Computers & Operations Research 41 (2014) 150–166

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Determining operations affected by delay in predictive train timetables

Robert Burdett, Erhan Kozan*

Decision Sciences Discipline, Queensland University of Technology, Australia

ARTICLE INFO

Available online 24 August 2013

Keywords:
Sensitivity analysis
Robustness
Train scheduling
Job shops
Topological ordering
Transitive closure

ABSTRACT

Constructing train schedules is vital in railways. This complex and time consuming task is however made more difficult by additional requirements to make train schedules robust to delays and other disruptions. For a timetable to be regarded as robust, it should be insensitive to delays of a specified level and its performance with respect to a given metric, should be within given tolerances. In other words the effect of delays should be identifiable and should be shown to be minimal. To this end, a sensitivity analysis is proposed that identifies affected operations. More specifically a sensitivity analysis for determining what operation delays cause each operation to be affected is proposed. The information provided by this analysis gives another measure of timetable robustness and also provides control information that can be used when delays occur in practice. Several algorithms are proposed to identify this information and they utilise a disjunctive graph model of train operations. Upon completion the sets of affected operations can also be used to define the impact of all delays without further disjunctive graph evaluations.

1. Introduction

A train timetable is a plan of all train movements that are supposed to occur in a railway system over a given period of time. Unfortunately there is no way of knowing beforehand what the journey times will be on any given day with absolute certainty. The sectional running time for example is a complex function of many variables. These include the state of the section, the gradient and curvature, the locomotive type, the train driver, the weight of the train, the weather and so forth. Therefore it is impossible for some specified sectional running time to be exactly and repeatedly achieved. Consequently actual events may deviate quite considerably from the timetable. The timetable is therefore only a prediction of what will happen. In this environment the creation of a robust schedule is of considerable importance, if not a necessity.

A predictive timetable may be regarded as robust if for a given performance criteria the schedules performance can be shown to be insensitive to delays of a specified level. In other words, robustness is a measure of a timetables tolerance to delays of a prescribed level (see [1]). Furthermore a timetable that is insensitive to multiply occurring delays may be regarded as more robust than one that is insensitive to single delays.

Methods for quantifying the “sensitivity” of a timetable to single delays, was developed in [2]. The effect of sectional running time deviations and additional dwell (i.e. stopping) time in particular were quantified for three pertinent objective criteria, namely makespan, total train delay and total time window violation which was used to measure schedule nonadherence. The outcome was a complete “profile” of performance for different levels of expected delay. The profile is essentially a function and there is a unique one for each considered objective criterion. The function shows when and how the objective function value increases, decreases, or remains static. It also signifies if and when the timetable becomes infeasible. This information can be used as part of a “proactive scheduling approach” to either alter the predictive timetable in advance or define suitable courses of action for specific “bad behaviour”. In other words it is used to determine whether there is a need for preventative or corrective action.

In this paper the identification of those operations whose start time is postponed as a consequence of a “forced” delay in another operation is considered. Developing efficient algorithms for this task is not trivial and poses significant challenges. Which operations have their exit time postponed is not considered because it is known that any operation whose entry time is postponed as a result of a delay in another operation will have its exit time postponed too, unless that operation had been “blocked” for a period of time greater than the delay.

Several sources of delay are addressed, namely sectional running time (srt) deviations, additional dwell (stopping time) and additional blocking. In the srt deviation case, the effect of both increases and decreases is considered. A significant feature of our approach is that all delays that have an effect are identified. This allows the immediate effect of delay and the effect of prescribed levels of delay to be obtained. For each operation in the timetable, the outcome is a list of affected operations. A value of delay that causes the effect also accompanies each operation in the list.

The determination of affected operations provides another component that can be added to the sensitivity analysis of [2].

---

* Corresponding author.
E-mail addresses: r.burdett@qut.edu.au (R. Burdett), e.kozan@qut.edu.au (E. Kozan).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.08.011

It is also the basis for a separate sensitivity analysis. The determination of affected operations is independent of the timetable performance (objective) criteria, which means that the analysis is only performed once. Obviously the effect on different objective criterion will be different but the list of affected operations will not change.

The number of affected operations, for a particular value of delay, signifies the extent of that delay, and provides usable "control" information. However this information does not always provide a direct and concrete means of judging the timetables overall performance, in the face of delays, nor does it provide a means of comparing the delays to different operations. For example, is a delay that immediately affects ten operations better or worse than a delay to another operation that immediately affects five operations? Similarly is an operation delay of one minute which affects ten operations better or worse than a delay of two minutes to another operation that affects four operations? This information is applicable for judging a performance criterion based upon the disturbance to other operations. It is also very useful in the later stage of the sensitivity analysis of the previous paper which is to refine the original predictive timetable to make it more tolerant (insensitive) to delays, and hence more robust.

The affected operation information could have been partially provided by the sensitivity analysis of Burdett and Kozan [2]. The focus of those algorithms however was not the identification of affected operations. Consequently they would be very computationally inefficient at this task. Therefore this aspect is focused upon in this paper.

The hybrid job shop scheduling framework of Burdett and Kozan [3,4] was used to characterise train timetables and all techniques in this paper are based upon it. That scheduling approach has also been extended in [5–7] so that additional trains can be inserted into existing timetables, trains can be scheduled on parallel lines with crossover points, and multiple overtaking conflicts are removed. Another noteworthy approach for inserting trains into existing timetables is [8]. The train scheduling model and approach of Liu and Kozan [9,10] could also be used as the basis of our approach and is a source of future research. Train scheduling can also be used for capacity level identification in railways. A review of capacity determination methods can be found in [11]. A critical aspect of our train scheduling approaches is the representation of trains as jobs, sections as machines, and train movements across sections and other section occupancies as operations. Index i, j and k are used to signify job (i.e. train), machine (i.e. section) and stage respectively. Each job $i \in I$ has K_i operations (i.e. stages). The kth operation of job i is denoted as $o_{i,k}$ and has a planned sectional running time of $p_{i,k}$ and a dwell time of $\delta_{i,k}$ on machine $m_{i,k}$. Each job that enters a machine for processing has an entry and exit time denoted by $entry_{i,k}$ and $exit_{i,k}$ respectively. Sectional running times and dwell times of trains contribute to the processing time of train operations as does the length of the train by way of a time lag denoted by $lag_{i,k}$. The extent of this lag primarily depends on the length of the train and the speed it is travelling at when it departs the section. Part of this value may also include planned dwell time if the length of the train exceeds the length of some sections of rail. Passing loops that separate adjacent sections of rail and allow trains to pass each other are represented as capacitated buffers and may contain more than one train. The timetable is represented by machine sequences σ_j. The operation in the kth position of sequence j is hence $\sigma_{j,k}$. The schedule is therefore a temporal realisation of the sequencing and is obtained by evaluating a disjunctive graph using a nondelay scheduling policy. The nondelay assumption prepares for unexpected events and ensures that trains are scheduled as early as possible and are not restricted from entering a section if it is feasible to do so (i.e. unforced idle time is not allowed).

In this paper we have developed a generic and powerful approach to identify the effect of all possible delays and a way of using this information to quantify the robustness and sensitivity of a train timetable. It is interesting to note that in other research, what to do in the event of a disruption has been considered (see [12–17]). For example how can a schedule be refined and improved by rescheduling or rerouting trains? A "real-time" optimisation model for resolving disruptions in a train schedule was developed by Walker et al. [12]. Refinements of a companion crew roster are also considered simultaneously. A branch and bound and column and constraint generation approach was utilised. A related train driver recovery problem was also considered in [13]. A solution method based on solving the LP relaxation of the set partitioning problem with a dynamic column generation approach was proposed. An advanced tabu search heuristic was proposed by Corman et al. [14] to adapt a timetable to delays and other unpredictable events occurring in real-time. Trains are rerouted and rescheduled in that approach.

A review of sensitivity analysis and robustness related research and associated limitations can be found in [1] and is not repeated here. This work is pertinent to the approach proposed in this paper.

Since [1] was published, we have observed a number of new papers that consider robust train scheduling, i.e. [18–20]. In [18] techniques to create delay resistant periodic train timetables were developed. They simulated delays and solved a corresponding delay management problem. In [19], an overview of the field was provided. Robust train scheduling has also been considered previously by Fischetti et al. [21], and Kroon et al. [22].

Transitive closure which provides a complete reachability analysis of a directed graph is an integral part of the procedures and theory of this paper. In recent years Koubkova and Koubek [23] presented a new algorithm with expected time complexity $\Theta(n^2)$ for constructing the transitive closure of an acyclic graph. It exploits the topological ordering of the directed acyclic graph. The algorithm is compared to other leading algorithms. A survey of dynamic algorithms has been provided by Demetrescu and Italiano [24] for path problems on general directed graphs. Two fundamental problems were primarily considered, that of dynamic transitive closure and dynamic shortest paths. Dynamic transitive closure approaches have increased in recent years and there has been a resurgence of interest in this problem. An algorithm is fully dynamic if both insertions and deletions are handled and partially dynamic if only one type of update is performed. Dynamic transitive closure and dynamic shortest path problems were also considered by Baswana et al. [25]; they presented improved "decremental" algorithms.

The format of the paper is as follows. In Section 2 definitions and preliminary theory is presented. In Section 3 a simple approach to determine the effect of a specific delay (or multiple delays) is first introduced and is used to validate later more efficient and generic algorithms. The immediate effect of operation delay is then addressed. Algorithms that identify the set of immediately effected operations for each operation in the timetable are developed. In Section 5 a more general analysis is proposed and determines when an operation is affected by a delay in another. In fact all possible effects and delays are identified from this analysis. This analysis also identifies immediately affected operations and makes the algorithms of the previous section redundant to some extent. A greater computational effort however is required by the general algorithms of Section 5. In Section 6 equations for explicitly calculating the impact of delays are proposed and utilise the information found in the sets of affected operations. In Section 7 the algorithms have been applied to two case studies and the results have been reported. Conclusions and future research directions are lastly provided.