) so as to arrive at a solution that is theoretically optimal but cannot be put to effect on the actually existing network.

Technology is nevertheless sufficiently advanced in principle at present to allow conflicts to be automatically detected and resolved, fig. 1. Already today, adopting heuristics allows priorities to be used for the purpose of optimising conflict resolution within the context of simulation procedures [5]. These priorities are currently static and are prescribed manually. It is not possible to adequately render the fluctuating significance of a train movement in terms of transport economics by this means. Performance regimes have only been in place at railways for a short while (if at all) and thus are not yet made direct use of for the purpose of optimising the regulation of services.

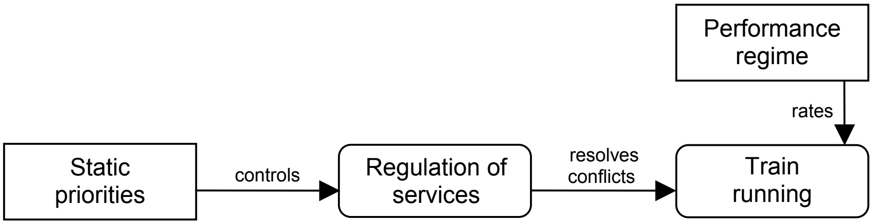

Figure 1: State of the art.

In the context of this paper, therefore, it is proposed reflecting advances made in technology by automatically adapting dynamic priorities in such a way that regulation can be optimised on the basis of a performance regime in future, fig. 2.

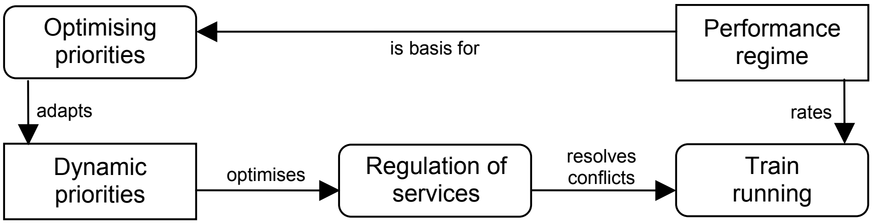

Figure 2: Proposal for new procedure.

Accordingly, a system for resolving conflicts is proposed in the following that combines the benefits of heuristic procedures (low computing time) with those of closed optimisations (quality achieved in rating traffic in terms of transport economics under the prescribed performance regime). To this end, the processes of priority-based traffic regulation, rating the outcome of regulation using the performance regime and the actual (longer-term) optimisation of priorities are divided into separate subsystems. Recourse can be had to "smart" database management systems for the optimisation process that merge deductive and active database techniques. Priorities are adapted in detachment from the day-to-day, time-critical conflict detection and resolution business of the dispatcher.

2 Modelling priorities

Priorities find application in the resolution of track-occupation, connectional and circulation conflicts by the regulation system, in which context it is also possible to regard circulation conflicts as being a special type of connectional conflict with fixed physical interlinking. It is also possible, by turn, to differentiate between train and connectional priorities.

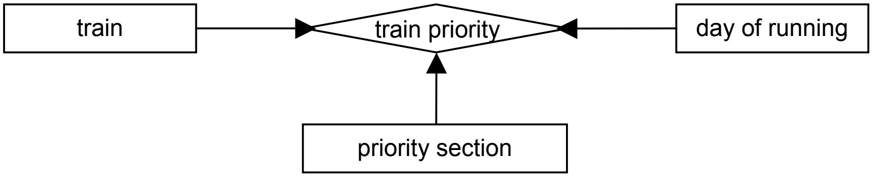

Figure 3: Train priorities.

Train priorities (see fig. 3) are defined for a train along specified, clearly demarcated sections of infrastructure on a given day of running, cf. [6]. Serving

WIT Transactions on The Built Environment, Vol 88, © 2006 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)

as the boundary for such a "priority section" is a station having sufficient buffer capacity to facilitate the requisite waiting events given a change of train priorities. Each train is accorded a specified and unequivocal ranking on a given priority section from which operational precedence amongst trains involved in any track-occupation conflict in that priority section can be derived. The higher-ranking train involved in the conflict is given precedence, whilst the lower-ranking train is forced to wait in some way. Train priorities can be used to straightforwardly model lines on which certain types of traffic (long-distance passenger services, local passenger transport, freight) have precedence, a phenomenon that is steadily gaining in significance. Moreover, it is relatively easy in this way to take account of how the importance of a train movement fluctuates in transport economics terms in the course of its taking place.

Connectional priorities (see fig. 4) are defined for a specific connection between two trains at a station. They involve specifying maximum permissible waiting periods (beyond the scheduled time of departure) for the train maintaining the connection. It is possible by taking the delayed time of arrival of the connectional feeder train, the minimum interchange time and the maximum permissible waiting period to straightforwardly establish whether a connection is to be held or whether the connecting train is to depart on time - without waiting for the connection. The same applies by analogy for circulation transfer runs of stock from one train to another.

Figure 4: Connectional priorities.

3 Description of the overall architecture

As set out in the introduction, the overall architecture can be divided up into three separate subsystems existing alongside the outside world, specifically a regulation system, a monitoring system and a priority optimisation system (see fig. 5). The outside world represents actual train running on the existing infrastructure (railway operations). Elementary faults can arise on vehicles and the track that may result in conflicts that have to be remedied on a priority basis within the framework of traffic regulation. In addition, predefined types of traffic advice (notably automated or manual running advice) are generated and signalised in advance. The regulation system serves to detect and resolve actual conflicts on the basis of priorities prescribed by the priority optimisation system. The outcome of regulation is notified to the outside world for implementation. The regulation system also notifies the priority optimisation system of any advice of failure that may have a bearing on the according of priorities. Conflicts can in principle be resolved either manually, by centralised computer-aided means (train regulation centres) or by decentralised means (distributed regulation in large networks) using the priorities given. In the overall context, there is additionally a monitoring system for the purpose of assessing traffic against the

WIT Transactions on The Built Environment, Vol 88, © 2006 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)

values prescribed under the performance regime. Standardised observations in this respect are likewise notified to the priority optimisation system so any adaptation of priorities required can be carried out. Priorities are generated by the priority optimisation system on the basis of input events conveyed and are supplied to the regulation system in good time. This process takes place with the help of rules.

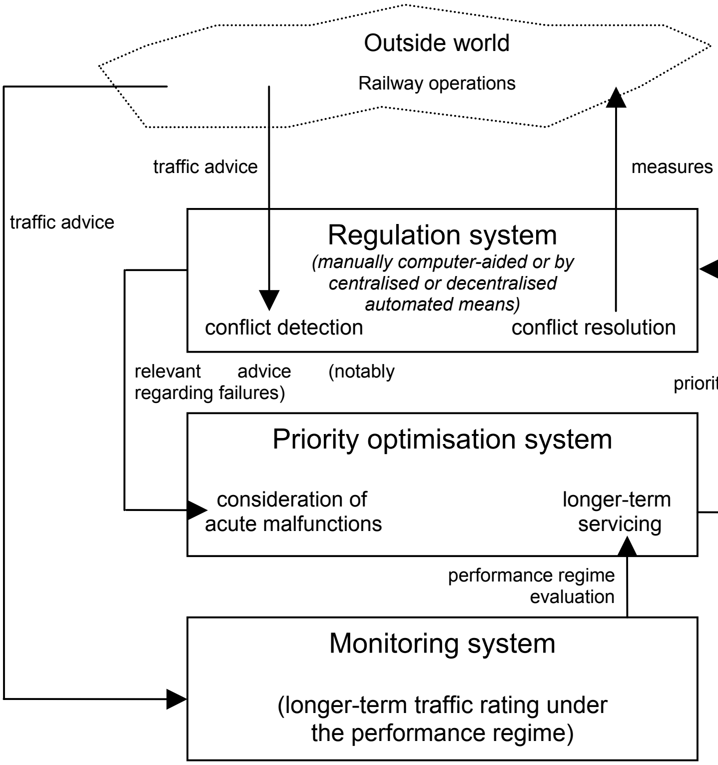

Figure 5: Overall system configuration.

The subsystems operate independently of one another and will be discussed in detail in the sections that follow. It is first intended to enumerate requirements for the regulation and monitoring system that are key to optimising priorities. Building upon this, important aspects of the priority optimisation system itself, which can be supported by both deductive and active database technology, will then be set out.

4 Key requirements for the regulation system

The (priority-based) regulation system has the function, in the event of traffic being regulated, of fine-tuning conflicting train paths by automated or manual computer-aided means on the basis of the original schedule, the priorities prescribed and traffic advice from the outside world. Key constituents of the regulation system are an advice filter for incoming traffic advice, a running-time forecasting computer plus the functions of conflict detection and resolution. It is essential for the affiliated priority optimisation system, moreover, that the regulation system additionally be in a position to signalise failure events in respect of vehicles and track installations outwards so as to be able to respond in suitable manner. It is possible to assign each failure event to a unique event category to this end. It is necessary in the process to synchronise the regulation system's very detailed data management capability with the significantly coarser modelling of the priority optimisation system.

The central purpose of the regulation system, however, relates to detecting and resolving conflicts on the basis of the priorities conveyed to it. There are three different strategies that can conceivably be adopted to resolve conflicts on a priority basis, those being manual computer-aided traffic regulation or else automated approaches of either a centralised or decentralised nature. In the case of manual computer-aided regulation, advice filtering, running-time computation and conflict detection are automated whereas actual resolution of conflicts remains the domain of the operative. With centralised automated regulation, conflict automation, too, is performed automatically by a central unit. Decentralised automated regulation, finally, involves the tasks of running-time forecasting and the detection and resolution of conflicts being assumed by several local rescheduling computers networked with one another.

5 Key requirements for the monitoring system

The monitoring system serves to rate traffic processed with the aid of the priority-based regulation system. Rating is effected on the basis of a predefined performance regime co-administered by the monitoring system. The monitoring system operates in the background and is primarily tasked with evaluating traffic notices received from the outside world, in support and independently of operational activities, to the specifications set forth in the performance regime. In the light of this evaluation, the monitoring system detects firmly preset penalty events implying a potential need for a longer-term adaptation of priorities. These events, which are each assigned to a unique event category, are then notified to the priority optimisation system.

Under the performance regime, predefined deviations from the schedule (in particular, delays and lost connections) are numerically weighted by means of penalties. Hence the value of a railway's service offerings and performance will be expressed in terms of transport economics. Reducing the penalties incurred will become a target of operating practice. This is the focus of interest towards which the according of priorities and thus regulation itself must be directed.

Penalties can feasibly be levied relative to the delay arising at a predefined reference point (as a function of delay minutes) or relative to the holding of a given connection.

6 Principal aspects of a priority optimisation system

6.1 Purpose of the priority optimisation system

The priority optimisation system acts to internally optimise priorities in regulation as well as to supply these priorities to the regulation system. To this end, the priority optimisation system obtains failure advice from the regulation system and evaluation advice (penalty payments) from the monitoring system.

Internally, the priority optimisation system distinguishes between "standard" and "degraded mode" or "disruption" priorities (see fig. 6). Standard priorities are the train and connectional priorities, firmly defined for each point in time and for each location, for use in the event of traffic being free of disruptions or only slightly disrupted. Disruption priorities, by turn, involve predefined deviations from "standard" priorities in respect of specified major degraded-mode scenarios signalised by the regulation system ("emergency plans"). In the event of traffic being largely free of disruptions, it is the respective standard priority that is conveyed to the regulation system, otherwise it is the disruption priority defined for the disruption incident concerned that is conveyed.

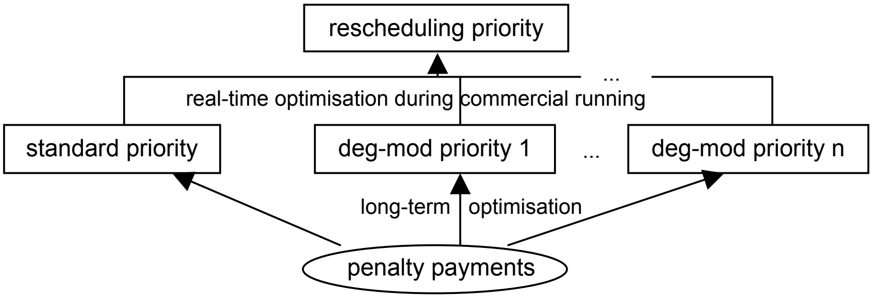

Figure 6: Priority optimisation.

Three possible kinds of change of priority controlled by a priority optimisation system are conceivable in this connection, specifically a scheduled change in the course of regulation, a change on account of a degraded-mode scenario in the course of regulation and, finally, a longer-term adaptation of internal standard and disruption priorities.

Scheduled changes of priority can arise where train priorities are concerned when a train passes from one priority section to another. In such an instance, the regulation system replaces the standard or degraded-mode priority applicable for the former section with the standard or degraded-mode priority applicable for the new section. A change of priority on account of a degraded-mode scenario, by contrast, corresponds to a short-notice temporary switching of rescheduling priorities between standard and degraded-mode priorities or between differing degraded-mode priorities owing to a sudden degraded-mode incident. The switching of priorities in respect of a degraded-mode scenario is determined beforehand by means of the degraded-mode priorities for train and connectional priorities. It is ensured in this way that there is no need to carry out extensive, time-critical ad-hoc calculations or optimisation work "with wheels rolling". Where longer-term priority adaptations are concerned, on the other hand, it is the underlying (internal) standard and degraded-mode priority data used in specific instances themselves that are optimised and not the rescheduling priorities. This is effected by means of an evaluation of actual operations by the monitoring system, observing the values prescribed by the performance regime, that is not carried out in real time and is hence largely non time-critical.

In what follows, a practical proposal for putting changes of priority to effect with the aid of active and deductive database technology is made.

6.2 Effectuation with the aid of active and deductive database technology

Recourse can be had to existing database-technology concepts for the purpose of putting a priority optimisation system into practice. Interest centres in this respect on deductive, active and supplementary normative constituents of database management systems (DBMS). Deductive DBMSs allow new deducible facts to be specified, administered and specially derived from explicitly introduced facts, cf. [7, 8]. An active DBMS allows the definition of reactions to be automatically initiated by the DBMS in response to the detection of specific database-related events, cf. [9]. Additionally, a normative constituent of the DBMS enables fixed integrity constraints to be defined that are required to be kept in each consistent state by the database.

The deductive constituent is in particular used in the context of the explicit definition of combinations having a bearing on the according of priorities. The elementary events transmitted by the traffic-regulation or monitoring system are generally unsuitable for use in priority management on their own. It is usually necessary to specify the event and thus resulting internal status changes more precisely as a function of the environment in which it occurs (constraints) before using it in the priority optimisation system. This can be done using deductive rules having the following simplified base structure:

$$\textit{combination}\ (p_1, p_2) \leftarrow \textit{status}\ (p_1, p_2, p_3),\ \textit{constraint}\ (p_3, p_4).$$

The active constituent is needed to execute the responses of the priority optimisation system (priority manipulations). These responses are to specific events. These events correspond with insertions or deletions of status information from the traffic-regulation or monitoring system or out of it deductively defined combinations. The active rules have the following simplified base structure:

ON insert (combination (p_1, p_2)*)*	[Event]
IF status (p_2, p_3)	[Condition]
DO change standard priority (p_1, p_3).	[Action]

By support of the normative constituent it is possible to define general fixed regulations of relevance to the according of priorities for regulation in the relevant network (e.g. statutory precedence provisions, waiting time instructions etc.) with the aid of normative rules or integrity constraints. It is essential that the latter be adhered to at all times regardless of any optimum levels specified under the performance regime. A normative rule contains details to this end of parameters covered by the normative process together with a range of values cleared for the priority with the appropriate normative credentials.

Clearly, it would also in principle be possible to implement a priority optimisation system with the aid of a suitable declarative programming language and hence do without these DBMS constituents. The advantage of the approach delineated here, however, is that the input for implementation and, above all, for future servicing can be significantly reduced given an altered rules scenario. Instead of having to laboriously process the source code, it is merely necessary to adapt the rules data to the database interface.

7 Conclusion

This paper sets out the development of a rules-based system of traffic regulation for railway operations involving priorities optimised on the basis of rules. The priority optimisation system delineated serves an interfacing regulation system as a decision support tool in the event of conflicts arising and operates in detachment from the actual conflict resolution process. With the aid of the priorities prescribed, the interfacing regulation system is able to independently identify potential forms of concrete conflict resolution and put them into practice. It is possible using an affiliated monitoring system to rate regulated traffic on the basis of a performance regime. This provides the wherewithal for further optimising priorities. Recourse can be had to existing deductive, active and normative constituents of a database management system for the core processes of managing and optimising priorities.

The approach portrayed pursues the goal of combining the benefits of optimisation and heuristic procedures for train service regulation into an all-embracing concept. The configurational concept presented here is currently being implemented and tested in prototype form by the Institute of Transport Science at RWTH Aachen University. Further findings are to be anticipated soon therefore.

References

[1] Pferdmenges, S. & Schaefer, H., Automatische Konflikterkennung und wissensbasierte Konfliktlösung in der Streckendisposition, *Signal + Draht*, 87, pp. 174-177, 1995.
[2] Martin, U., Verfahren zur Bewertung von Zug- und Rangierfahrten bei der Disposition, *Eisenbahntechnische Rundschau*, 45, pp. 582-583, 1996.
[3] Oh, S. M., Hong, S. H. & Choi, I. C., Railway conflict detection and resolution in the Korean railway system, *Computers in Railways IX*, pp. 675-684, 2004.

[4] Kosemund, M., *Beitrag zum Einsatz von Prozeßrechnern für die optimale Zuglaufregelung auf Eisenbahnstrecken*, Dissertation, HfV "Friedrich List", Dresden, Fakultät für Technik und Naturwissenschaften, 1975.
[5] Jacobs, J., Reducing delays by means of computer-aided 'on-the-spot' rescheduling, *Computers in Railways IX*, pp. 603-612, 2004.
[6] Wunderlich, W. & Kockelkorn, E., Bessere Disposition durch schnellere Information, *Eisenbahntechnische Rundschau*, 24, pp. 283-294, 1975.
[7] Ullman, J. D., *Principles of Database and Knowledge-Base Systems*, Computer Science Press, 1988/89.
[8] Bry, F., Manthey, R. & Schütz, H., Deduktive Datenbanken, *Künstliche Intelligenz*, 3/96, pp. 17-23, 1996.
[9] Widom, J. & Ceri, S., Introduction to Active Database Systems (Chapter 1), *Active Database Systems: Triggers and Rules For Advanced Database Processing*, Morgan Kaufmann, pp. 1-41, 1996.