# Integrating of optimization and data mining techniques for high-speed train timetable design considering disturbances

Ting-Wu Ho [1]*, Te-Che Chen [1,2], Chien-Cheng Chou [1]

*[1] Department of Civil Engineering, National Central University, Taiwan*
*[2] Taiwan High Speed Rail Corporation, Taiwan*
** Corresponding author (ncuceit@gmail.com)*

**Purpose** The Taiwan high-speed railway (THSR) system plays an important role in maintaining efficient transportation of passengers around Taiwan. However, the control mechanisms of THSR and the traditional railway systems are quite different. Drivers on THSR-trains cannot control the cars by themselves; only the control center of THSR can give the commands, which are based on the train timetables and should be followed by the drivers to operate the cars. Moreover, when a disaster occurs, the control center needs to prepare a rescheduled timetable in accordance with current situations that drivers can follow. **Method** This study presents a methodology to establish a set of optimal operation rules which are tree-based rules for real-time train timetable control for the THSR-system. The rules can be used to determine the optimal real-time operation during disturbances. Steps of the proposed methodology involve: (i) building of train timetable optimization model, (ii) generation of optimal input-output patterns, and (iii) extraction of tree-based rules for designed scenarios using the decision-tree algorithm. **Results & Discussion** The model could generate a timetable result that was as good as a real timetable. This means it has potential as a simulation analysis for predicting the effect of disruptions on the timetable without doing the real experiment with train timetables during disruptive events.

***Keywords***: *timetable, optimization model, data mining, high speed rail*

## INTRODUCTION

Nowadays, railway transportation has become a good alternative in many countries as an efficient and economic public transportation mode. It plays an important role in the passenger and freight transportation market. The railway has grown by over 40% in both freight and passenger sectors over the past 10 years[1]. All railway companies try to provide good services in order to satisfy their customers. One way to realize this is by improving the quality of the train control process or scheduling so that the railway company could optimize these services as well.

In addition, the train timetable is the basis for performing the train operations. It contains information regarding the topology of the railway, train number and classification, arrival and departure times of trains at each station, arrival and departure paths, etc. More formally, the train scheduling problem is to find an optimal train timetable, subject to a number of operational and safety requirements.

Scheduling problem in this research was emphasized to High-speed rail (HSR) system; a type of passenger rail transport that operates significantly faster than the normal speed of rail traffic. Specific definitions by the European Union include 200 km/h (120 mph) for upgraded track and 250 km/h (160 mph) or faster for new track. Commonly, HSR system has different system with normal speed railway due to the safety issues. It uses lines without branches and minimizes the amount of stoppage time to keep the speed constant.

Managing circulation of trains during disturbances, including regular inspection, car cleaning times and turning back operations, has become important as the mitigation effort for minimizing consequences especially for high-speed system; which is known as a type of passenger rail transport that operates significantly faster than the normal speed of rail traffic and sensitive to disturbances. The Taiwan High Speed Rail (THSR) system already has cyclic patterns of daily train circulation, but these patterns have not been modeled yet especially during disturbances. Moreover, based on a review of the literature, only few researchers in the railway field have considered train circulation model, especially in high-speed rail systems, even though it is an important requirement and the arrangement of circulation time in a given timetable could be such a way that the timetable becomes maximally robust against stochastic disturbances[2].

The second issue is that railway systems are often characterized by high traffic density and heterogeneous traffic that is sensitive to disturbances, especially in high-speed railway system; thus, rescheduling activity for updating an existing production schedule in response to disruptions or other changes is needed[3]. The general problem is to decide when the trains get access to the tracks as well as if and where the trains should meet and overtake while ensuring that safety restrictions and other considerations are maintained[4]. In normal speed railway system, this rescheduling could be done manually by the traffic management. It monitors the network and traffic status via a screen and has a pencil and a paper-version of the time–distance diagram of the timetable to adjust in line with the re-scheduling plans. However, in high-speed system, as the name high-speed, predicting disturbance propagation and minimizing the consequences is a challenging problem because of the speed itself.

The benefits and challenges of the train scheduling method can contribute to improve the quality of the train control process. Thus, the optimization model that represents the real problem needs to be designed to support the train operator's creation of an optimal schedule in limited computation times. Moreover, predicting the impact of disturbances on the timetable also needs to be addressed in order to anticipate and mitigate the worst condition. Those two problems are being faced by the THSR Company today; therefore, the objective of this research is to solve them by doing three tasks as follows: 1) designing an optimal timetable using an optimization-based approach which capability to accommodate basic requirements (railway topology, traffic rules, user requirements) and train circulation requirements (turning back, regular inspection and car cleaning times); 2) checking the model using real data from THSR Company; and 3) analyzing the responses of model results to the disturbances using sensitivity analysis.

The real problem in the THSR system is its use of a contingency timetable to solve the timetabling problem under disturbances. It considers two conditions of disruption effects: (1) low-speed running operation at a certain location, and (2) line track or station blocked. To solve the first condition, THSR creates a timetable with speed restrictions and applies it to the train running time and headway calculations. Each speed would create different timetable results. Moreover, to solve the second problem (line track blocked), the THSR Company uses five types of blocked track possibilities and one type of bi-directional solution. This system generates the timetable according to a specified train running route. Each condition also has different timetable results. Nevertheless, as many contingency timetables can be prepared as planners can make in the THSR system. In fact, it still has a big problem, in that it is especially time consuming to maintain and select the appropriate contingency timetable for one disruption and create an optimal timetable during the disturbances. There are five thousand types of contingency timetables in the THSR Company, and the train operator should select the appropriate timetable within a limited time[5]. Furthermore, based on the data in the contingency timetable, THSR prefers to cancel many trains and operates only two trains per hour in many cases of disturbances. On the other hand, creating an optimal timetable, which means optimal journey time, is important since the THSR Company has to preserve the maximal profit during disturbances. In addition, in order to mitigate the impact of disturbances instead of cancelling many trains on their system, THSR needs a method for analyzing how disturbances propagate within the original timetable and which actions to decide to minimize the consequences. In the end, the train operator could predict the effects of disruptions on the timetable without doing real experiments.

## RELATED WORKS

Train scheduling problems have been studied by researchers over the years. They have been formulated using operation research (OR)-based techniques including mixed integer programming[6] and network optimization models[7]. Among the solution techniques developed to solve the problems were branch and bound[8], heuristics[9], and lagrangian relaxation[10]. In addition, researchers have proposed some methods to solve scheduling problems as the Job-Shop problem[11], program evaluation and review technique or PERT[12], and satisfaction constraint[13].

The strengths and weaknesses of each method depend on research objectives. Many researchers prefer to use a heuristic method to speed up the computation time and direct the search towards a feasible solution, but a heuristic needs more evaluation in practice due to the optimality issue of the results. There is no guarantee that the heuristic will generate an optimal solution in every test case[13].

The scheduling and rescheduling problem in this research is formulated as mixed-integer-programming (MIP), in which some of the variables are real-valued and some are integer-valued. There are two ways to solve MIP: exact solution, including branch and bound, and dynamic programming and approximation, including heuristic and lagrangian relaxation techniques. The most widely used method for solving MIP for exact solution is branch and bound, and no researchers have used dynamic programming as a solution technique, although it could obtain the optimal solution as well. Therefore, this research proposes the exact-solution algorithm, which is dynamic programming, to solve train-scheduling problems and get the optimal solution exactly.

Since the train-scheduling problem is being used to find an optimal train timetable, it has many rules to consider. Early work by Chiu[13] has considered six types of scheduling rules in railway systems: speed constraints, station occupancy, station entry and exit constraints, stopover, and line time constraints. In addition, railway topology, traffic rules, and user requirements have been considered in mathematical formulations modeled[6]. This research has contributed several additional constraints that arise in real-world timetabling applications. In particular, they address the following constraints: automatic and manual signaling, station capacities, maintenance operations, and periodic trains.

However, based on the literature review, only few researchers in the railway field have never considered train circulation, especially in high-speed rail systems, even though it is an important requirement. Therefore, a scheduling model which has the capability to accommodate not only basic requirements (railway topology, traffic rules, and user requirements) but also train circulation requirements needs to be formulated.

In this paper, we use an optimization method to solve the train-timetabling problem similar to the one presented before[6], but more complex, and discuss the problem of sensitivity analysis. Sensitivity analysis has been used in train timetabling fields, e.g., sensitivities of station spacing and sidings, showing its importance for a single-track line, or the influences of train speed on the performance of the line[8]. Castillo et al. also have analyzed sensitivity analysis in timetabling optimization of single-railway track line problems[8]. They applied sensitivity to add quality for solving methods, and it should be a routine complement to any optimization problem.

Based on the results of sensitivity analysis, Castillo, et al, concluded a corresponding among the number of stations, speeds, and total travel-times. Reducing the number of station would enhance speeds and travel-time among stations, thus total travel-time in line would reduce automatically. In the other hand, enhancing the number of station (intermediate) improves the performance of the line substantially, even though the dwell-time at station is considered.

This research also used sensitivity to investigate the maximum relative travel-times with respect to dwell-times. The results concluded that only some trains modify the optimal solution locally if the dwell-times of those trains are modified. Otherwise, no such local affect, that is, a small change in the corresponding dwell times does not modify the optimal relative travel time.

In short, the sensitivity analysis provides important information about critical resources and trains, which used to improve the line and, indirectly, the timetable design. For example, in many important decisions, such as the number of stations, speed restrictions, departure times and dwell-times modifies, could be derived from the sensitivity analysis results.

## PROBLEM DESCRIPTIONS

The proposed HSR system is about a 335.2 km intercity service line without branches along the western corridor of Taiwan. Railway topology of the THSR system consists of two connecting lines between southbound and northbound. It connects two major cities, Taipei and Kaoshiong with eight stations along the line as illustrated in figure 1. Each station has multi-tracks (at least two tracks) which are used as platforms, waiting time terminals, and free passes. In one day, the THSR Company provides 120 services with 29 trains running.

After knowing the THSR topology, we should understand the timetable components and rules in order to develop a robust mathematical model for the scheduling problem. A timetable contains information regarding railway topology (stations, tracks, distance among stations, traffic control, etc), and the schedules of the trains that use this topology (arrival and departure times of each train at stations, dwell-times, crossing times, car cleaning times, regular inspection times, and turning back operation). The timetabling design in this research is described as follows. Given the THSR railroad system and set of services, then the problem performs a timetable as well as a track assignment plan for these services.

The goals of the optimization model in this research are to let the trains depart as close to their target departure times as possible, at the same time minimizing the operation times of services. Since the operation times of each train as well as required headway between consecutive trains depend on the track assignment, railway topology and train circulation issues have to be considered simultaneously to obtain a realistic result which is close to the real timetable.

Therefore, the objective function in designing the timetable in this research is to minimize the total operation time for all train services subject to basic requirements (railway topology, traffic control, user requirements) and train circulation requirements (regular inspection, car cleaning, and turning back operation).

## MODEL DEVELOPMENT

A model was developed to analyze a particular problem and may represent the different parts of the system. The scheduling model in this research was developed based on the real problem in THSR system and also from some assumptions below:

- Optimal timetable in high-speed railway system means that the timetable could perform the effective journey time to the passenger. Therefore, the objective functions for scheduling model in this research is to minimize the total journey time for all trains in the system.
- The scheduling problem occurs in double lines railway system with multi tracks at each station. The different topology of railway will create different formulation of model.

- The upper bound (UB) value of this model is formulated as maximum operation speed of THSR (300 km/ hour) or 186 mph. Since the distance between Taipei to Zuoying is 335.2 km, it takes 67 minutes travel and 15 minutes as the start point time to rich optimal speed of THSR. Therefore, the reasonable UB for this HSR line is 82 minutes. Under disturbances, this value of UB would be increase, since the speed decrease.
- Train circulation includes car cleaning activity, regular inspection, and turning back operation. Only Taipei and Zuoying provides train circulation activities, because these two stations have multi tracks that are possible to perform it.

**The proposed mathematical model**

Suppose a railway system with r station, n trains going down and m trains going up. Minimizing the operation times for all trains means minimizing the journey times (arrival and departure times) for all trains going-down, initialized as i (1 to n) plus the journey time of trains going-up as j (1 to m) in every station (1 to r). Thus, the mathematical constraint for representing the objective function in this research is presented as Equation (1) below to minimize total operation time:

$$Z = \sum_{i=1}^{i=n}\left(T_iA_r - T_iD_1\right) + \sum_{j=1}^{j=m}\left(T_jA_1 - T_jD_r\right)$$

$\forall i=1\text{-}n$ and $\forall j=1\text{-}m$ (1)

The variables of this research are journey time, arrival and departure time of all trains with travel time, station time, headway, car cleaning time, regular inspection time and turning back operation, as parameters. Variables and parameters will be explained as constraints below:

- Travel time between two contiguous stations' constraints

Travel time constraints restrict minimum time to travel between two contiguous stations (k to k+1) for all trains going up initialized as i (1 to n) and trains going down initialized as j (1 to m).

$$T_iA_{k+1} - T_iD_k \ge time\ i_{k\rightarrow(k+1)} \forall i = 1-n\ and\ \forall k = 1-r \quad (2)$$

As represented by Equation (2), the arrival time for train i in the station k+1 minus departure time in the station k (origin station) should be greater or equal to the needed time for trains i to travel between two contiguous stations (k to k+1).

$$T_jA_k - T_jD_{k+1} \ge time\ j_{(k+1)\rightarrow k} \forall j = 1-m\ and\ \forall k = 1-r \quad (3)$$

The arrival time for train j in the station k minus departure time in station k+1 should be greater or equal to the needed time for trains j to travel between two contiguous stations (k+1 to k). This research uses minimum travel time between two contiguous stations, because different types of trains have different speeds and travel time would automatically differ.

- Dwell time constraints

As explained before, running time is calculated from departure times in the timetable minus dwell times. Therefore, station time for each train i and j at station k (1 to r) should be greater than departure time minus arrival time, as shown in Equations (4) and (5). This condition represents that the model uses maximum station time at each station, because not all trains will stop at every station.

$$T_iD_k - T_iA_k \le TS_{i_k} + CS_{i_k} \quad \forall i = 1-n\ and\ \forall k = 1-r \quad (4)$$

$$T_jD_k - T_jA_k \le TS_{j_k} + CS_{j_k} \quad \forall j = 1-m\ and\ \forall k = 1-r \quad (5)$$

- Headway constraint

Headway constraint restricts to the departure times differences between two consecutive trains in the same station. The headway time in this research is fixed to one value because we want to keep the time spacing between two trains exactly.

$$T_{i+1}D_k - T_iD_k = headway \quad \forall i = 1-n\ and\ \forall k = 1-r \quad (6)$$

$$T_{j+1}D_k - T_jD_k = headway \quad \forall i = 1-m\ and\ \forall k = 1-r \quad (7)$$

- Travel time in line constraints

Travel time in line determines the total travel time for one train to travel through a line southbound and northbound plus allowed time margin. Maximum travel time has been applied in the model; thus, the difference between arrival and departure time for one train in the same station should be less or equal to this travel time, as formulated in Equations (8) and (9) below:

$$T_iA_r - T_iD_1 \le \left(1 + \partial/100\right) \times time\ i_{1\rightarrow r} \quad (8)$$

$$T_jA_1 - T_jD_r \le \left(1 + \partial/100\right) \times time\ j_{r\rightarrow 1} \quad (9)$$

In the THSR system, allowed time margin was set to different numbers for different types of train. Therefore, this parameter would be a good input in sensitivity analysis to reveal the effects of changes in this parameter on objective value.

- Crossing time constraints

Although the THSR system has multiple tracks at stations, sometimes crossing operations become necessary for one train to allow another train to pass through the station. In this constraint, it is assumed that crossing time would be performed by two trains headed in different directions (southbound and northbound trains). Thus, the difference between arrival time for train i and departure time for train j at the same station k+1 (because the second train had already departed from its original station) should be less or equal to the upper bound time minus buffer time in the available segment.   is the decision variable for the availability of track in one segment. The value is 1 if there is a track available between station k to k+1 and 0 otherwise, as formulated in Equation (10) below:

$$T_iA_{k+1} - T_jD_{k+1} \le UB \times (1 - Y_{i-j;k\rightarrow(k+1)}), \forall i = 1-n\ \&\ \forall j = 1-m. (10)$$

- Train circulation constraints

Like many railway companies, THSR has a cyclic timetable in order to manage the resources comprising its infrastructure. In this research, it is assumed that the cycle time is 120 minutes (maximum travel time). Thus, if the headway time is set to 12 minutes, it means there are ten trains in the first cycle, and the next train (11th train) would be the same train as number one. It means the timetable at every two hours is the same pattern; the daily timetable is obtained by carrying out this pattern repeatedly. A train circulation model is needed to deal with these requirements. In the THSR system, train circulation takes 30 minutes maximum time, including car cleaning, regular car inspection, and turning back operations. Consequently, departure time for another train which will use the same track as those operations should be greater or equal to arrival time plus train circulation times as formulated in Equations (11) and (12) below:

$$T_{j\left(i+\frac{traveltime}{headway}\right)} D_r \ge T_i A_r + (InsT_r + ClT_r + TB_r) \qquad (11)$$

$$T_{i\left(j+\frac{traveltime}{headway}\right)} D_r \ge T_j A_1 + (InsT_1 + ClT_1 + TB_1) \qquad (12)$$

If there were an event that required long travel times and headway between two contiguous trains, then the train circulation pattern would change as well.

## Data collection

After the mathematical model for scheduling problem was formulated, a collection of data regarding scheduling requirements from the THSR Company began. Data requirements list are number of stations, number of trains in southbound and northbound line, headway time, allowed margin time, upper bound in line, station time, distance between stations, and the operation time regarding train circulation.

Primary data was collected from interviews with senior engineers in the THSR Company, and secondary data was gathered from THSR documents including the Equipment and Facilities Operations Manual, and existing timetables from THSR.

To accommodate data modification during model development process, a scheduling database has been built using Microsoft Office Access Database. This databases consists of stations ID, trains ID, train directions ID, train sequences ID, model parameters ID, and the table value for all parameters.

## Sensitivity analysis

Figure 1 shows the results of the THSR timetables using the proposed model. Figure 2 is the result of delay event for each train delay 9 minutes in High Speed Rail system, and it can be bring that the effect for train timetable delay. In this figure, the train ID's are the 10(time for morning), 30(time for noon) and 50(time for afternoon) for south bound train respectively. It can be indicate that the timetable delay time with three trains delay 8 minutes at BANCIAO station is more than the timetable delay with three trains delay 8 minutes at Tainan station. The start station of south bound train is TAIPEI station, and for south direction are BANCIAO, TAOYUAN, HSINCHU,

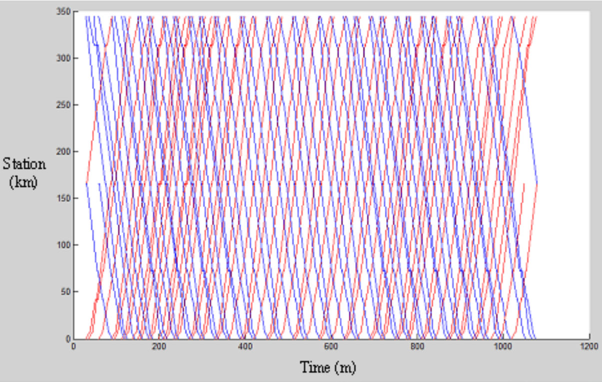

*Fig.1. Timetable diagram of the purposed model.*

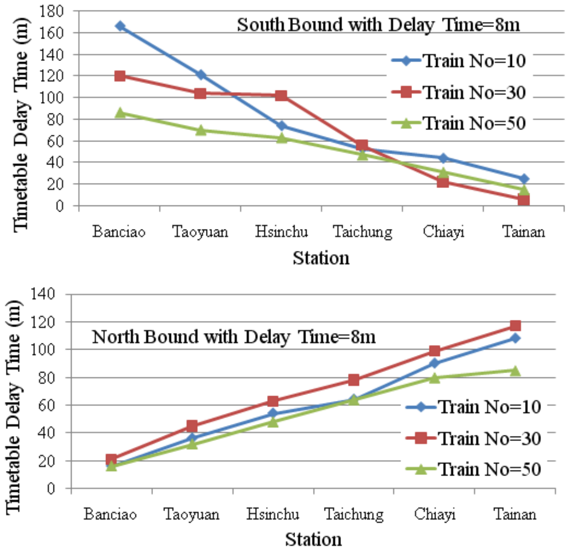

*Fig.2. Different trains vs. fixed delay time.*

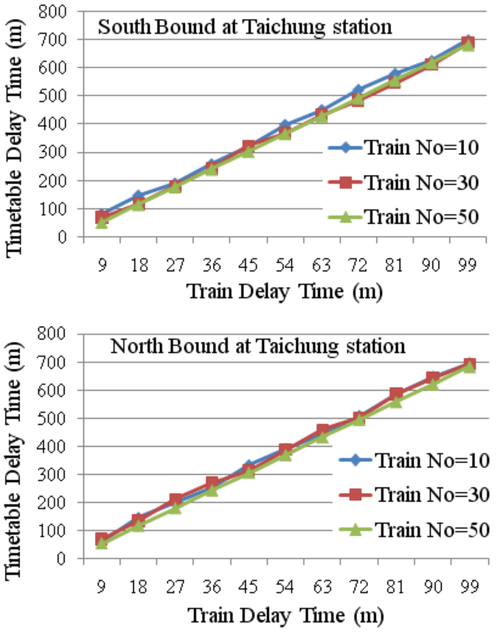

*Fig.3. Different trains vs. different delay time.*

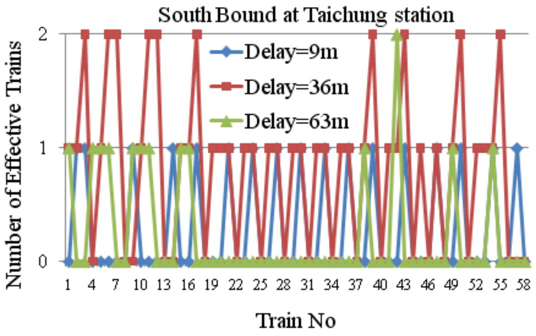

*Fig.4. Effective trains at the middle station.*

TAICHUNG, CHIAYI, TAINAN and ZUOYING. If a train is stop delay at some station, the other stations are also delay. Therefore, the train is delay near at first station, it can be effected the timetable delay time many more. For example, the south bound train ID is 50 for stop-all-station, which stop 1 minute at each station. If this train (train ID=50) stop 9 minutes at TAICHUNG station (means that is stop delay), the stop will also delay after TAICHUNG station (CHIAYI, Tainan and ZUOYING station). This train stop 9 minutes at Taichung that is be more over 8 minutes delay time. For the train timetable delay time, the total delay time is 8 m (delay at Taichung station) + 2×8×2 m (the departure and arrival delay time at CHIAYI and TAINAN station) + 8 m (the departure and arrival delay time at ZUOYING station) = 48 minutes.

Figure 3 is the result of train delay time effect to timetable delay time for south bound and north bound. The transverse axle is different train delay time for train ID=10, 30 and 50 at Taichung station, and the vertical axle is the timetable delay time. It can be indicate that if the train delay increasing, the timetable delay time also increase. Because the minutes of train delay time is positive depended for timetable delay time, train delay time many more for timetable delay time many more also.

ble delay time.

The Figure 4 is the result of the south bound train stop delay to the number of effective train. The transverse axle is train number for south bound, and vertical axle is the number of effective train. In this figure, the train number that is 1 to 20 and 35 to 58 occur stop delay at some station, the number of effective train are increasing. This result is depend on the time for morning and afternoon is rush hour in Taiwan, so the passengers travelling by train is many more. The other thing is that the train delay time increasing, and the number of effective train is increasing also. But for delay time is 63 minutes, the number of effective train is not increasing (it's decreasing), that is the delay time is so long for effect train at the back. In this figure that can explain that effect large of effective train number is by train stop delay about 30 minutes.

## CONCLUSIONS

This research developed an optimization model for designing timetables in high-speed railway systems that consider basic requirements as well as special requirements regarding train circulation, including car cleaning, regular inspection, and train turning back operations. The model could generate a good timetable result as good as a real timetable in a short computation time (0.10 seconds). Furthermore, the model could generate train circulation patterns as illustrated in the timetable diagram results.

## References

1. Railway Safety and Standard Board, *The Rail Industry:□ A Way Forward on Sustainable Development*, RSBB London, 2006, www.rssb.co.uk (viewed on 1st January 2012).
2. Kroon, L., Maróti, G., Helmrich, M.R., Vromans, M., and Dekker, R., "Stochastic improvement of cyclic railway timetables", *Transportation Research Part B*, 42(6), pp. 553–570, 2008.
3. Vieira, G.E., Herrmann, J.W., and Lin, E., "Rescheduling Manufacturing Systems: A Framework of Strategies, Policies, and Methods", *Journal of Scheduling*, Vol. 6, pp. 39-62, 2003.
4. Tornquist, J. and Persson, J.A., "N-tracked railway traffic re-scheduling during disturbances", *Transportation Research Part B*, 41(3), pp 342–362, 2007.
5. Taiwan Shinkansen Corporation, *Equipment and Facilities Operations Manual*, Vol. 3, 2007.
6. Salido, M. and Barber, F., "Mathematical Solutions for Solving Periodic Railway Transportation", *Mathematical Problems in Engineering*, 2009.
7. Martland, C.D., Zhu, Y., Lahrech, Y., and Sussman, J.M., "Risk and Train Control: A Framework for Analysis", *Transportation Research Record*, No. 1742, pp. 25-33, 2001.
8. Higgins, A., Kozan, E., and Ferreira. L., "Optimal scheduling of trains on a single line track", *Transportation Research Part B*, 30(2), pp.147-161, 1996.
9. Lee, Y. and Chen, C.Y., "A Heuristic for the Train Pathing and Timetabling Problem", *Transportation Research Part B*, 43(8-9), pp. 837-861, 2009.
10. Brännlund, U., Lindberg, P.O., Nõu, A., and Nilsson, J.E., "Railway Timetabling Using Lagrangian Relaxation", *Transportation Science*, 32(4), pp. 358-369, 1998.
11. Mladenovic, S. and Cangalovic, M., "Heuristic Approach to Train Rescheduling", *Yugoslav Journal of Operations Research*, Vol. 1, pp. 9-29, 2007.
12. Norio, T., Yoshiaki, T., Noriyuki, T., Chikara, H., and Kunimitsu, M., "Train Rescheduling Algorithm Which Minimize Passengers' Dissatisfaction", In M. Ali and F. Esposito (Eds.), IEA/AIE 2006, *LNAI* 3633, pp. 829-838, 2006.
13. Chiu, C.K., Chou, C.M., Lee, J.H.M., Leung, H.F., and Leung, Y.W., "A Constraint-Based Interactive

Train Rescheduling Tool", *Constraints*, Vol. 7, pp. 167–198, 2002.