PRIORITY BASED TECHNIQUE FOR RESCHEDULING TRAINS

Z. Alwadood¹, A. Shuib¹ and N. A. Hamid²,*

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
²Arshad Ayub Graduate Business School, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Published online: 17 October 2017

ABSTRACT

This paper presents a novel approach to solve a railway rescheduling problem using a Mixed Integer Goal Programming (MIGP) model. The model is developed to find a new rescheduling timetable based on train priorities whenever a service disruption occurs. The objectives are to minimize the total delay time of all trains in the rail line and maximize the service reliability. The experiments were done on Malaysian double track railway system and the model was solved using a heuristics approach. The result shows that train priority has a great influence on the total delay time and service reliability. This research has proposed a more advanced and practical model as it is able to produce the provisional timetable in short computing time and the solution generated satisfies all the restrictions posed by the rail operator.

Keywords: rescheduling; mathematical modelling; service disruptions; priority.

Author Correspondence, e-mail: zuraida794@salam.uitm.edu.my
doi: http://dx.doi.org/10.4314/jfas.v9i5s.32
1. INTRODUCTION

For many countries, rail transportation is a major service provider for both passenger and freight transportation. Occasionally, there occur unexpected events which lead to the inability of the train to run within their scheduled timetable. Such events are termed as disruptions. Disruption is defined as an event or a series of events that renders the planned schedules for aircraft, crew or other timetable to be infeasible [1]. Service disruptions in most cases often lead to non-adherence to scheduled timetable and thereon leading to service delays. Rail service disruptions do not only occur in less developed countries but also in highly developed countries, where the services are advanced and equipped with sophisticated technology and infrastructures.

To deal with these problems, this research aims to develop an innovative optimization model which is able to manage railway traffic in real-time whenever disruption occurs. Specifically, it is concerned with developing a Mixed Integer Goal Programming (MIGP) model for solving disruption-related railway rescheduling problem. The model has two goals namely to minimize total delay time of trains in the whole network and to maximize trains service reliability. Heuristic computational methods using MATLAB Solver is implemented to obtain the optimal solution. The model is intended to produce quantifiable quick solution to the real-time rescheduling problem and to offer service recovery strategy which can help the railway services to maintain an efficient and reliable mode of transportation.

2. RELATED WORKS

In [2] has reported that French Railway commuter faced service disruptions in May 2016 due to the rail workers’ strike over the unsatisfied management plans on working conditions. The operator had to cancel half of its high speed train service. Barber’s findings from the National Rail Passenger Survey for Autumn 2015 on 300,000 train travellers reported that the overall commuter rail passenger satisfaction for Thameslink Train was only 73% of which the complaints were mainly due to punctuality and overcrowding [3].

These cases of rail service disruption have caused a wide range of mathematical model being formulated to focus on the train rescheduling problem in managing train service delays. Several model approaches for rescheduling have been proposed. A Mixed Integer
Programming (MIP) model which was established by [4] includes disruption and conflicts-resolving constraints in the model itself. The novelty of the method ensures that only disrupted trains will be rescheduled, leaving alone the unaffected trains. This is done by partitioning train sets into conflicting and non-conflicting trains by means of linear constraints.

An Integer Programming (IP) model for a partial blockage and a complete blockage which aimed at maximizing the service level to the train passengers was proposed by [5]. The model determines which trains need to be cancelled and it presents the disposition timetable for the neighbouring trains. They defined delay for each train sub series and used exact time between two operated trains in the same direction instead of the maximum time. IP-based railway capacity management model was also presented by [6] with the objective of minimizing the sum of total travel time and delay for trains, utilizing rail track capacity through efficient routing and scheduling. Other IP models are established by [7-8] which determine a timetable for a set of trains, subject to some operational constraints within the track capacities, while minimizing the number of cancelled and delayed trains.

In [9] have formulated a new technique to obtain an efficient decomposition mechanism through modelling track capacities as side constraints using IP model. These side constraints are then dualised through a new Lagrangian relaxation solution technique. The special feature of the solution method is the ability of the dual solution to be transformed to feasible solutions through a set of priority rules. A model predictive control approach which has been proposed by [10] attempts to reschedule trains by a discrete-time control. In this approach, a set of alternative blocking-stairways was used as the basis in each rescheduling step. This was followed by several planning steps which were linked to each other by different temporal scopes. The concept of bi-level multi-objective formulation sees that three criteria are considered separately in the first level. They were then aggregated into one objective function as a weighted sum.

In [11] did some modification on some constraints to form groups of blocks, known as a section. As a section consists of more than a single block, therefore, the constraints need to be amended to accommodate spacing area between blocks. Due to the large number of variables and constraints, a Constraint Programming (CP) model was used as an alternative to the
former MIP method. A CP engine performs a set of logical inferences to reduce the available options for the remaining variables’ domains which allows the problem to have less variables and constraints as compared to the MIP.

In [12] have formulated a Goal Programming (GP) model in the attempt to satisfy two different objective functions. The first goal of minimizing the total service delay is for the use of current practice while the second objective of minimizing the total service cost is valued for future planning. They used a multi-criteria decision analysis (MCDA) and discovered a relationship between certain disturbance characteristics and the ability to find good solutions within short time.

3. THE OPTIMIZATION MODEL

The model presented in this research is an extension of [12] MIP model. It is chosen as the reference model in this study due to the strong formulation of rescheduling railway traffic problem in the concepts of multi train types, multi-track lines, different blocking systems and highly interacting traffic. In this paper, the modified MIGP model for train rescheduling uses basically the same variables and parameters and focuses on minimizing the service delay time and maximizing service reliability. The model incorporates a new objective function which optimizes the service reliability. Besides the additional new parameters, other new features include the different number of parallel tracks at stations, new block-oriented headway restrictions posed by signalling switches and a real-time schedule modification. These new characteristics represent the novelties of the modified mathematical model. The optimization model is presented as follows:

Indices

\begin{itemize}
  \item[i] \text{Index for trains}
  \item[j] \text{Index for segments}
  \item[k] \text{Index for events}
  \item[t] \text{Index for tracks}
\end{itemize}

Model sets

\begin{itemize}
  \item[B] \text{Set of segments}
\end{itemize}
T : Set of trains
E : Set of events
\( K_{fi} \) : Ordered set of train i events having \( f_i \) events
\( L_{gj} \) : Ordered set of segment j events having \( g_j \) events
\( \mathcal{R} \) : Set of parallel tracks

Model parameters

\( d_{ik} \) : Scheduled departure time of event k
\( a_{ik} \) : Scheduled arrival time of event k
\( \Delta_k \) : Fixed minimum running time for event k
\( l_j \) : Location of segment j

\[
l_j = \begin{cases} 1, & \text{if } j \text{ is a station segment} \\ 0, & \text{otherwise} \end{cases}
\]

\( p_i \) : Priority of train i

\[
p_i = \begin{cases} 1, & \text{if } i \text{ is a normal train} \\ 0, & \text{if } i \text{ is a high priority train} \end{cases}
\]

\( \delta_j \) : Average time taken to pass a kilometre of a single station segment j

\( \eta_j \) : Average time taken to pass a kilometre of a single non-station segment j

\( \alpha \) : The minimum safety distance between two consecutive trains

M : A large positive constant

Model decision variables

\( d_{ik}^{\text{res}} \) : Rescheduled departure time of the event k
\( a_{ik}^{\text{res}} \) : Rescheduled arrival time of the event k

\( z_{ni} \) : Total train service delay time for all trains, up to the final train event \( n_i \)

Occupancy of track t to the event k,

\[
r_k^j = \begin{cases} 1, & \text{if event } k \text{ uses track } t, \text{ where } k \in L_j, t \in \mathcal{R}_j, j \in B \\ 0, & \text{otherwise} \end{cases}
\]
Dispatch routine for events $k$ and the proceeding event $\hat{k}$

$$G_{ik} : \quad G_{ik} = \begin{cases} 1, & \text{if event } k \text{ occurs before event } \hat{k}, \text{ where } k \in L_j, j \in B, k < \hat{k} \\ 0, & \text{otherwise} \end{cases}$$

Pre-set time windows for train $i$ in arriving to the final destination,

$$\tau_i : \quad \tau_i = \begin{cases} 1, & \text{if train } i \text{ arrives with a delay less than } w_i \text{ time units, where } i \in T \\ 0, & \text{otherwise} \end{cases}$$

Objective functions

\begin{align*}
\text{Minimize} \quad Z &= \sum_{i \in T} z_{n_i} \\
\text{Maximize} \quad R &= \frac{\sum_{i \in T} \tau_i}{\sum_{i \in T} T_i}
\end{align*}

Subject to:

\begin{align*}
a_k^a &\leq d_{k+1}^a, \quad k \in K_j, i \in T, k \neq n_i \\
a_k^a &\geq d_k^a + \Delta_k, \quad k \in E, k \neq \psi \\
a_k^* &\geq d_k^*, k \in \psi, \quad k \in E \\
a_k^a - d_{k+1}^a &= 0, \quad k \in K_j, i \in T, k \neq n_i \\
a_k^a - d_k^a &= \Delta_k, \quad k \in E \\
d_k^a - d_k^a &\geq 0, \quad k \in E \\
d_k^a - d_k^x &= 0, \quad k \in E, d_k^x > 0 \\
a_k^a - a_k^x &= 0, \quad k \in E, a_k^x > 0 \\
a_k^a - a_k^x &\leq z_k, \quad k \in E \\
\sum_{i \in T} r_k^i &= 1, \quad k \in L_j, j \in B \\
r_k^i + r_k^i - 1 &\leq G_{ik} + G_{ik}^x
\end{align*}
The objective function (1) intends to determine the minimum total delay, while the objective function (2) aims to find the service reliability when trains arrive at the final destination. Constraint (3) translates this condition and indicates that a successor of a train event must wait until its predecessor has been completed before it can start. The minimum running time for each train event is guaranteed by Constraint (4). It implies that a normal train must complete its running time $\Delta_k$ before the train event ends. Constraint (5) is introduced to represent the dynamic feature of the real time scheduling. It forces the disrupted event to resume journey once it recovered.

Constraint (6) guarantees that each of the train events must be directly succeeded by the next one, as far as the original schedule is concerned. In addition to this, Constraint (7) specifies that trains should strictly depart and arrive according to the planned scheduled as soon as it has completed the minimum running time $\Delta_k$. Constraint (8) indicates that the rescheduled departure time should never be earlier than the original time scheduled. The earliest it can depart is at least equivalent to the respective scheduled departure time of that event at station. Constraint (9) and Constraint (10) ensure that events that have already started and ended before disruption occurs must follow the original timetable. Constraint (11) defines the total
delay of all trains as the deviation between the rescheduled and the original arrival times. Constraint (12) guarantees a single use of track by one train at a time. To accommodate the different number of parallel tracks at different segment, Constraint (13) ensures that these concurrent events at a segment must not exceed the track capacity. Constraint (14) checks the sequence between an event $k$ and its proceeding event $\hat{k}$ so as to ensure that either one of the binary variables $G_{\hat{k} k}$ or $G_{k \hat{k}}$ takes value of ‘1’. Constraint (15) and Constraint (16) impose a restriction for the minimum headway between two consecutive trains using the same piece of track. If event $k$ is followed by event $\hat{k}$, then Constraint (15) and Constraint (16) are active. Otherwise, Constraint (17) and Constraint (18) are operative. Constraint (19) up to Constraint (22) define the domain of the decision variables.

4. THE COMPUTATIONAL EXPERIMENT

The developed mathematical programming model intends to solve railway traffic conflict as fast as possible so as to assist the train dispatcher in the rescheduling process. The MIGP model developed is a complex combinatorial problem that would consume a large amount of computation time and requires a huge memory space to produce solution with inconsistent accuracy. Solution to the traffic conflicts may involve a number of stations, departure and arrival times, direction of routes and location of conflicts especially when the disruption involves neighbouring trains. Therefore, it is difficult to attain optimal solutions in large-scale and complex instances, or the number of feasible solutions may be very large depending on the chosen solution. Some approaches may be able to yield optimal solutions, however the computing complexities normally increased exponentially and take a long computing time. As the matrix involves thousand of rows and columns, it is very difficult to solve. Finding the optimal track to use and keeping a safe distance between trains are NP-Hard problems, while the generation of the new rescheduling timetable is NP-Complete [13]. Such a complex combinatorial problem would consume a large amount of computation time and requires a huge memory space to produce solution with inconsistent accuracy. Due to these intractable problem instances, heuristic methodis used in this research experiment. As heuristicis able to solve large and complex problems, they are considered very powerful tools in solving
optimization problems.

A novel heuristic technique proposed by this research is called Headway and Order Scheme (HOS) which is comprehensively introduced to solve the rescheduling model. HOS is a solution method which considers the headway restriction and the sequence order of conflicting trains as its main feature in the method to solve the MIGP model. The main objective of the heuristics is to ensure that a solution close to the optimal feasible solution is produced within a reasonable computing time. To reduce the computational complexity, the MIGP model is reduced to a relaxed model by keeping the binary value \( G_k \) and \( G_k^\hat{} \) fixed. As a result of this action, the relaxed MIGP model has less constraints and solution process can be accelerated. An experiment utilizing the novel HOS was done on a sample data.

The test data is composed of 23 track segments including 10 stations taken from a railway network of a local railway company. The experimental analysis was carried out by considering few incident scenarios involving disruptions on track segments. For all cases, a partial blockage was considered in which only one track of a segment is blocked. The number of active tracks for Segment 1 up to Segment 21 is two, while Segment 22 and Segment 23 have four active tracks. It was specified beforehand that all trains going south and north are using Track 1 and Track 2 respectively. Any additional tracks available can be used when needed. For partial blockage which only one single track is left, any two consecutive trains using the single piece of track need to adhere to the predetermined minimum headway distance.

There are two types of train in the experiment which are commuter trains and the fast train. Suppose a commuter train experiences a breakdown and it resumes journey once it has recovered. In this paper, this train is termed as prime train. In rescheduling practices, the rail operator has established its priority settings in two aspects. First, in any conflict circumstances between types of train, a commuter train must always give priority to the fast train. The fast train is set to run according to the timetable as long as there is at least one track available. It is assumed that the fast train or the high priority train does not experience any service delay or break down. Second, in case when two commuter trains of the same direction are separated by less than 5 minutes, then the prime train has to be given priority over the
other normal-scheduled commuter train (NST).
There are nine commuter trains running in both directions including one high priority train heading south. Each train is assumed to be able to fit on any track and a maximum of six-car train is assumed. The location of all trains in the network is known at all times. For simplicity, the speed of trains is assumed constant and the dwell time of the trains at stations is embedded in the event duration. In addition, a minimum headway distance of 4 minutes is set at non-station segments and 3 minutes at station segments. The term service reliabilities used in this research denotes the percentage of train services which arrive within a time window of 5-minutes. The proposed model was tested on the instance when a high priority train was ahead of the disruption site. This computational experiment involved the utilization of the HOS approach to examine the result when the prime train is to give priority to the fast train. There are a total of 2018 to 2020 number of constraints generated from a list of 124 events and 1145 decision variables altogether. The time horizon is limited to 42 minutes and incident cases with disruption duration of 5 to 15 minutes were created to capture the rescheduling scenario. The cases were randomly selected with the aim to get solutions, which satisfy the model constraints. The mathematical model is solved using MATLAB R2014a. The computational tests were run on a 2.20GHz Intel (R) Core (TM) i5 CPU with 4Gb RAM.

5. RESULTS AND DISCUSSION
The research intends to solve railway traffic conflict as close as possible to the actual rescheduling process, as practiced by the train dispatcher in a real railway system. The mathematical model is expected to produce prompt solution in terms of a provisional timetable with a minimum service delays when disruption takes place. The experiments are categorized according to some possible scenarios that take place due to service disruption. The rescheduling strategies are set according to the real practice of the rail company. The situation is the scenario when there is a fast train approaching the disruption site. The solution generated by the solver presents the new departure and arrival times, the total service delay with the corresponding service reliabilities, the track used and the sequence of the new events. The results are then examined to determine the relationship between the disruption duration and the total delay generated by the proposed MIGP model.
The experiment was done on two selected events namely k53 and k55, to analyse the total delay time that is generated when there is a high priority train involved in the rescheduling activity. The next priority is given to the recovered prime train which was first affected by the service disruption. The experiment was initiated by supposing an event k53, which was scheduled to be followed by a high priority train, experiences a disruption at a station segment, Segment 18. The rescheduling model was run based on the HOS heuristics method. When the model was run for a disruption of 5 minutes on event k53, the total delay time recorded is 17 minutes. This disruption was then increased by 1 minute and for each experiment the total delay time recorded does not show any difference up to the 12th minute. It recorded an increasing total delay time for disruptions beyond the 13th minute.

Table 1. The total delay generated for different duration of disruptions, giving priority to the fast train

<table>
<thead>
<tr>
<th>Disruption Duration (min)</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total delay (min) when the event k53 is disrupted</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Service Reliability (%)</td>
<td>88.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total delay (min) when the event k55 is disrupted</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>25</td>
<td>27</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Service Reliability (%)</td>
<td>88.9</td>
<td>77.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the total delay time generated for cases when the recovered prime train is prioritized over the NST, with the presence of the fast train. The disruption is set to 5 minutes up to 15 minutes. The leftmost column is the name of the events which have been rerun in this experiment. The top row of the table shows the length of disruption duration while the total delay times, in minutes, are recorded for each of the two events. The corresponding service reliabilities are stated for the respective total delay time. The total delay time started to show increment at the disruption length of 13 minutes and above.

The experiment setting is repeated on another event of the same attributes, k55, which shows similar trend when the initial disruption length of 5 minutes is increased by one minute each time. The total delay time generated is fixed to 23 minutes for each case but started to increase for the disruption length of 13 minutes and beyond.
Fig. 1 shows the total delay time resulted from the experiments done on the two selected events k53 and k55 with respect to the duration of service disruption. The horizontal lines indicate that the total delay time is kept constant for the first 12 minutes of disruption and they started to increase for longer disruption duration.

![Fig. 1. Total delay of trains with respect to the disruption duration](image)

The reason of the high amount of total delay is due to the existence of the fast train in the experiment. A fast train event is governed by the ‘no-wait’ Constraint (6) and Constraint (7), which guarantees that it must strictly adhere to the original timetable.

\[
a^R_k = d^R_i, \quad k \in K_{ij}, i \in T : k \neq n_i.
\]

\[
a^R_k = d^R_i + \Delta_k, \quad k \in E.
\]

In addition to this, the next priority is given to the recovered prime train. This implies that the prime train is allowed to resume its travelling right after the fast train is at a safe distance ahead. This is then followed by the NST as specified in the headway Constraint (15).

\[
(\alpha \delta_j + M)G_{ij} - d^R_k + a^R_k \leq M, k, \hat{k} \in L_j, j \in B, k < \hat{k}.
\]

This particular headway constraint takes effect as \( k < \hat{k} \), which implies the prime train event occurs before the NST event as scheduled in the timetable.
As the prime train event is blocking a track, there is only one piece of track available and both fast train event and NST event are competing for the resource. To deal with this conflict, the track capacity Constraint (13) enforces that the concurrent events at a segment must not exceed the track capacity.

\[
  r_i' + r_i' - 1 \leq G_{i\alpha} + G_{i\alpha}, k, l, j \in P, j \in B: k < \hat{k},
\]

The mathematical model guarantees that fast train event must be directly succeeded by the next one and it should strictly depart and arrive according to the planned schedule. Being restricted by all these constraints is the reason of the high total delay time recorded in the experiment. Even for a very short disruption of 5 minutes, the prime train is not able to resume journey because the priority has to be given to the high priority train.

Based on the incident cases that have been analysed, the MIGP model has successfully generated the rescheduling timetable in a short computation time. The total service delay of each affected train is attainable, together with the service reliabilities and the list of all departure and arrival times for each event. The new rescheduling timetable practically considers the aspects of track capacity and the minimum headway requirement, besides other model constraints. Ultimately, the fast train runs according to the original schedule and did not experience any delay in all cases. For each of the experiment done, the processing time recorded is only 36 seconds.

The model validation and verification are also carried out to ensure that the results obtained from the completed experiment resemble the true practice of the company. The research works are presented to the operational management of the rail company and a series of consultation with the experts is carried out. This ensures that the service delays generated fall within their operational tolerance and the service reliabilities obtained meet their standard.

6. CONCLUSION

This research has proposed a more advanced and practical model as the rail company does not have the systematic or structured mechanism to measure train service reliabilities. A comparison of the total delay time between the model output and the real output could not be made because the total delay time are not systematically recorded by the service operator in
its case of service disruptions. Therefore, the total delay time and the service reliabilities should be part of the evaluation in their operational services in the near future.

When the total service delay is kept at minimum level and the service reliability of the train service is satisfactorily high, then to some extent, this will give a significant impact on the overall level of efficiency of the train service. A higher level of efficiency of train service will strongly allow a higher degree of attraction amongst the potential riders of the train service on a particular route. The desire to use the service is even more if comparisons were to be made between their present mode of rail transportation and the available alternative forms of transportation such as the private vehicles or bus services. From the perspective of the rail operations, these findings are indeed important as there are savings in terms of the utilization of resources such as labor and maintenance cost. With lower operational cost, a higher profit can certainly be generated.

7. ACKNOWLEDGEMENTS

The work presented in this paper is financed by Universiti Teknologi MARA Malaysia. This research is also made possible with the support of the anonymous railway company who provides the research data. The authors gratefully acknowledge these supports.

8. REFERENCES


How to cite this article: