Cost functions for mainline train operations and their application to timetable optimization

Aris Pavlides
Centre for Transport Studies
University College London
Gower Street, London WC1E 6BT
United Kingdom

Andy H. F. Chow
(corresponding author)
Centre for Transport Studies
University College London
Gower Street, London WC1E 6BT
United Kingdom
andy.chow@ucl.ac.uk
Tel: +44 207 679 2315, Fax: +44 207 679 3042

5168 words + 7 figures = 6918 words

August 1, 2015

ABSTRACT

This paper discusses a set of cost functions for timetabling mainline train services. Mainline train services are generally heterogeneous which consist of passenger and freight trains, slow and express services, domestic and international connections, etc. The feasibility of a timetable is subject to a number of factors including availability of trains and crew, infrastructure capacity, and travel demand. With the complex nature of modern railway systems and the heterogeneity of rail traffic, deriving satisfactory train service schedules for passengers, train operators, and infrastructure manager is always a challenge. The cost functions presented here are used as indicators for evaluating different performance associated with the corresponding timetable. The performances of interest include carbon, capacity, cost, and customer satisfaction. These four performance indicators are also identified as the '4C' criteria by the railway industry in Great Britain. These 4C criteria are set in order to address the need to improve customer satisfaction (e.g. by providing more punctual service) and operational capacity, while decreasing operational cost and carbon emission. We will also demonstrate the application of these cost functions into an optimization framework which derives optimal timetable for heterogeneous train services. The method is applied to Brighton Main Line in south-east England as a case study. The results reveal that overall performance of the railway systems can be achieved by re-scheduling and re-sequencing the train services through the optimization framework, while this may have to come at the expense of slow and local train services if the optimization is not properly formulated.

Keywords: train scheduling, capacity, punctuality, multi-objective optimization, genetic algorithm

INTRODUCTION

Railways are generally considered to be sustainable and green compared with other modes of transport. The significance of railway systems can be reflected by the amount of investment made around the globe, exemplified by the number of high speed railways (HSR) projects, in recent years. In the UK, we have seen recently a number of large investment programmes including the Crossrail project, focusing on improving reliability, journey times, and capacity of the London transport network (*1*). In Hong Kong, over the next decade the MTR Corporation will complete five new strategic rail extensions in Hong Kong and mainland China (*2*). Nevertheless, this sustained demand has been placing tremendous pressure on the railway infrastructures. Due to the tight fiscal, physical and environmental constraints, continuous construction of new tracks and purchase of rolling stocks will not be a sustainable solution. Consequently, we will have to rely on effective utilization of existing infrastructures in terms of timetabling the train services.

This study looks at the issue of timetabling mainline train services for improving the overall efficiency of the rail systems. Mainline train services generally refer to connections between cities as opposed to the local metro services, while timetabling is regarded as the process of deriving a feasible schedule for a given set of train lines over a specific route through specifying the associated arrival and departure times at each designated point. Mainline services are generally heterogeneous which consist of passenger and freight trains, slow and express services, domestic and international connections, etc. Moreover, the feasibility of a timetable is also subject to a number of external factors including availability of trains and crew, infrastructure capacity, and travel demand. As a consequent, deriving a satisfactory timetable for different stakeholders including passengers, train operators, and infrastructure manager is always a challenge.

In this paper, we address the mainline train timetabling problem by using an optimization approach. A prerequisite for formulating the optimization problem is to define a set of cost (or objective) functions that can reflect different performances of interest to different stakeholders. Following (*3*), we identify Carbon, Capacity, Cost, and Customer satisfaction as the four main aspects of interest. Chen and Roberts (*4*) and Roberts et al. (*5*) further categorize and discuss them according to the associated relevance to different stakeholders. These four aspects are regarded as the '4C' criteria by the railway industry in UK. The 4C criteria are set in order to address the need to improve customer satisfaction (e.g. by providing more punctual service) and operational capacity, while decreasing operational cost and carbon emission. We believe these four are also among the main objectives in railway sector in other countries apart from the UK. A set of cost functions is formulated to reflect the performance of each timetable in terms of the '4C'. The cost functions are then incorporated in a multi-objective optimization framework (see e.g. (*6, 7, 8, 9, 10*)) for deriving an optimal timetable following the setting of the cost functions or objectives.

The optimization framework is applied to the Brighton Main Line (BML) in south-east England as a case study. It is noted that different cost functions have different dimensions. This study adopts the monetary values suggested by the Department for Transport (*11*) in the UK to convert and integrate all costs into monetary units. However, the proposed approach is generic and will be applicable to different systems by revising the conversion factors according to different operators' or countries' needs. The results obtained from BML reveal that overall performance of the railway systems can be achieved by re-scheduling and re-sequencing the train services through the optimization framework, while this may have to come at the expense of slow and local train services if the optimization is not properly formulated.

The rest of the paper is organized as follows: the next section starts with introducing the specification of timetable in an optimization framework and its associated operational constraints. It is then followed by discussion of different performance indicators related to 4C and formulation of the associated cost functions. The cost functions are then used to formulated a multi-objective optimization problem for train timetabling. We also discuss the complexity of the timetabling problem and present a genetic algorithm (GA) based solution approach. The optimization framework is applied to a case study of Brighton Main Line which is used to demonstrate the proposed method and the results are discussed. Finally, the paper concludes with some final remarks and suggestion for future work.

SPECIFICATION OF TIMETABLE AND ASSOCIATED CONSTRAINTS

A timetable is typically incorporated through specifying the arrival $\tau_{n,s}$ and departure times $\sigma_{n,s}$ of each train n over a set of control points s (which can be a station, junction, etc.) along its service route. An example is shown in Figure 1 in which the horizontal and vertical axes represent the time and position along the train route respectively. Each line on the diagram represents a train run which is specified by a series of departure $\sigma_{n,s}$ and arrival times $\tau_{n,s}$ at station s for each train n as specified by the timetable. Given a set of $\sigma_{n,s}$ and $\tau_{n,s}$, we can derive the running time $T_{n,s}$ of each train n between station s and $s+1$ as

$$T_{n,s} = \tau_{n,s+1} - \sigma_{n,s}, \tag{1}$$

and also the dwell time $D_{n,s}$ of train n at station s

$$D_{n,s} = \sigma_{n,s} - \tau_{n,s}, \tag{2}$$

The setting of the variables $\sigma_{n,s}$ and $\tau_{n,s}$ will be subject to a set of operational constraints in practice. We first have the minimum section running time constraints to reflect the speed limit imposed on each track section $(s, s+1)$:

$$\tau_{n,s+1} \geq \sigma_{n,s} + \frac{\Delta_{s,s+1}}{v_n^*}, \tag{3}$$

where $\Delta_{s,s+1}$ is the distance between stations s and $s+1$, v_n^* is the maximum speed limit for train n traveling from station s toward s + 1. Moreover, we also have the minimum dwell time constraints which define the minimum time have to be spent by each train n at station s:

$$\sigma_{n,s} - \tau_{n,s} \geq d_{n,s}^*, \tag{4}$$

The minimum dwell time $d_{n,s}^*$ imposed here will typically be determined by a number of factors on the demand side such as demand level of passengers or freight for that specific train at that specific station, and/or the consideration of connectivity where it is necessary to ensure a long enough dwell time for passengers or goods to transfer from one train to another at the station or interchange (*12*).

Finally, to implement the signaling system, each track section is further disaggregated into a series of blocks. Under the current fixed block signaling systems in practice, each block can only accommodate up to one train at a time to ensure safe operations (see Figure 2). Referring to Figure 2, denote the arrival and departure times of train n at block j between station pair $(s, s+1)$

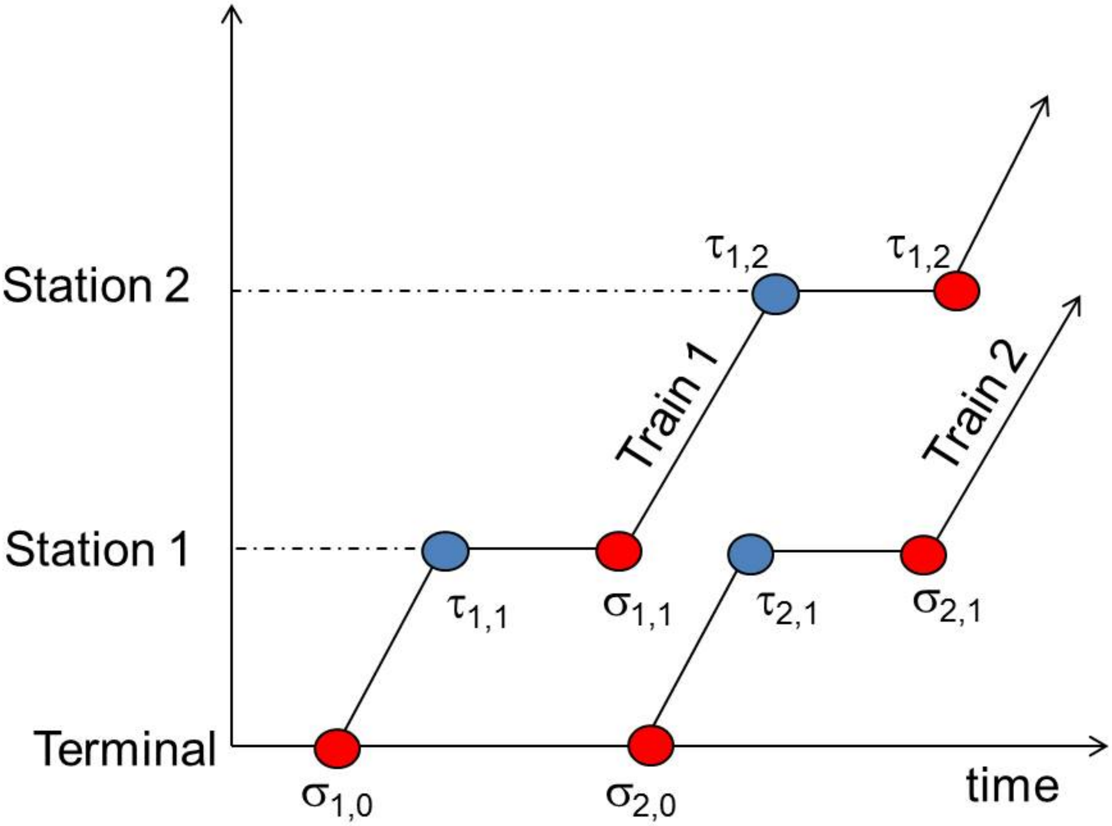

FIGURE 1 A realization of timetable with train diagram

as $\sigma_{n,s,j}$ and $\tau_{n,s,j}$ respectively. The shaded region in the figure represents the location and time period (during times t_{in} and t_{out}) that is occupied by the train of interest during which other trains are prohibited from entering. Following the specification in the current UIC (International Union of Railways) operational code (*13*), we have

$$t_{in} = \tau_{n,s,j} + \frac{\delta_{n,j}}{v_{n,s,j}}, \tag{5}$$

where $\delta_{n,j}$ is the visual distance of train n to the entrance of block j; $v_{n,s,j}$ is the nominal speed of train n traveling through block j. The time t_{in} represents the time when the driver of train n observes the signal aspect at block j and starts to take action(s) accordingly. Moreover,

$$t_{out} = \sigma_{n,s,j} + \frac{L_n}{v_{n,s,j}}, \tag{6}$$

where L_n is the length of train n. The time t_{out} represents the time when the tail of the train n clears from the block section. Because of the signaling system, it is expected congestion will occur when the train volume on a track section is high (*14, 15*). Following (5) and (6), the signal blocking constraint can then be written mathematically for all station pairs $(s, s+1)$ and signal blocks j as

$$\tau_{n+1,s,j} \geq \sigma_{n,s,j} + \frac{L_n}{v_{n,s,j}}, \tag{7}$$

in which train $n + 1$ is the train following immediately after train n.

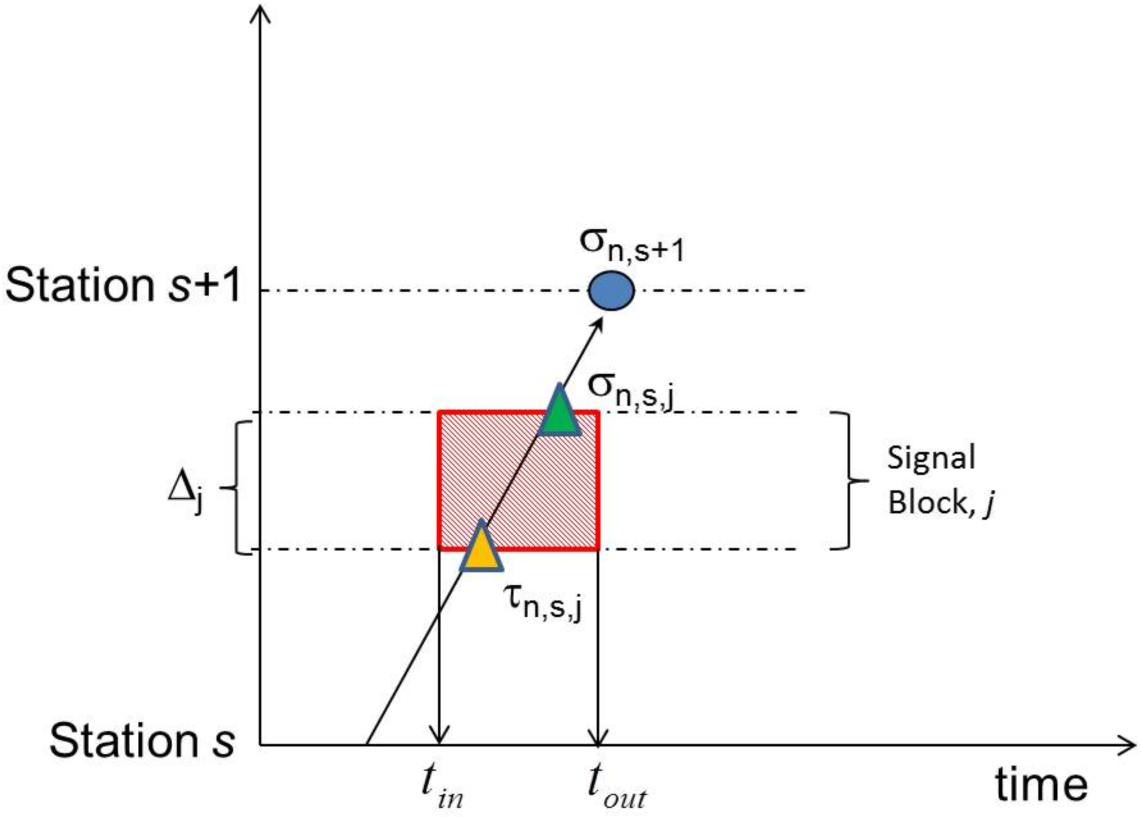

FIGURE 2 Representation of fixed block system

PERFORMANCE INDICATORS AND COST FUNCTIONS

With the timetable and the associated constraints specified, we can then formulate the cost functions to be used in the optimization framework which reflect various performance in railway timetabling and operations. Following the comprehensive review in (*4*) and (*5*), we have selected five representative performance indicators in the railway industry: train running times, customer waiting times, service punctuality, utilization of trains and track resources. It is recognized that the first three performances will be specifically interesting to customers (passengers and freight companies), utilization of trains will be interesting to Train Operators, and utilization of track will be interesting to Infrastructure Manager.

Running times of trains

Running times $T_{n,s}$ of trains n over all section $(s, s+1)$ can be obtained from Equation (1) in the previous section following the specification of timetable variables $\sigma_{n,s}$ and $\tau_{n,s}$ as discussed. Given all running times $T_{n,s}$, we define the cost associated with the running time components as

$$C_T = \hat{c}_T \sum_{n=1}^{N} \sum_{s=1}^{S} T_{n,s} p_{n,s}, \tag{8}$$

where N and S represent the total number of trains and stations in the system respectively. The variable $p_{n,s}$ is a quantity associated with the demand which represents the number of passengers (or amount of goods) on train n running between stations s and $s+1$. Determining this $p_{n,s}$ will require detailed origin-destination (OD) survey which can be difficult in practice. The quantity $p_{n,s}$ may be dropped from (8) if such OD information is not available, and this will result in the optimizer treating each train n equally. With this $p_{n,s}$, the corresponding timetable will then give higher priority to trains carrying more passengers or goods after optimization. Finally, the notation $\hat{c}_T$ represents a monetary cost associated with running times, where some examples can be found in (*11, 16, 17*). We will have further discussion on the choice of this c_T and other monetary cost coefficients in latter section.

Waiting times of passengers (or goods)

Estimating the cost associated with waiting times first requires knowledge of $\lambda_s(t)$ which denotes the profile of demand for service at station s over time t. Fundamental queueing analysis (e.g. (*18*)) gives the total waiting time W (in the unit of [persons-time] or [goods-time]) as

$$W = \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \iint_{\tau_{n,s}}^{\tau_{n+1,s}} \lambda_s(t) dt^2, \tag{9}$$

where N_s is the total number of trains serving station s over the study time period. The time interval between $\tau_{n,s}$ and $\tau_{n+1,s}$ specify the headway of train service at station s. Equation (9) can be simplified by assuming a uniform demand $\bar{\lambda}_s = \lambda_s(t)$ for all times t during the study period as:

$$W = \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \bar{\lambda}_s [\tau_{n+1,s} - \tau_{n,s}]^2. \tag{10}$$

As reflected from (10), the total waiting time grows linearly with the average demand rate $\bar{\lambda}_s$ but quadratically as the service headway increases (i.e. frequency of service decreases). However, the uniform demand assumption made in deriving (10) may be valid for high frequency service (e.g. metro) while it may not be appropriate for low frequency mainline services as it is known that the arrival of passengers will cluster around the publicized scheduled service times in the timetable. Hence some detailed survey will be needed for obtaining the demand pattern if one wants to have a reasonable estimate of waiting times when deriving mainline timetable.

Finally, following the calculation of W, the eventual cost associated with waiting times is determined as

$$C_W = \hat{c}_W W, \tag{11}$$

where $\hat{c}_W$ is the monetary cost associated with waiting times. The purpose of incorporating the waiting time into the optimization framework is to ensure that there are enough services for number of passengers or goods at the station without creating excessive waiting times. Empirical studies conducted by the UK Department for Transport (e.g. (*11, 16, 17*)) suggest that this $\hat{c}_W$ will be around two or three times larger than $\hat{c}_T$ as the waiting time is generally regarded as a dead time.

Punctuality of service

Punctuality is measured herein as the time discrepancy between the scheduled and the actual arrival times of the train services. To quantify the punctuality in monetary unit (see (*19*), (*20*)), we adopt a schedule cost function as shown in Figure 3. In the figure, τ^* denotes the ideal arrival time of the train service while Φ is a time allowance for lateness (e.g. Φ is considered to be three minutes under the UK railway operational regulations (*20*)). If the corresponding train is delayed by more than Φ from the ideal arrival time τ^*, a schedule delay cost will be imposed on the Train Operator by the Infrastructure Manager for lateness. It is considered here that this schedule delay cost increases linearly with a slope of $\hat{c}_P$ over arrival time τ, where $\tau \geq \tau^* + \Phi$. This penalty rate $\hat{c}_P$ represents the value of lost time of customers (passengers or freight companies) per unit lateness in time (*20, 21*). Following this linear specification, the total schedule delay cost associated with punctuality can be determined, taking the arrival of passengers and/or goods into account, as

$$C_P = \hat{c}_P \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \int_{\tau_{n,s}}^{\tau^*_{n+1,s}} \lambda_s(t)(\tau_{n+1,s} - \tau^*_{n+1,s} - \Phi)^+ dt, \tag{12}$$

where $\tau^*_{n+1,s}$ is the ideal arrival time for train $n+1$ at station s, $(\tau_{n+1,s} - \tau^*_{n+1,s} - \Phi)^+ = \max[(\tau_{n+1,s} - \tau^*_{n+1,s} - \Phi), 0]$. Similar to (10), Equation (12) can be simplified by assuming uniform arrival $\bar{\lambda}_s = \lambda_s(t)$ for all times t as

$$C_P = \hat{c}_P \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \bar{\lambda}_s(\tau^*_{n+1,s} - \tau_{n,s})(\tau_{n+1,s} - \tau^*_{n+1,s} - \Phi)^+. \tag{13}$$

Finally, it is noted that this punctuality cost analysis is generally applicable to other schedule cost functions, apart from the linear assumption in Figure 3, by revising the cost function term '$(\tau_{n+1,s} - \tau^*_{n+1,s} - \Phi)^+$' in (12) and (13) accordingly.

Utilization of trains

If the on-board loading $p_{n,s}$ of each train n between each station pair $(s, s+1)$ is available, we can also derive a cost associated with the utilization of trains as

$$C_L = \hat{c}_L \sum_{n=1}^{N} \sum_{s=1}^{S} \Big(1 - \frac{p_{n,s}}{p^*_n}\Big), \tag{14}$$

where p^*_n is the physical holding capacity of train n for passengers or goods, $\hat{c}_L$ is the monetary cost associated with per unit lost due to inefficient use of train holding capacity. The cost C_L will be an useful component to be included from Train Operators' perspective for deriving effective strategies transporting passengers and goods with the least number of trains.

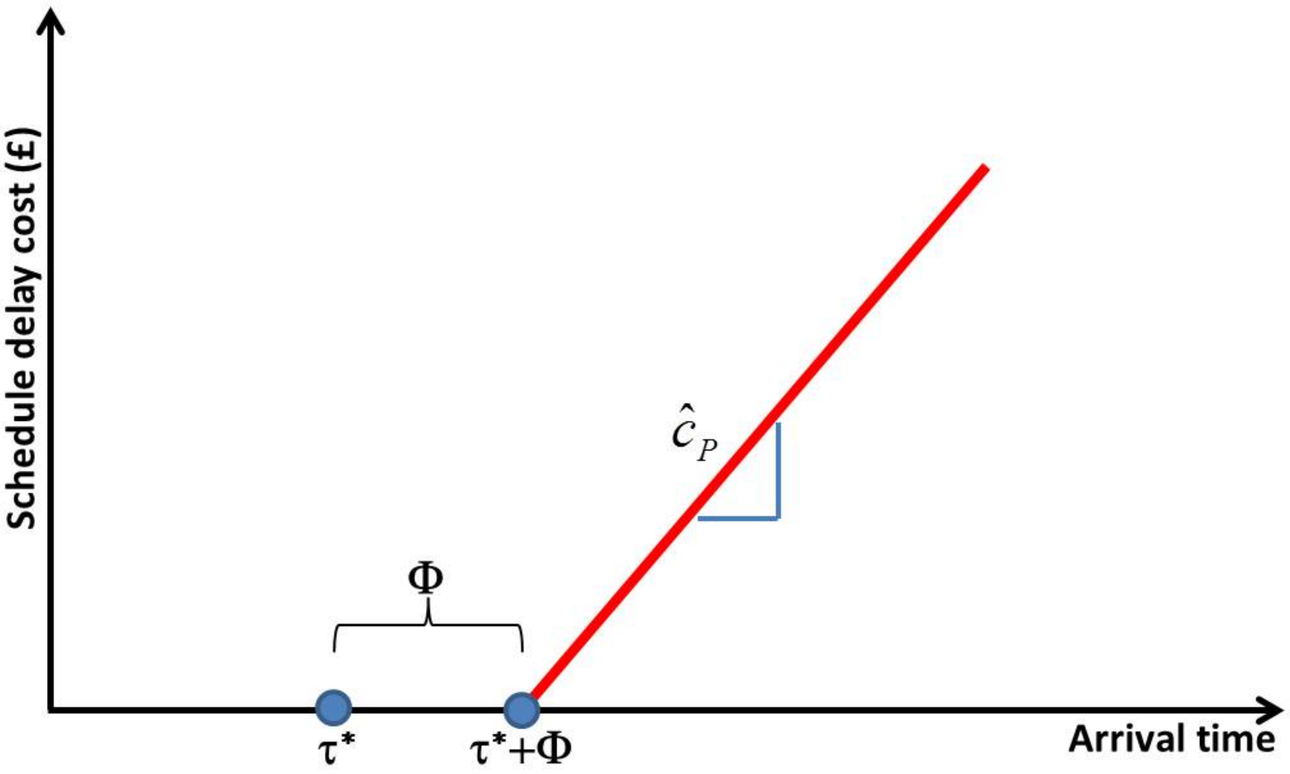

FIGURE 3 Schedule delay cost function

131 **Utilization of track capacity**

Utilization of track capacity is measured by using the occupation measure specified in the UIC 406 'Capacity' code (*13*). The occupation of a block section is defined as the total time that the block is occupied by trains within a specific study period, divided by the length of the study period. Referring to an example in Figure 4 which shows two trains passing through a block section within a specific time period T. Denote $(t_{in})_{n,j}$ and $(t_{out})_{n,j}$ respectively the entry and exit times of train n to the block (j) of interest. The occupation ratio for this block j is calculated as

$$occ_j = \frac{\sum_{n=1}^{N_j}[(t_{out})_{n,j} - (t_{in})_{n,j}]}{T}, \tag{15}$$

where N_j is the total number of trains passing the block in T. We can then come up with a network-wide measure of track utilization as over all track sections between stations s and $s+1$ in the system as:

$$OCC = \sum_{s=1}^{S}\sum_{j=1}^{J_s} occ_j, \tag{16}$$

where J_s is the number of blocks along track section between stations s and $s+1$. Different from previous UIC 405 standard (*22*) which only considers only the number of trains passing the block sections, the UIC 406 approach captures the heterogeneity and speed differentials among trains through considering the time occupied by trains. Following the OCC calculated by (15), the monetary cost associated with track utilization is determined as:

$$C_U = \hat{c}_U(1 - OCC), \tag{17}$$

where $\hat{c}_U$ is recognized as the cost per unit lost of track occupation. This C_U will be an useful indicator for Infrastructure Manager which aims to maximize the efficiency of utilizing limited infrastructure capacity.

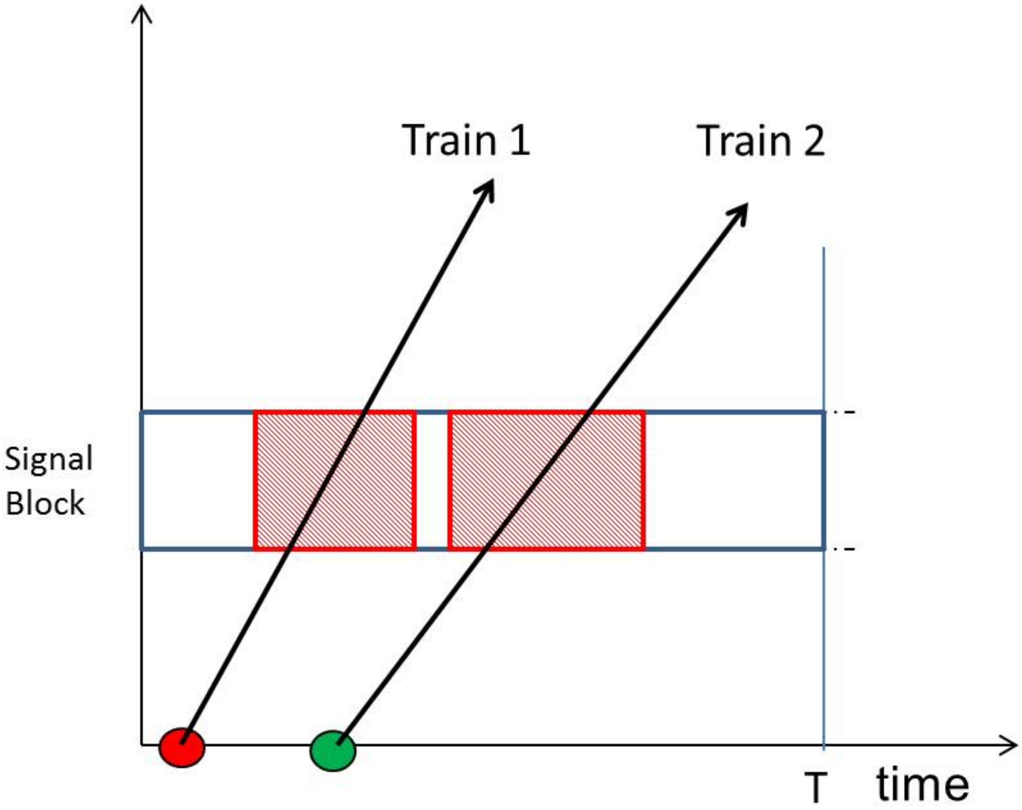

FIGURE 4 Measure of utilization of track (*13*)

Discussion

This section presents five performance indicators widely used in the railway industry and the formulations of the corresponding cost functions. It should be emphasized that the five performance indicators considered herein are mainly for illustration purposes and readers can incorporate other performances of interest such as energy consumption and connectivity in the proposed optimization framework. Detailed discussions of other performance indicators and the associated cost functions can be found in (*4*) and (*5*).

It is understood that different stakeholders have different level of interest in different performances. For example, customers will obviously be concerned about the running time, waiting time, and punctuality of their services, while they will be less interested in how the track or rolling stock resources are utilized. Infrastructure manager and train operators on the other hand will be very careful about planning the use of their resources (track and trains) in order to come up with the most cost-effective operational strategy. Unfortunately, different performances are often in conflict of each other. Utilization of track can be enhanced by running more trains within a given time period, while this could have an adverse effect on punctuality as it will be more likely generating delays due to congestion. The conflict between different performance indicators and stakeholders' interest calls for the use of multi-objective optimization technique to come up with a timetable that can maximize the overall performance of the system while taking the conflicts into account.

APPLICATION TO TIMETABLE OPTIMIZATION

The cost functions developed in previous section are applied to formulate a multi-objective optimization problem. The optimization aims to determine the train timetable, in terms of arrival $\tau_{n,s}$ and departure times $\sigma_{n,s}$ for all trains n over all stations s, such that the following total cost is minimized:

$$C = C_T + C_W + C_P + C_L + C_U. \tag{18}$$

The cost in (18) is in monetary unit and its cost components are integrated through the monetary cost coefficients: $\hat{c}_T, \hat{c}_W, \hat{c}_P, \hat{c}_L$, and $\hat{c}_U$ as discussed. The cost minimization problem is subject to the operational constraints (3), (4), and (7).

The train timetable optimization problem is combinatoric that involves different feasible combinations of $\tau_{n,s}$ and $\sigma_{n,s}$ representing different sequencing and scheduling of trains (*23, 24*). Considering a scenario where there are N trains to schedule, the number of possible sequences for scheduling these trains will be $N!$. This has not included the infinite number of ways of setting the departure and arrival times of these trains along the service route given a sequence.

To derive a solution within a reasonable time, an optimal sequence and times of departures of trains from their terminals is searched by using a genetic algorithm (GA). The genetic algorithm starts with a population (e.g. with a size of around 100) of randomly generated sequences of trains which are regarded as 'chromosomes'. Each chromosome is a combination of binary (0-1) bit representing different train sequences. For example, consider there are three trains (A, B, C) with different service paths and characteristics to schedule. This gives a total of $3! = 6$ possible sequences: ABC, ACB, BAC, BCA, CAB, and CBA. This can be represented by a set of 3-bit binary chromosomes (which gives a total of possible 8 (=2^3) combinations). Given the train sequence, the corresponding departure $\sigma_{n,s}$ and arrival times $\tau_{n,s}$ of each train is then computed by using a greedy search approach in the second stage. The greedy search strategy determines the $\sigma_{n,s}$ and $\tau_{n,s}$ as the earliest times that each train can proceed subject to constraints (3), (4), and (7). In case of a conflict occurs when two (or more) trains meet at a junction along their service lines, priority is given based upon the first-come-first-serve principle.

With the set of chromosomes containing information of sequence and departures of trains, the optimizer starts with the 'reproduction' step which reproduces chromosomes according to their 'fitness' values in the next iteration. The fitness value is calculated based upon the value of total cost (18) associated with the train sequence and departures specified in the chromosome. Essentially a higher fitness value will be assigned to a chromosome if the chromosome achieves lower total cost, and the fitness function FIT_i for each chromosome i is defined as:

$$FIT_i = \frac{A_i}{\sum A_i}, \tag{19}$$

where

$$A_i = \exp\Big(\Big(\frac{C_{max} - C_i}{C_{max} - C_{min}}\Big)p\Big), \tag{20}$$

in which C_i is the value of total cost calculated from (18) based on the sequence and departures of trains specified in chromosome i, C_{max} and C_{min} are respectively the maximum and minimum cost values identified in the current iteration of optimization, p is a parameter tuning the fitness function for maximizing the efficiency of the optimizer where it is set to be 5 here. The chromosomes are 'reproduced' in proportion to their fitness value FIT_i calculated above.

Following the reproduction step, the 'crossover' operation will then randomly select and 'mate' two chromosomes (regarded as 'parents'). The GA optimizer separates each 'parent' chromosome into two parts, swaps with each other, and forms a new pair of chromosomes (which are regarded as 'children'). This crossover process is for generating the next set of population with some entirely new characteristics with respect to the previous population and hence avoiding the optimisation process from trapping into local optima. Finally, the 'mutation' process randomly selects some bits in the population with a predefined probability (typically 0.005 - 0.01) and 'mutate' (i.e. a '0' bit will be changed to '1' arbitrarily, and a '1' bit will be changed to '0'). This is again to prevent the optimization process from trapping into local optima. The GA optimization process above (reproduction-crossover-mutation) will continue until the predefined maximum number of iterations (e.g. 20 - 30) is reached. Further details of GA can be found in a number of literature including (*25*).

CASE STUDY - BRIGHTON MAIN LINE (UK)

The optimization framework is applied to the Brighton Main Line in southeast England (Figure 5). The Brighton Main Line is approximately 80-km long electrified connection linking London Victoria and London Bridge with Brighton via East Croydon and Gatwick Airport. The line itself has a complex structure with a variable number of tracks (four tracks from London down to Balcombe Tunnel Junction and two tracks thereafter), different speed limits along the line, multiple branch lines (e.g. at Junctions Horsham, Lewes), and sidings (e.g. along Ardingly, Lovers Depot). Passenger operators that operate on the BML include Southern and First Capital Connect. We select the section between Gatwick Airport and Brighton which is highlighted in Figure 5. This is one of the busiest sections along BML. The study period is 08:00 - 10:00, which is regarded as the morning peak, on weekdays. During the study period there is currently a total of 22 trains running from Brighton toward Gatwick and hence Central London (the 'Up' direction) and 18 trains running from Gatwick toward Brighton (the 'Down' direction). We derive this 'base case' train timetable with information obtained from Network Rail. The idea is to derive an optimized timetable from the proposed optimization framework with the same number of trains within the same study period. We then compare the 'optimized' timetable with this 'base case' timetable to see how much improvement, in terms of reduction in costs, can be achieved in different aspects through re-sequencing and re-scheduling. There are two different train classes running through the section during the study period: Classes 375 and 442 with Class 375 used for the express connection. Both train classes are used for passenger transport, while it should be noted the proposed optimization framework presented in this paper can capture any number and type of train classes including freight train. Finally, it is noted that the actual origin-destination demand matrix is not made available so that an average demand rate ($\bar{\lambda}_s$) and train loading ($p_{n,s}$) will have to be estimated from field observations on a weekday. In general it is recognized that the demands at the major stations including Gatwick Airport, Three Bridges and Brighton are higher than other stations which is expected as these are some major hubs or interchanges along the line

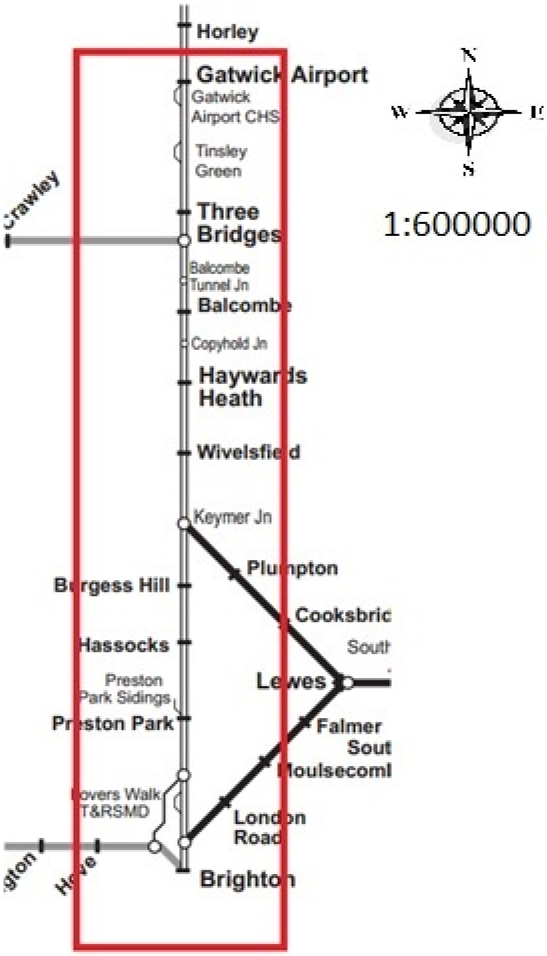

FIGURE 5 Test network - Brighton Main Line (UK)

The coefficients in the cost function (18) are set here with official documents by the British government organizations. On the customer side, the monetary cost $\hat{c}_T$ associated with running times of train is set to be £5.76 (per person-hour), while the monetary costs $\hat{c}_W$ and $\hat{c}_P$ are both £14.4 (per person-hour) for waiting times and punctuality respectively. The figures are set according the 'webTAG Unit 3.5.6' guidance (*11*) published by UK Department for Transport which specifies the values of time of travelers based on an empirical study conducted by University of Leeds (*26*). The monetary costs of waiting times and punctuality are around two times higher than the one for running time. It is because waiting times and delays due to lateness are generally regarded as non-productive dead loss. From the perspective of Train Operators and Infrastructure Manager, the costs $\hat{c}_U$ and $\hat{c}_L$ respectively for utilization costs of trains and track are set to be £350 (per train) following the current track usage price published by UK Network Rail (*27*) which specifies the cost for deploying a train on the track.

Given the network configuration and cost coefficients, the total cost (18) associated with the existing timetable is determined as £66.7k. The breakdown of this total cost into its components is shown in Figure 7. It is shown that the majority ($\sim$85%) of the cost is associated with the customer related components: running times, waiting times, and punctuality. As a consequence, it can be expected that the eventual optimized timetable would favor customers over Train Operators and Infrastructure Manager. This however can be modified with revised formulation of the cost functions and coefficients.

Figure 6 shows the progress of the optimization process in which the value of total cost is reduced gradually from the initial value £73.6k with randomly generated timetables to eventual £62.9k with the optimized one after 15 iterations given the same number of trains to schedule, the same number of passengers to serve, over the same period. The optimization process takes five minutes to complete on a standard Windows 7 (64-bit) desktop computer. Similar to other implementations of genetic algorithm (e.g. (*25*)), the most significant improvements are observed in the first few generations while the optimization process gradually converges slowly to the ultimate final solution at latter iterations.

Figure 7 further compares the cost components before and after the optimization. As aforementioned, the optimization mainly benefit the customers' costs due to their large portion in the cost components. The reduction in waiting times comes from assigning more priority to trains (e.g. the express or 'fast' trains) serving major stations with higher demand over other trains serving local area. This can be revealed from Figure 8 which compares the train diagrams under the nominal and optimized timetable toward the end of the study period (after 09:00). Under the original timetable, the fast trains (Class 375) are hindered by the slow train (Class 442) highlighted in the figure. This leads to higher costs associated with running times and hence potentially waiting times and punctuality. After optimization, more slow trains are scheduled toward the end of the study period with an objective to give way to the faster Class 375 trains in the front. As the number of available trains is considered to be fixed, the improvements in punctuality as well as utilization of trains and track, are insignificant. This however can be modified by allowing more (or less) trains to be scheduled in the optimization process. One can also estimate the marginal cost of adding or reducing a train with respect to the overall system performance.

CONCLUSIONS

This paper presents a multi-objective optimization framework which derives optimal timetables for mainline train service that maximizes the system efficiency in various performance aspects. The performances considered herein include running times of trains, waiting times of customers for service, punctuality, utilization of trains and track. The performances considered cover different stakeholders: customers, Train Operators, and Infrastructure Manager. The contributions of this paper include specification of timetable and its associated operational constraints, formulations of cost functions reflecting the corresponding performances, and multi-objective optimization with a GA-based solution method.

The optimization framework is applied to the Brighton Main Line in southeast England.

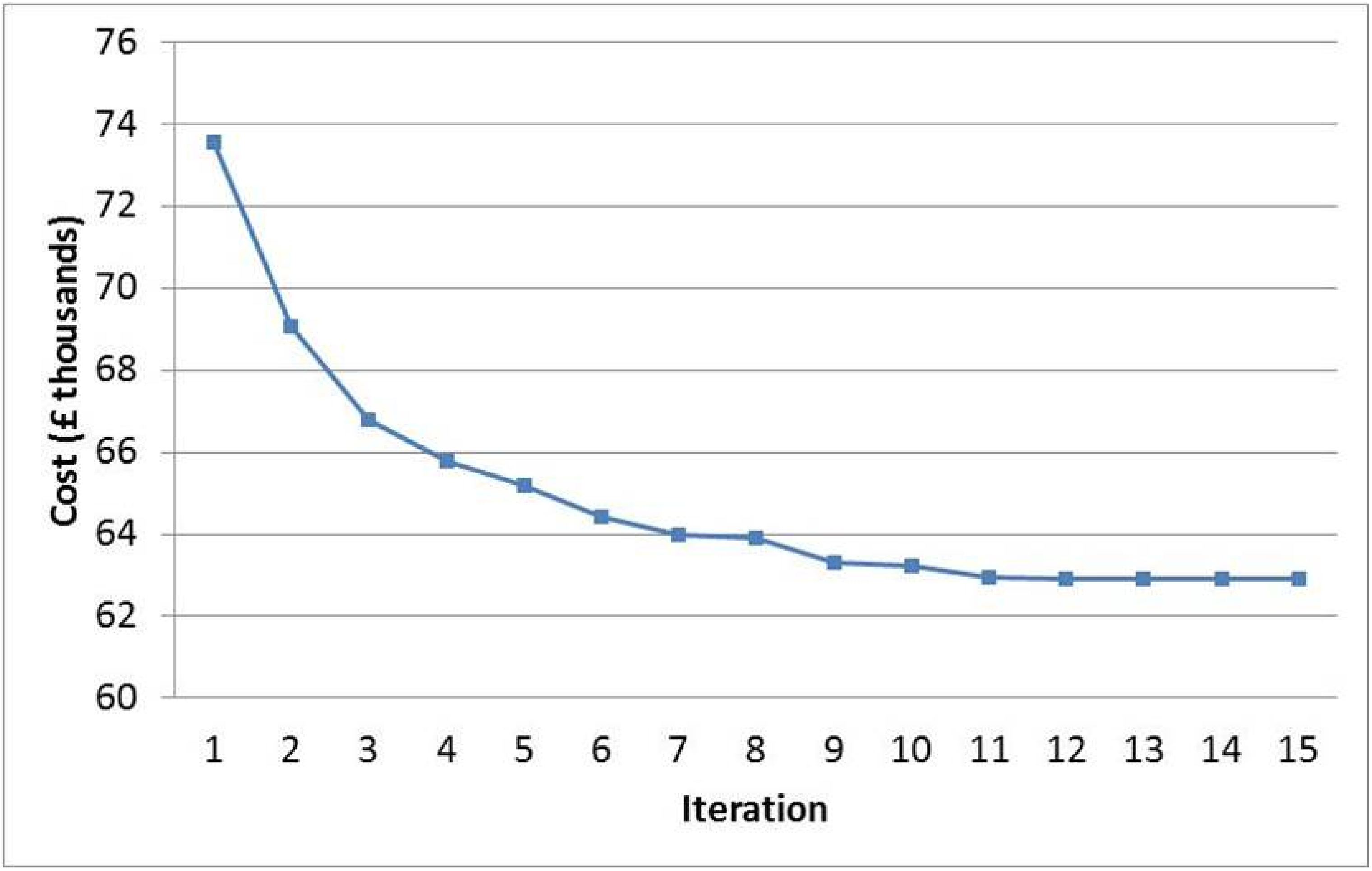

FIGURE 6 Progression of optimization process

Given the network configuration and demand, the optimal timetable is solved by a two-stage solution procedure based upon genetic algorithm. The optimizer is shown to be able to reduce the cost of operations in particular in the aspects of running times, waiting time, and punctuality. Nevertheless, it is revealed that this is achieved by assigning higher priority to fast express trains at the expense of slow local trains. This may not be a desirable result if one is interested in improving the equity of different service types. In particular, it is found that current policies of many Infrastructure Managers around the world tend to favour passenger train operations over freight ones due to the higher demand for passenger trains, higher speeds, and less energy consumed. Such timetabling and capacity allocation policy however can hurt the freight train industry in the long run. Incorporating the equity of train services will be a future research direction. Finally, it is noted that the focus of the present paper lies on the formulation of cost functions and their application to timetabling instead of the optimization algorithm. We agree that it will be worthy of conducting further research on alternative algorithms for improving the quality of the optimal solutions.

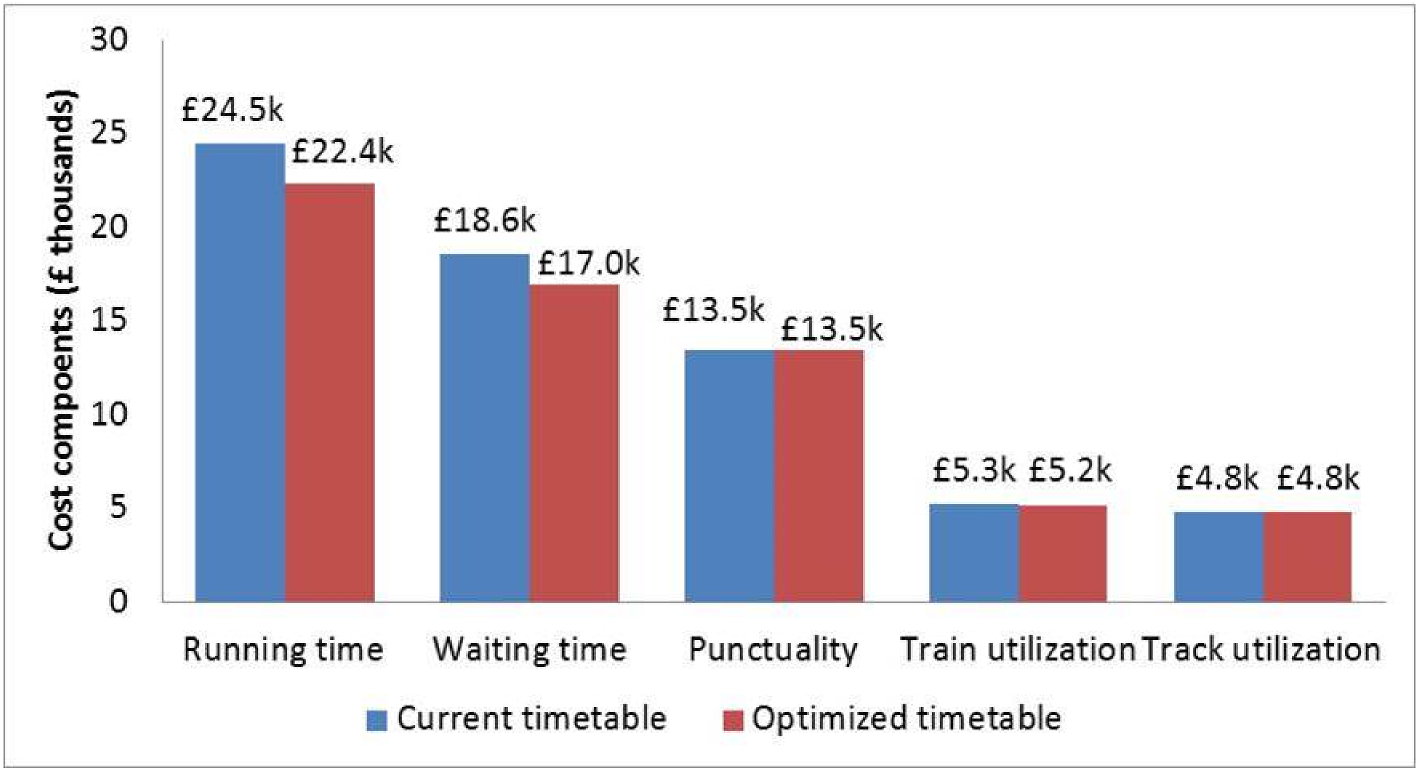

FIGURE 7 Cost components before and after optimization

ACKNOWLEDGMENTS

This study is funded by an Industrial CASE studentship awarded by UK Engineering and Physical Sciences Research Council (EPSRC) and Rail Safety and Standard Board (RSSB). We would like to thank Catherine Baker at RSSB for her valuable comments. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not reflect the official views or policies of RSSB, Network Rail, or any other organization. This paper also does not constitute a standard, specification, or regulation.

REFERENCES

[1] UK HM Treasury, *Investing in Britain's future*, 2013.

[2] Hong Kong MTR Corporation Limited, *Sustainability Report*, 2013.

[3] UK Rail Safety and Standards Board (RSSB), *The Future Railway - The Industry's Rail Technical Strategy (RTS 2012)*. Technical Strategy Leadership Group (TSLG), RSSB, London, UK, 2012.

[4] Chen, L. and Roberts, C., *Defining the Optimisation Function for Timetabling as Part of Future Traffic Regulation (FuTRO)*. Report submitted to Railway Safety and Standards Board, United Kingdom, 2012.

[5] Roberts, C., L. Chen, M. Lu, L. Dai, C. Bouch, G. Nicholson, and F. Schmid, *A Framework*

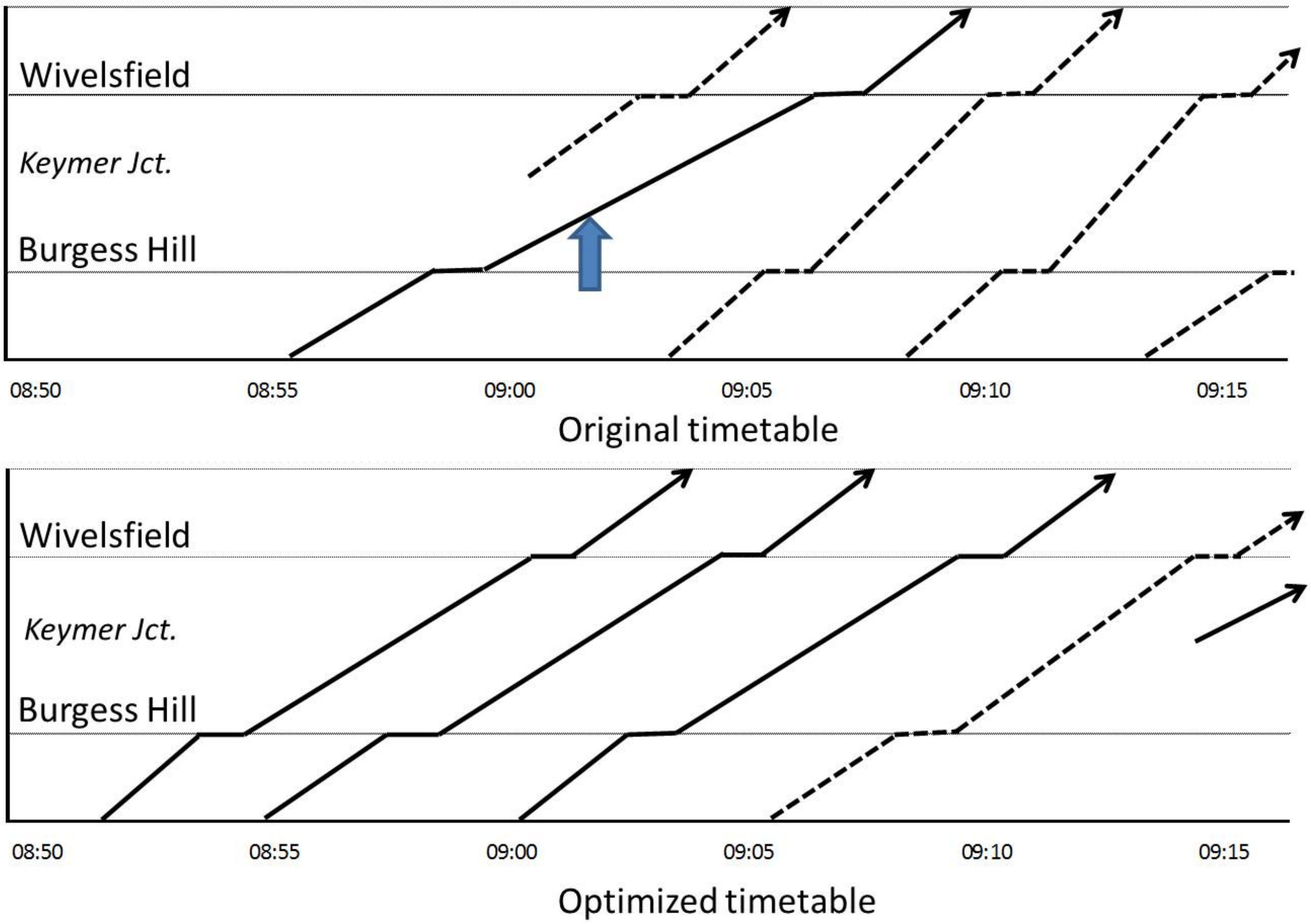

FIGURE 8 Train diagrams before and after optimization (dotted line: fast trains; solid line: slow trains)

for Developing an Objective Function for Evaluating Work Package Solutions (Cost function). Deliverable 1.2: Optimal Networks for Train Integration Management across Europe (ONTIME), European Commission Seventh Framework Programme (FP7), 2013.

[6] Ghoseiri, K., Szidarovszky, F. and Asgharpour, M. J., A multi-objective train scheduling problem and solution. *Transportation Research Part B*, Vol. 38, No. 10, 2004, pp. 927–952.

[7] Bussieck, M. R., Winter, T. and Zimmermann, U. T., Discrete optimization in public rail transport. *Mathematical Programming*, Vol. 79, No. 1-3, 2009, pp. 415–444.

[8] Vansteenwegen, P. and Van Oudheusden, D., Developing railway timetables which guarantee better service. *European Journal of Operational Research*, Vol. 173, No. 1, 2006, pp. 337–350.

[9] Vansteenwegen, P. and van Oudheusden, D., Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B*, Vol. 41, No. 4, 2007, pp. 478–492.

[10] Albrecht, T., Automated timetable design for demand-oriented service on suburban railways. *Public Transport*, Vol. 1, No. 1, 2009, pp. 5–20.

[11] UK Department for Transport, *TAG UNIT 3.5.6: Values of time and vehicle operating costs*. `www.persona.uk.com/a5dunstable/deposit-docs/DD-096.pdf`, 2012, [Accessed on: 22 Jul 2015].

[12] Pavlides, A. and Chow, A. H. F. and Baker, C., A study of cost functions for timetabling national railway operations. *Proceedings of the 47th Annual Conference of Universities Transport Study Group, January 5-7. London, UK*, 2015.

[13] International Union of Railways (UIC), *UIC Code 406: Capacity*. `banportalen.banverket.se/Banportalen/upload/1753/HandbokUIC406.pdf`, 2004, [Accessed on: 22 Jul 2015].

[14] Gibson, S., Cooper, G. and Ball, B., The evolution of capacity changes on the UK rail network. *Journal of Transport Economics and Policy*, Vol. 36, No. 2, 2002, pp. 341–354.

[15] Dingler, M. H., Lai, YC. and Barkan, C. P. L., Population Synthesis for Microsimulating Travel Behavior. *Transportation Research Record*, Vol. 2117, 2009, pp. 41–49.

[16] Wardman, M., The value of travel time - A review of British evidence. *Journal of Transport Economics and Policy*, Vol. 32, No. 3, 1998, pp. 285–316.

[17] Wardman, M., Public transport values of time. *Transport Policy*, Vol. 11, No. 1, 2004, pp. 363–377.

[18] Daganzo, C. F., *Fundamentals of Transportation and Traffic Operations*. Pergamon, New York, 1997.

[19] Carey, M., Reliability of interconnected scheduled services. *European Journal of Operational Research*, Vol. 79, No. 1, 1994, pp. 51–72.

[20] UK Network Rail, *Schedule 8: Performance Regime*, 2015.

[21] UK Office of Rail Regulation, *Passenger & Freight rail performance quality report*, 2015.

[22] International Union of Railways (UIC), *UIC Code 405: Links between Railway Infrastructure Capacity and the Quality of Operations*, 1996.

[23] D'Ariano, A., D. Pacciarelli, and M. Pranzo, A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, Vol. 183, No. 2, 2007, pp. 643–657.

[24] Xu, F., B. Heydecker, A. H. F. Chow, and T. Fujiyama, Optimisation framework for rail traffic regulation and speed optimisation at a single junction. *Proceedings of International Conference of RailTokyo*, 2015, March 23-26. Tokyo, Japan.

[25] Lo, H. and A. H. F. Chow, Control strategies for over-saturated traffic. *ASCE Journal of Transportation Engineering*, Vol. 130, No. 4, 2004, pp. 466–478.

[26] Mackie, P. J., M. Wardman, A. S. Fowkes, G. Whelan, J. Nellthorp, and J. Bates, *Values of travel time savings in the UK*, 2003.

[27] UK Network Rail, *Track Usage Price List, 2015/16 prices*. `www.networkrail.co.uk/using-our-network/cp5-access-charges/`, 2014, [Accessed on: 22 Jul 2015].