wjert, 2022, Vol. 8, Issue 3, 231-253. *Review Article* ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org SJIF Impact Factor: 5.924

BIDIRECTIONAL SINGLE TRACK TRAIN SCHEDULING: "MOMBASA-NAIROBI" SGR CASE

Niyitanga Irene*[1], Bizumuremyi Justin[2], Twiringiyimana Alexandre[3], Umugwaneza Laetitia[4] and Musabe Clement[5]

[1]Department of Mechanical Engineering, Integrated polytechnic Regional College-Kigali, Rwanda Polytechnic, KK15 Rd, Po. Box:6579 Kigali-Rwanda.

[2]School of Information, Communication and Technology, College of Science and Technology, University of Rwanda, KN 67 Rd, Po. Box:3900 Kigali-Rwanda.

[3]Department of Information, Communication, and Technology, Integrated polytechnic Regional College-Tumba, Rwanda Polytechnic, Po. Box 35 Rulindo-Rwanda

[4]Department of Mechanical Engineering, Integrated polytechnic Regional College-Ngoma, Rwanda Polytechnic, Po. Box 35 Kibungo-Rwanda.

Article Received on 08/02/2022 **Article Revised on 28/02/2022** **Article Accepted on 20/03/2022**

***Corresponding Author**

Niyitanga Irene

Department of Mechanical Engineering, Integrated polytechnic Regional College-Kigali,Rwanda Polytechnic, KK15 Rd, Po. Box: 6579 Kigali-Rwanda.

ABSTRACT

Provision of punctual and reliable services are one of the major goals in railway operation and management. Due to the complexities involved in railroad operations planning, there are several limited possibilities to improve railroad infrastructure respective to the increase of rail traffic and customer demands. It is very important to develop innovative operation management strategies to make optimum use of the existing capacity, improve profitability and the overall level of rail service Train scheduling is an important stage in railway operations planning process and it is used as the basis for railroad organization. In this paper a mathematical programming model is developed as a support tool to schedule trains on single-track railway networks as well as a planning tool to assess the impact of changes in traffic demand and railroad infrastructure. The single-track train scheduling problem is formulated as a variable-based cumulative flow model with to minimize the total train travel time, under a set of operational and safety constraints. By reformulating the infrastructure capacity using a vector of cumulative flow variables on a time-space network, the model enables the decomposition of the original complex train routing and scheduling problem into a sequence of multiple single-train optimization sub-problems to optimize the routes and passing times of each train at each station along the route. The physical railway network is constructed in NEXTA-Rail Network Editor and an open- source train scheduling package FastTrain is used to solve the proposed model. FastTrain combines an effective time-dependent shortest path algorithm in Lagrangian relaxation solution framework and a priority rule-based implementing algorithm to provide feasible solutions with useful quality measures. The developed model is verified on Mombasa-Nairobi SGR line. The impact of varying the traffic demand by increasing the number of trains as wellas opening the reserved passing stations on this network is discussed. In the network constructed with 33 stations, the average travel time of trains is generally higher than in the network with 45 stations. The average train travel time increases with increasing traffic demandin both networks. The network with more passing stations could also accommodate more traffic demand. The results obtained from this research can be directly used as a basis for railroad operations and infrastructure planning, as well as framework for further studies on the capacityof this line.

KEYWORDS: Single-track, Train scheduling, Cumulative flow, Lagrangian Relaxation, mixed integer linear programming.

1. INTRODUCTION

Railway transportation is an energy efficient mode of transportation for people and cargo, it plays a key role in the development of a country's economy. In many countries railway transport has facilitated passenger transportation, large-scale freight movement and helped to alleviate highway congestion. Being a sustainable and environmentally friendly mode of transportation, many countries are improving and expanding their railway networks as an alternative mode of transportation.

Therefore, rail transport is a capital-intensive means of transport and proper management, and planning is essential to ensure the profitability of railway enterprises in highly competitive transportation markets. To increase the market share, it is important for rail service providers to offer reliable services while at the same time ensuring the safety, economic and environmental sustainability of the rail transport system. With the dynamically changing environment, technological advancement and increasing transport demands, railway companies must constantly upgrade the efficiency of their operations. Due to the complexity of these operations, limited possibilities of improving railroad infrastructure, increasing railroad traffic and customer demands, developing innovative operation management strategies while making use of the existing capacity is key in improving the level of rail service. Timetabling is an important planning stage in railway operation management and has a significant contribution to the attraction of travelers and shippers and the general level of rail service. With increase in traffic demand on a railway network, problems with safety, punctuality, reliability, and service frequency begin to arise, hence the need for proper traffic management techniques.

Until recently, train scheduling process was carried out manually in the developing countries based on the experience and expertise of timetable planners. However, as the rail networks become more complex and real-time dispatching operations more complicated, scheduling based on manual calculation becomes so time-intensive and ineffective, affecting reliability, punctuality, and overall service level. Due to the consequent need for improved techniques to solve complex scheduling problems, several computer-based methods have been studied and developed. Currently, several automated railway timetabling systems are being used in practice, thanks to the recent improvements in the computational power of computers and the available optimization techniques. Railway companies can achieve improved quality of the train operation diagram, improved service levels and reduced operational costs, while making optimum use of the available infrastructure.

Due to high initial capital costs, a railway line must be designed as economically as possible and at the same time have sufficient capacity to meet the forecast demand.(Higgins, Ferreira et al. 1995). During the construction of the Mombasa-Nairobi Railway line, some of the designed passing stations were reserved for future construction because the line presently has low traffic demand and therefore train scheduling is not a major problem. However, passenger and freight volume on this section is predicted to increase every year and thus there is a need to study the implications on train scheduling that might arise due to increased traffic demand and opening of the reserved passing stations.

The main goal of this research is to study the train scheduling problem for use as a decision support tool for transportation planners to schedule trains on a single-track line and to help in railroad operations planning. The problem will be formulated as a variable-based cumulative flow model with the objective of minimizing the total train travel time and solved in an open-source train scheduling package, in which a train-based Lagrangian relaxation framework which provides an easy decomposition mechanism of the problem is used to obtain feasible schedules. The study will also evaluate the implications of the increase in traffic demand and railroad infrastructure changes on a railway network. Specifically, the study aims to: (1) To describe the development of a variable-based cumulative flow optimization model for the single-track train scheduling problem, to schedule trains subject to a set of operational and safety constraints. (2) To solve the proposed model to simultaneously optimize the routes and schedule of trains and output feasible solutions. (3) To investigate the impact of increasing the number of trains and the construction of more passing stations on the average train travel time.

2. Related past work and motivation of the study

Railway train scheduling problem is an important issue in railway operations planning and thus it has attracted considerable attention. Many studies devoted towards solving railway traffic management problems have been done in the past few decades. An overview of models and algorithms for real-time railway traffic management is presented in(Cacchiani, Huisman et al. 2014, Corman and Meng 2014). A recent study by (Caimi, Kroon et al. 2017) provides an overview of the railway timetable design approaches with a comparison of the different optimization models and solution methods that have been proposed for solving the railway timetabling problem.

The single-track train scheduling problem is known to be a NP-hard problem(Caprara, Fischetti et al. 2002) and optimal solutions are normally unattainable in large-scale and complex railway networks. , (B. 1973)formulated a mixed-integer programming model to determine the crossing and overtaking positions for trains with given routes and departure times on a single-track railway and designed a branch-and- bound algorithm to solve it. Considering the objective of minimizing delays and yet meeting traffic demands, (Carey 1994) developed a train pathing and timetabling mathematical model for complex rail networks with choice of lines, platforms, and routes. This model was further extended in(Carey 1994)for networks with one-way and two-way tracks.(Pellegrini, Marlière et al. 2014) proposed a mixed- integer linear programming formulation to search for the best train route and schedule in the event of real-time railway traffic management disruption. An integer programming model reformulation with cumulative flow variables based on the network is presented by(Meng and Zhou 2014), to simultaneouslyreroute and reschedule trains in an N-track network. Using a similar method, (Zhou and Teng 2016)developed an ILP model which was further reformulated as a path-choice model of trainsbased on a space-time discretized network to simultaneously route and schedule passenger trains on a rail network with both unidirectional and bidirectional tracks.

Simulation and some heuristic approaches have been widely applied to solve the large- scale and real-world train timetabling problems in the recent years. This is due to the reason that they can usually obtain a satisfactory train timetable rather than the optimal one within an acceptable computing time.(Zhou and Teng 2016)A discrete event model of railway traffic was proposed by(Dorfman M J 2004)in which a local feedback-based travel advance strategy (TAS) is developed to simulate train advances along lines of the large-scale railway network and can quickly handle perturbations in the schedule. As far as heuristic algorithms are concerned, (Carey M 2007) developed an effective heuristic algorithm to help find and resolve conflicts in the draft train schedules in complex rail networks.

Many studies have been devoted to efficient decomposition mechanisms to reduce the complexity of the models and heuristic algorithms to obtain feasible solution in reasonable computational time, e.g., in the train-based decomposition by(Lee Y 2009, Zhou and Teng 2016, Liu L 2017). Other studies have employed classical Lagrangian relaxation of the conflicting constraints to decompose the problem into shortest path problems on time discretized networks. For instance, in (Brännlund, Lindberg et al. 1998) a Lagrangian relaxation solution approach has been used to separate the original train scheduling problem into train-based dynamic programs by relaxing track capacity constraints and assigning usage prices for them.

A heuristic algorithm based on a Lagrangian relaxation of track capacity constraints is presented in(Cacchiani, Caprara et al. 2010) for a timetabling problem with both passenger and freight trains. The problem is modelled by a means of a space-time graph and using a generalization of the approach presented in(Caprara, Fischetti et al. 2002, Caprara, Monaci et al. 2006).

3. Model formulation

This section describes the formulation of a mathematical model for the single-track line train scheduling problem. The chapter introduces the conceptual illustration and notations, parameters, and variables, followed by the objective function and constraints considered in the formulation.

3.1.The Single-Track Train Scheduling Model Formulation

A mathematical formulation for the is proposed based on a time-space network structure. Cumulative flow variables are taken to model the temporal and spatial occupancy of trains on railway tracks and safety time headways, jointly optimizing the routes and passing times at each station along the selected route of each train. This way, the original complex problem is thus decomposed into a sequence of multiple single train optimization sub- problems, which are easier to solve. The network cumulative flow model framework proposed by [3] is adapted. Whereas their model focuses on rerouting and rescheduling trains during perturbations, the formulation presented in this study addresses the tactical scheduling problem, since real time traffic management is beyond the scope of this research.

3.2.Conceptual Illustration

In this research, a rail network is viewed as a set of nodes and links. Nodes represent the intersections of station tracks, switch lines or a point where tracks are merging or diverging in the physical railway network. A station is represented as a sub-network consisting of a main track and several siding tracks corresponding to a set of links. In the proposed model, the track is modelled as a link and only one train is allowed on a link any given time. Each link connects two nodes, and it is assumed to be bidirectional so that trains can traverse the track from both directions. The length of a train is assumed to be zero for simplicity. Even though the model is flexible as regards the spatial granularity, it is proposed for a macroscopic network view, and the granularity of time taken as one minute for this network.

A simple example of a rail network representation is illustrated in Figure 3-1. The single- track rail network consists of two stations with 10 nodes connected by bidirectional links.

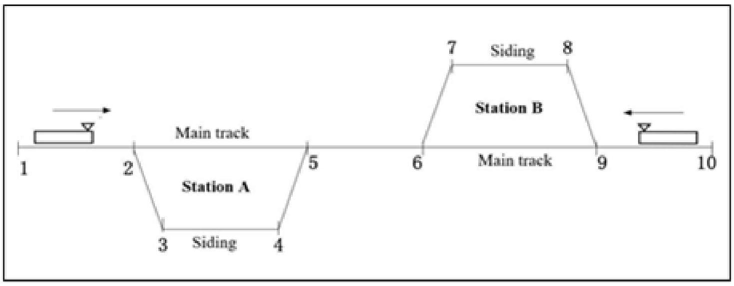

Figure 3-1: Bidirectional network with 10 nodes.

Nodes (2, 3, 4, 5) represent station A, and nodes (6, 7, 8, 9) represents station B. With the route being modelled as a series of nodes and the track being modelled as a link in the proposed model, the station minimum and maximum dwell times can be mapped as constraints on train traveling time in the corresponding link(s). For each link, input data such as free flow running time, safety headways and dwell time requirements are given. The earliest departure time, origin and destination for each train is also given. The train scheduling problem on the single-track railway network can be defined as follows: Assuming a network of railway stations and segments, the problem requires the determination of the arrival/departure times at every station for a set of trains $f \in F$ from pre-specified origin stations to destination stations in each planning horizon $t = 1, \ldots . T$, where T is the length of the planning horizon. To capture the practical safety operational rules, the network is represented as a directed graph $G = (N, E)$ with a set of nodes N and a set of links E.

3.3.General Subscripts, parameters, and variables

The general subscripts, parameters and decision variables of the proposed formulations are introduced in Tables 3-1, 3-2 and 3-3.

Table 3-1: General subscripts.

Symbol	Description
$\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$	Node index, $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \in \boldsymbol{N}$, N is the set of nodes
$\boldsymbol{e}$	Link index, $(\boldsymbol{i}, \boldsymbol{j})$, $\boldsymbol{e} \in \boldsymbol{E}$, E is the set of links
$\boldsymbol{p}$ $\boldsymbol{m}$	Route index,$\boldsymbol{p} \subset \boldsymbol{P}$, P is the set of all routes on a railway network Link sequence number along a route$\boldsymbol{p}$, $\boldsymbol{m} \leq \boldsymbol{np}$, np is the number of links
	in route p
$\boldsymbol{t}$	Scheduling time index, $\boldsymbol{t} = \mathbf{1}, \ldots . \boldsymbol{T}$ T is the planning horizon
$\boldsymbol{f}$	Train index,$\boldsymbol{f} \in \boldsymbol{F}$, F is the set of trains

Table 3-2: Input parameters.

Symbol	Description
$\boldsymbol{p}$ $\boldsymbol{P_f}$ $\boldsymbol{E_f}$ $\boldsymbol{FT_f(i, j)}$ $\boldsymbol{EST_f}$ $\boldsymbol{w^{min}(i, j)}$ $\boldsymbol{f}$ $\boldsymbol{w^{max}(i, j)}$ $\boldsymbol{f}$ $\boldsymbol{g}$ $\boldsymbol{h}$ $\boldsymbol{o_f}$ $\boldsymbol{s_f}$ $\boldsymbol{E^{os}(i)}$	Set of sequenced links of route $\boldsymbol{p}$, B$\boldsymbol{E}\boldsymbol{p}$B = $\boldsymbol{np}$| Set of possible routes on which train f may run, $\boldsymbol{P_f} \subset \boldsymbol{P}$ Set of links train $\boldsymbol{f}$ may use, $\boldsymbol{E_f} \subset \mathbf{E}$ Free flow running time for train $\boldsymbol{f}$ to traverse link $(\boldsymbol{i}, \boldsymbol{j})$ Predetermined earliest start time of train f at its origin nodeMinimum dwell time for train $\boldsymbol{f}$ on link$(\boldsymbol{i}, \boldsymbol{j})$, $(\boldsymbol{i}, \boldsymbol{j}) \in \boldsymbol{\Omega}$ Maximum dwell time for train $\boldsymbol{f}$ on link $(\boldsymbol{i}, \boldsymbol{j})$, $(\boldsymbol{i}, \boldsymbol{j}) \in \boldsymbol{\Omega}$ Safety time headway between occupancy and arrival of trains Safety time headway between departure and release of trains Origin node of train $\boldsymbol{f}$ Destination node of train $\boldsymbol{f}$ Set of links starting from or ending at node $\boldsymbol{i}$ Set of links starting from node $\boldsymbol{i}$

$\boldsymbol{E^o(i)}$	Set of links ending at node $\boldsymbol{i}$
$\boldsymbol{E^s(i)}$	Set of cells that allow dwell time, representing siding tracks in stations
$\boldsymbol{\Omega}$	Flow capacity on link $(\boldsymbol{i}, \boldsymbol{j})$ at time t, $\boldsymbol{cap}(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{t}) = \boldsymbol{0}$ due to maintenance of
$\boldsymbol{cap(i, j, t)}$	link $(\boldsymbol{i}, \boldsymbol{j})$ at time $\boldsymbol{t}$, otherwise $\boldsymbol{cap}(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{t}) = \boldsymbol{1}$.

Table 3-3: Decision variables.

Symbol	**Description**
$TT_{+}(i,j)$	Running time of train f on link $(\boldsymbol{i}, \boldsymbol{j})$
$x_{+}(i,j)$	Binary train routing variables, $x_{+}(i,j) = 1$ if train f selects link $(\boldsymbol{i}, \boldsymbol{j})$, otherwise, $x_{+}(i,j) = 0$
$y_{+}(i,j,t)$	0-1 binary time-space occupancy variables for time-space network, $y_{+}(i,j,t) = 1$ if train f occupies link $(\boldsymbol{i}, \boldsymbol{j})$ at time t, and otherwise, $y_{+}(i,j,t) = 0$.
$a_{+}(i,j,t)$	0-1 binary cumulative arrival flow variables, $a_{+}(i,j,t) = 1$ if train f has already arrived at link $(\boldsymbol{i}, \boldsymbol{j})$ by time t, and otherwise $a_{+}(i,j,t) = 0$.
$d_{+}(i,j,t)$	0-1 binary cumulative departure flow variables, $d_{+}(i,j,t) = 1$, if train f has already departed from link $(\boldsymbol{i}, \boldsymbol{j})$ by time t, and otherwise $d_{+}(i,j,t) = 0$.

3.4. Objective Function

The objective function in the model aims to minimize total trip completion time of all trains from the origin node to the destination node.

$$Z = min \sum_{f} \left\{ \left| \sum_{t} t \times \sum_{i:(i,S_f)\in E^S(S_f)\cap E_f} \left[d_f(i, S_f, t) - d_f(i, s_f, t-1) \right] \right| \right\} \tag{1}$$

3.4.1. Flow balance constraints

Constraints (2), (3) and (4) ensure flow balance on the network at the origin node, intermediate nodes, and the destination node of train f respectively.

(1) Flow balance constraints at the origin node

$$\sum_{i,j:(i,j)\in E^0(O_f)\cap E_f} x_f(i,j) = 1, \quad \forall f \tag{2}$$

(2) Flow balance constraints at intermediate nodes

$$\sum_{i,j:(i,j)\in E^S(j)\cap E_f} x_f(i,j) = \sum_{k:(j,k)\in E^0(j)\cap E_f} x_f(j,k), \ \forall f, j \in N - O_f - S_f \tag{3}$$

(3) Flow balance constraints at the destination node

$$\sum_{i,j:(i,j)\in E^S(S_f)\cap E_f} x_f(i,j) = 1, \ \forall f \tag{4}$$

3.4.2. Time-space network constraints

Constraints (5) and (6) ensure that trains do not depart earlier than predetermined earliest starting time at their origin nodes. While constraints (7) represent the transition within the link, constraints (8) ensure link-to-link transition by guaranteeing that $a_f(j, k, t) = d_f(i, j, t)$ if the adjacent links i, j and $j, k)$ are both used by train f.

Constraints (9) are imposed to map the variables $a_f(i, j, t)$in time-space network to the variables $x_f(i, j)$ in the physical network, hence describing whether link i, j is selected by train f to traverse the network from its origin to destination.

(1) Starting time constraints at the origin node

$$\sum_{j:(o_f,j)\in E_f} a_f(o_f, j, t) = 0, \quad \forall f, t < EST_f \tag{5}$$

$$\sum_{j:(o_f,j)\in E_f} d_f(o_f, j, t) = 0, \quad \forall f, t < EST_f \tag{6}$$

(2) Within link transition constraints

$$d_f\left(i, j, t + FT_f(i,j)\right) \le a_f(i, j, t), \qquad \forall f, (i,j) \in E_f, t \tag{7}$$

(3) Link-to-link transition constraints

$$\sum_{i,j:(i,j)\in E_f} d_f(i,j,t) = \sum_{j,k:(j,k)\in E_f} a_f(j,k,t), \ \forall f, j \in N - O_f - S_f, t \tag{8}$$

(4) Mapping constraints between time-space network and physical network

$$x_f(i,j) = a_f(i,j,T), \qquad \forall f, (i,j) \in E_f \tag{9}$$

(2) Mapping constraints between time-space network and physical network

$x_f(i, j) = a_f(i, j, T)$, $\forall f, (i, j) \in E_f$

Running time and dwell time constraints

While the running time for train f at link i, j can be calculated by equation (10), constraints (11) and (12) enforce the required minimum running time as well as the minimum and maximum station dwell times respectively. In this study, $w^{max}(i, j)$ is taken as 1 hour.

(1) Running time constraints

$$TT_f(i,j) = \sum_t \{t \times [d_f(i,j,t) - d_f(i,j,t-1)]\} - \sum_t \{t \times [a_f(i,j,t) - a_f(i,j,t-1)]\} \forall f, (i,j) \in E_f \quad (10)$$

(2) Minimum running time constraints

$$TT_f(i,j) \geq FT_f(i,j) \qquad \forall f, (i,j) \in E_f \quad (11)$$

(3) Minimum and maximum dwell time constraints

$$w_f^{min}(i,j) + FT_f(i,j) \leq TT_f(i,j) \leq w_f^{max}(i,j) + FT_f(i,j), \quad \forall f, (i,j) \in E_f\Omega \quad (12)$$

+

3.4.3. Safety headways and capacity constraints

Constrains (13) link time-space occupancy variables and cumulative arrival/departure variables of train f by mapping $y_+(i, j, t)$ with $a_+(i, j, t + g)$and $d_+(i, j, t - h)$if train f has started occupying link i, j by time t, $y_+(i, j, t) = 1$, otherwise $y_+(i, j, t) = 0$. If train f has ended occupying link i, j by time t, then $d_+(i, j, t) \geq d_+(i, j, t - 1), \forall f, (i, j) \in E_+, t$, otherwise 0.

Furthermore, constraints (14) enforce safety time headways by ensuring that the number of trains occupying link i, j is less than the capacity of the respective link.

(1) Link occupancy indication constraints

$$y_f(i,j,t) = a_f(i,j,t+g) - d_f(i,j,t-h), \qquad \forall f, (i,j) \in E_f, t \quad (13)$$

(2) Link capacity constraints

$$\sum_{f:(i,j)\in E_f} y_f(i,j,t) + \sum_{f:(i,j)\in E_f} y_f(j,i,t) \leq Cap(i,j,t), \qquad \forall i,j,t \quad (14)$$

3.4.4. Time-connectivity constraints

Constraints (15**Error! Reference source not found.**) and (16) represent time connectivity for cumulative flow variables.

If train f has arrived at or departed from link i, j by time t, then either $a_+(i, j, t)$ or $d_+(i, j, t)$ must have a value of 1 in all later time periods, such that $t^{\prime} \geq$ t.

$$a_f(i,j,t) \geq a_f(i,j,t-1), \quad \forall f, (i,j) \in E_f, t \quad (15)$$

$$d_f(i,j,t) \geq d_f(i,j,t-1), \quad \forall f, (i,j) \in E_f, t \quad (16)$$

4. Solution Approach

4.1. Space-Time Representation of Physical Rail Network

The train scheduling problem requires to precisely model spatial and temporal occupancy of trains on the physical infrastructure with respect to various safety headway constraints. The solution approach presented in (Meng and Zhou 2014) is adopted. In the train scheduling software package FastTrain, the input physical network is transformed into a space-time network according to discretized time units and constructed arcs. The network G is extended into a space-time network $TSG = (V, A)$ for each train f. Each node i in set N, is extended into a set of vertices (i, t) in the set of time-space network at each interval t in the planning horizon, $t = 1,2, \ldots, T$. Three types of arcs in the extended time-space network are defined to consider the feasible transitions allowed in the network, i.e., link traveling arcs (some allow dwelling while others do not), link waiting arcs at the origin link and dummy arcs at the destination node.

Through the different types of arcs, the state transition is restricted by setting an infinitely large cost for arcs which are invalid or infeasible so that the standard shortest path algorithm can be adapted for train path choice. The mapping constraints (9) between the physical network and the time-space network, and the flow balance constraints (2)-(4) on each link and the origin/destination nodes are considered by the network representation. The link occupancy capacity constraints (13) and (14) will be considered through the resource costs in the label correcting algorithm discussed in section 4.5.

As illustrated in Figure 4-1, the physical network with 4 nodes and 5 links (on the left) is transformed into the link-based space-time network on the right.

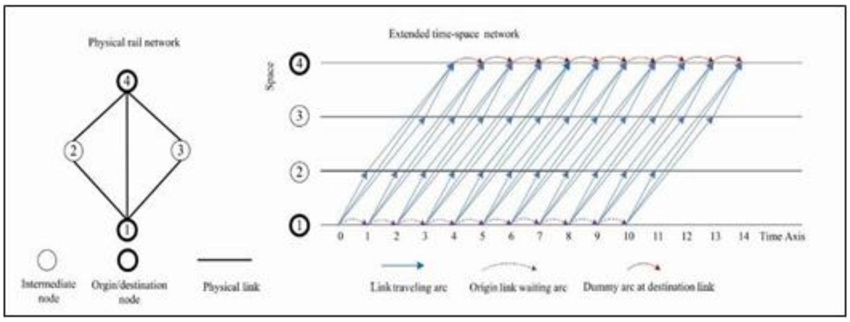

Figure 4-1: Extended space-time network representation of the physical network.

A set of binary-based cumulative flow variables $a_{+}(i, j, t)$ and $d_{+}(i, j, t)$ is introduced to represent link occupancy in the extended space-time network, where; $a_{+}(i, j, t) = 1$ if train

f has already arrived at link (i, j) by time, t, and otherwise $a_+(i, j, t) = 0$. $d_+(i, j, t) = 1$ if train f has already departed from link (i, j) by time, and otherwise $d_+(i, j, t) = 0$.

4.2. Transformation of Network Inflow Variables into Cumulative Flow Variables

To model both temporal and spatial occupancy of trains on tracks as well as safety time headways between trains, inflow variables are linked with cumulative flow variables. The cumulative flow decision variables enable a simultaneous train routing and scheduling solution approach, which implicitly enumerates all possible routes in the extended space-time network and jointly optimizes train routes as well as the train arrival and departure times.

A set of network flow variables $u_+(i, j, t)$and u'_+ (i, j, t) are used into represent the route selection and the corresponding arrival and departure times of train f. These binary network inflow variables are linked to the cumulative flow variables by equations (17) and (18).

$$u_f(i,j,t) = a_f(i,j,t) - a_f(i,j,t-1) \quad (17)$$

$$u'_f(i,j,t) = d_f(i,j,t) - d_f(i,j,t-1) \quad (18)$$

Where $u_+(i, j, t) = 1$ represents train f arriving at the upstream node i of link (i, j) at time t, $u_+(i, j, t) = 0$ otherwise; u' $(i, j, t) = 1$ represents train f departing from the downstream node i of link i, j at time t'', u' $(i, j, t) = 0$, otherwise. Figure 4-2 depicts an illustration on the usage of cumulative arrival/ departure variables to describe variables to describe link selection and arrival/departure times for train f at link (1,2).

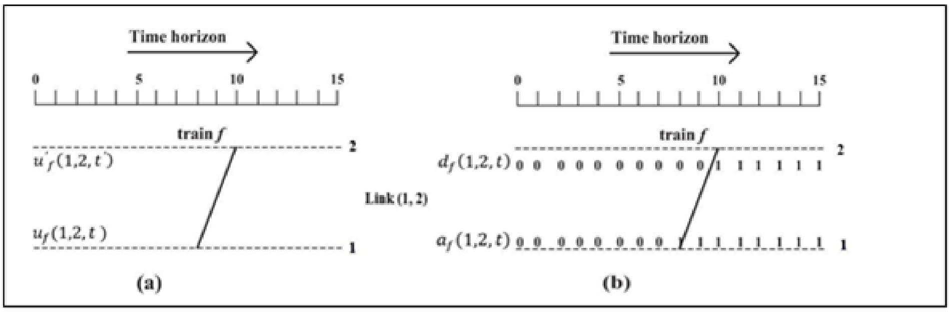

Figure 4-2: Transformation of inflow variables into cumulative flow variables.

In figure 4-2 (a) above, train f arrives at link (1,2) at time $t = 8$ and departs at time $t = 10$ with $u = (1,2,8) = 1$ and $u'(1,2,10) = 1$. In terms of cumulative flow variables as depicted in figure 4- 2(b), $a_+(1,2, t) = 0$ for $t < 8$ and $a_+(1,2, t) = 1$, for $t \geq 8$; $d_+(1,2, t) = 0$ for $t < 10$, and $d_+(1,2, t) = 1$ for $t \geq 10$. Moreover, $a_+(i, j, T) = 1$ demonstrates that the link (i, j) is used by train f to traverse the network, where T represents the planning horizon.

4.3. Modelling Safety Headways by Cumulative Flow Variables

To model the safety headway and spatial occupancy of trains, a set of shifted cumulative flow variables $a_+(i, j, t + g)$ and $d_+(i, j, t - h)$ is introduced to represent whether train f starts or ends occupying link by time t, by considering minimum safety time headways g and h . the spatial occupancyof train f is represented through the equation $y_+(i, j, t) = a_+(i, j, t + g) - d_+(i, j, t - h)$; where $y_+(i, j, t)$ is a set of 0-1 binary occupancy variables with $y_+(i, j, t) = 1$, if train f occupies link (i, j) at time t, and otherwise $y_+(i, j, t) = 0$. The planning horizon is discretized and denoted byintegers from time index 1 to T. For instance, if $g = h = 1$, the grey rectangular block in figure4-3 corresponds to $y_+(i, j, t) = 1$ for $t = 7 \ldots 10$, and $y_+(i, j, t) = 0$ otherwise, which implies that train f occupies link (i, j) from 7 minutes to time 10 minutes.

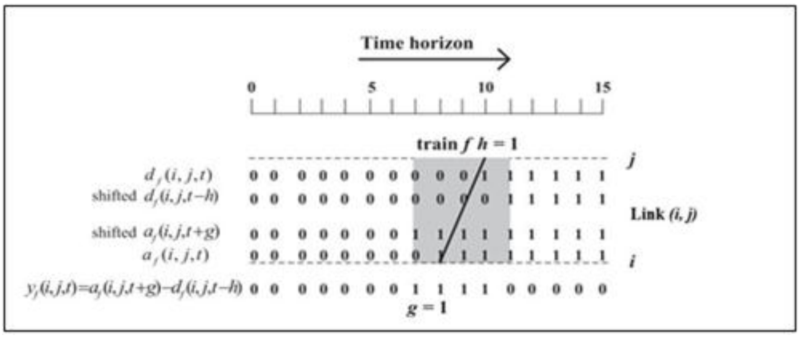

Figure 4-3: Spatial occupancy of link i, j by train f.

An illustration of a single-track case is depicted in figure 4-4. A directed e from station i to j and $e^{\rangle}$from station j to i is introduced to allow trains to run on opposite directions. Considering train $cap(i, j, t)$ using link e and train $f^{\rangle}$using link $e^{\rangle}$. Since links e and $e^{\rangle}$correspond to the same segment, a constraint $y_+(i, j, t) + y_+F\ (i, j, t) \leq 1$ can be used to model the safety headway requirement between the two trains. Specifically, $y_+(i, j, t) + y_+F(i, j, t) = 1$ for t between 3 and 9,11 and 16. Furthermore $y_+(i, j, t) + y_+F(i, j, t) = 0$ for t between 0 and 3, 9 and 11 (indicates 2 time units buffer time), 16 and 25.

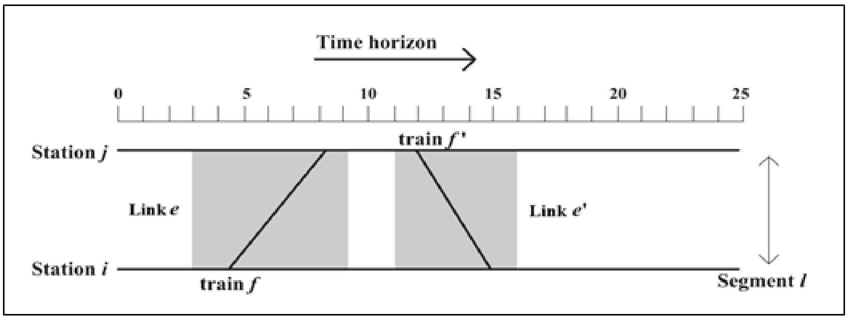

Figure 4-4: Two links corresponding to a single-track segment l from station i to j.

Based on the variables $a_{+}(i,j,t)$, $d_{+}(i,j,t)$ and $y_{+}(i,j,t)$ safety headways g and h, and $cap(i,j,t)$, the basic safety headway constraints are simply modeled by constraints (13) and (14) can decouple the original train scheduling problem into many train-specific sub-problems. The decomposition mechanism is later used in a Lagrangian Relaxation mechanism.

4.4. Lagrangian Relaxation Solution Framework

In the Lagrangian relaxation framework, the capacity constraints of the network are relaxed, and resource prices updated by Lagrangian multipliers. The original TSP is decomposed into a set of train-based sub-problems and a label correcting based algorithm is employed to determine the least time-dependent shortest path and compute the lower bound of each train traversing the network. Lagrangian profits for each of the trains are computed and a priority-based heuristic algorithm is then used to construct feasible solutions and calculate the optimality gap, upon which a termination condition is set. If the termination condition is met, the algorithm outputs the feasible solution and computational results at the current iteration.

Otherwise, a sub-gradient method updates the Lagrangian multipliers and moves to the next iteration.

$$minZ = \sum_{f}\left\{\left|\sum_{t} t \times \sum_{i:(i,S_f)\in E^S(S_f)\cap E_f}\left[d_f(i,S_f,t) - d_f(i,s_f,t-1)\right]\right|\right\} + \sum_{i,j}\sum_{t}\rho_{i,j,t} \times \left\{\sum_{f:(i,j)\in E_f} y_f(i,j,t) + \sum_{f'(j,i)\in E_{f'}} y_{f'(j,i,t)} - cap(i,j,t)\right\} \quad (19)$$

A set of non-negative Lagrangian multipliers $\rho_{i,6,t}$ is introduced as the cost incurred for utilizing a resource i.e., link (i, j) at time t; and ρ represents the iteration number. The original TSP is decomposed into a set of train-based sub-problems LR_{+} as in equations (20) and (21).

$$\max_{\rho_{i,j,t}} LR = -\sum_{i,j}\sum_{t}[\rho_{i,j,t} \times cap(i,j,t)] + min \sum_{f} LR_F \tag{20}$$

Where,

$$LR_f = \left| \sum_{t} t \times \sum_{i:(i,e_f)E^S(s_f)\cap(E_f)} [d_f(i,e_f,t) - d_f(i,e_f,t-1)] \right| + \sum_{(i,j)\in E_f}\sum_{t}[\rho_{i,j,t} \times y_f(i,j,t)] \tag{21}$$

In a sub-problem with train *f*, the objective is to find the time-dependent least generalized cost path of train *f* from its origin node to its destination node. The generalized cost includes the schedule cost and the resource cost. For train *f* traversing a network from the origin to the destination nodes, schedule cost refers to the total travel time of train *f*, while the resource cost (the second portion of equation (21) is computed by summing $\rho_{i,6,t}$ over all selected links within associated time spans.

A label-correcting based time-dependent shortest path algorithm is used to solve each sub-problem. Lagrangian profits for each train are computed after solving the train-based sub-problems, and then the trains are ranked by decreasing values of the Lagrangian profits. The Lagrangian profit of each train is the ratio of total free-flow travel time to the total travel time in the dual solution. A heuristic algorithm based on priority rules is then used to transform dual solutions into feasible solutions. Train priority is determined by the corresponding Lagrangian profits. The optimality gap at current iteration is computed based on the dual solutions and feasible solutions, and then the algorithm checks whether the termination condition is met. The termination criterion is set as: if $q > Q_{max}$ (a predetermined maximum number of iterations), then algorithm ends.

If the termination condition is met, the algorithm outputs feasible solutions along with the corresponding quality measures (i.e., optimality gap). Otherwise, a subgradient method is invoked to update Lagrangian multipliers and then move to the next iteration. The subgradient method iteratively adjusts the resource prices by setting:

$$\rho_{i,j,t}^{q+1} = max\left\{0, \rho_{i,j,t}^{q} + \alpha^{q} \times \left\{\sum_{f:(i,j)\in E_f} y_f(i,j,t) + \sum_{f':(i,j)\in E_{f'}} y_{f'}(i,j,t) - cap(i,j,t)\right\}\right\} \quad (22)$$

Where the superscript q is the iteration index used in the dual updating procedure, while $\rho^{q}_{i,j,t}$ and α^q denote the link multiplier values and step size at iteration q, respectively. In the optimumsearch process, the step size parameter is updated as $\alpha^q = 1/(q+1)$ and after a certain numberof iterations, we stop reducing α^q.

4.5. Time-Dependent Shortest Path Algorithm

The framework for the label correcting algorithm for solving the time-dependent least cost path problem is based on an extended time-space network. To compute $min \sum_{+} LR_{+}$in equation, the least cost path problem must be solved through a link-based network $minG = (N, E)$. All resource prices i.e., Lagrangian multipliers are submitted to be 0 and after label connecting process in step 2, each vertex has its least cost label and preceding vertex. A list of symbols is introduced in Table 4-1 and the shortest path algorithm is then detailed.

Table 4-1: Notation for the time-dependent shortest path algorithm.

Symbol	Description
s	Origin node, corresponding to o_{+} Destination node, corresponding to
r	s_{+} The corresponding node of vertex (i, t)
$\Theta(i, t)$	The least cost from vertex (s, EST_{+}) to vertex(j, t)
$\lambda_s(j, t)$	The preceding least cost vertex $(j, t^{`})$ denoted as time-space vertex
$\pi_s(j, t)$	(i, t)
	Free flow running time of link (i, j), corresponding to $FTf(i, j)$
$\sigma_{i,6}$	Waiting time of link (i, j) at time t
$\Delta_{i,6}(t)$	Resource cost of using link (i, j) at time t to $t + \sigma_{i,6} +$
$P_{i,6}(t, t + \sigma_{i,6}$	$\Delta \quad (t)$, P nt, $t + \sigma + \Delta\ (t)$p $= \sum^{tB_{\sigma}i,jB\Delta i,j}(t)$ ρ
$+ \Delta_{i,6}(t))$	i,6 i,6 i,6i,6 EHt i,6,E
$\Gamma(i, t)$	Set of outgoing vertexes of vertex

Input: Networks G and TSG origin node $s(i.e., o_{+})$, destination node $r(i.e., o_{+}$,starting time $t(i.e., EST_{+}$ and resource cost vector ρ at current iteration.

Output: The least cost path from s to r, at time t.

Step 1: Initialization

Create an empty SE list, set $\lambda_s(j,t) = \infty, \forall j \in N/\{s\}, t = 1, \dots, T; \lambda_s(s,t) = 0, \forall t = 1, \dots, T; \pi_s(s,t) = \emptyset, \forall t = 1, \dots, T$; insert the source vertex (s,t) into the SE list.

Step 2: Label updating While SE list is not empty **do**

Pop up the front vertex from the SE list, denoted by (i,t)

For vertex $(j, e') \in \Gamma(i,t)$, **do**

For $t = EST_s$ **to T**

For $\Delta_{i,j}(t) = w_f^{min}(i,j)$ **to** $w_f^{max}(i,j)$

If $\Theta\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) = r$ **Then**

Set candidate new cost label by $\lambda'_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) = \lambda'_s(i,t) + \vartheta_{i,j}\left(t, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) + \left|t + \sigma_{i,j} + \Delta_{i,j}(t)\right|$

Else

Set candidate new cost label by $\lambda'_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) = \lambda'_s(i,t) + \vartheta_{i,j}\left(t, t + \sigma_{i,j} + \Delta_{i,j}(t)\right)$

End

If $\lambda'_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) < \lambda_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right)$

Then

Set node cost label by $\lambda'_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right) = \lambda_s\left(j, t + \sigma_{i,j} + \Delta_{i,j}(t)\right)$

Update preceding vertex by setting π_s n$j, t + \sigma_{i,6} + \Delta_{i,6}(t)$p to time-space vertex (i,t)

If vertex (j, t'), i.e., vertex n$j, t + \sigma_{i,6} + \Delta_{i,6}(t)$p, **has been in the SE list,**

Then

Add vertex (j, t') to the front of SE list;

Else Add vertex (j, t) to the back of SE list;

End End// Updating node cost labelEnd // **for each link waiting time**

End // for each possible starting time

End// for each vertex

Remove vertex (i,t) from the SE list.

End Step 3: Fetch the time-dependent shortest path

Step 3.1: Sind the vertex (j^*, t^*) corresponding to the destination node r andwith the least cost; Set vertex (j^*, t^*) as the current vertex (k,t);

Step 3.1: Backtrack from the destination node r to node s;

While vertex k, t is not corresponding to the origin node s;

(1) Find the preceding vertex $(i, t^{\prime})$of the current vertex (k, t);

(2) Update the preceding vertex $(i, t^{\prime})$ as the current (k, t).

End

Step 3.3: Reserve the backward path and output the least cost path from s to (r) at (t);

Step 3.4: Terminate the algorithm.

4.6. Priority Rule-Based Algorithm

At each lagrangian iteration, a feasible solution based on priority rules is constructed to improve the upper bound estimate of the optimal solution. The priority rule implementing algorithm is detailed as below.

Priority rule-based implementing algorithm

Input: Network *G*, train set *F*, origin node o_+, destination node s_+, earliest departure time EST_+ for each train f.

Output: The routes and passing times at each station for each train f, and the updated upper bound.

Step 1: Train priority ranking

Rank the trains by decreasing values of Lagrangian profits. The Lagrangian profit of each train is the ratio of total free-flow travel time divided by total travel time in the dual solution.

Step 2: Schedule trains one by one

Step 2.1: For the train f^* with the highest priority, apply the shortest path algorithm introduced in Section 3.2.2 to find its route and passing times at eachstation;

Step 2.2: Fix the route and passing times at each station for train f^*; record thecapacity usage of train f^*on network *G*;

Step 2.3: If all trains have been scheduled, move to Step 3, otherwise, loop back to Step2.1.

Step 3: Update and output upper bound

Step 3.1: Compute the objective value of the heuristic solution obtained by step 2;

Step 3.2: Update the upper bound using the new objective value;

Step 3.3: Output the route, passing time at each station, and the new upper bound at the current Lagrangian iteration.

5. Numerical experiments, Results and Analysis

In this section, the proposed TSP model and solution approach proposed is applied to solve the train scheduling problem for the Mombasa-Nairobi SGR line.

The model was then implemented in FastTrain on a 1.61GHz Intel(R) Core (TM) m3- 7Y30 CPU with 4 GB of RAM. The current and long-term networks were used, and different number of trains were considered in each case (from 2 to 36) with a planning horizon, T = 1440 min. The program was allowed to terminate after 10000 iterations, before which feasible solutions were obtained in all instances. The output results provided values of the total travel time, total resource price, total trip time as well as the corresponding quality measures.

5.1. Results and analysis

The output results provided values of the total travel time, total resource price, total trip time, computational time, upper bound (UB) and lower bound (LB) values with a corresponding optimality gap. The lower bound and upper bound tend to become better with an increase of number of Lagrangian iterations. Considering the total travel time as the objective value, feasible results with the least optimality gap are considered. In the analysis, train schedules for the current (short-term) and long-term networks are obtained and compared.

Table 5-2 below shows the total travel times, upper bounds, lower bounds, and the corresponding optimality gap for different number of trains when scheduled in both the current and the long-term networks.

Table 5-2: Results.

	Current network (33 stations)				Long-term network (45 stations)			
No. of Trains	**Totaltravel time (min)**	**Lower Bound**	**Upper Bound**	**Optimality Gap**	**Totaltravel time (min)**	**Lower Bound**	**Upper Bound**	**Optimality Gap**
2	734	719	113	0.27	7	731	102	0.288
4	161	1554	260	0.36	1	1570	244	0.359
6	238	2286	332	0.31	2	2308	328	0.298
8	330	3473.	454	0.23	3	3395.5	449	0.245
1	426	4501.	561	0.19	4	4409.1	546	0.193
1	495	5411.	654	0.17	5	5263.3	646	0.186
1	570	6526.	754	0.13	5	6252.3	761	0.179
1	667	7520.	891	0.15	6	7135.6	891	0.199
1	754	8674.	996	0.12	7	8202.4	104	0.212
2	879	9863.	112	0.12	8	9297.2	114	0.189
2	961	11112	126	0.12	9	10387.	124	0.165
2	106	12134	145	0.16	1	11515.	138	0.167
2	111	13363	153	0.13	1	12478.	144	0.139

2	-	-	-	-	1	13604.	165	0.176
3	-	-	-	-	1	14710.	168	0.126
3	-	-	-	-	1	15920.	187	0.151
3	-	-	-	-	1	17657.	205	0.141
3	-	-	-	-	1	18255.	221	0.177

5.2. Analysis of Traffic Demand on Average Train Travel Time

Feasible schedules were obtained for a both the current network with 33 stations and the long-term network with 45 stations. As illustrated in figure 5-3 below, for a given number of trains, the average train travel time for the network with more sidings is lower than the average traiz travel time for the network with lesssidings. Notably, with an increase in traffic demand, there is a corresponding increase in the average train travel time.

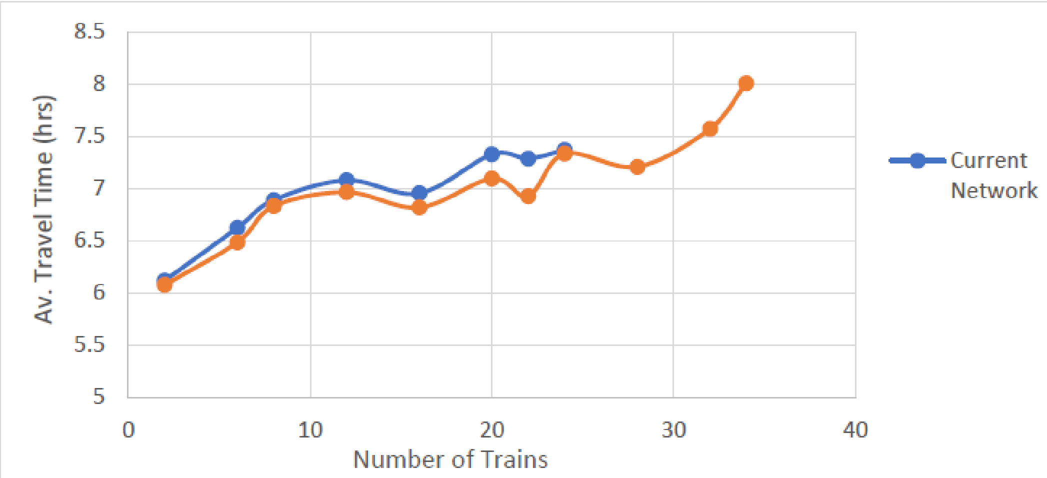

Figure 5-3: Effect of Traffic Demand on Average Train Travel Time.

As can be noted from Figure 5 above, the average train travel time increases with an increase in the number of scheduled trains for both networks under consideration. Moreover, for any given number of trains, the corresponding total travel time on the current network is generally higher than the total travel time on the long-term network. For the current network, optimal schedules could only be obtained for a maximum of 26 trains while the long-term network could take more trains.

5.3. Analysis of Traffic Demand on the Optimality Gap

The optimality gap is a relative difference between upper and the lower bound, given by the equation

$$Optimality\ Gap = \frac{Upper\ Bound - Lower\ Bound}{Upper\ Bound}$$

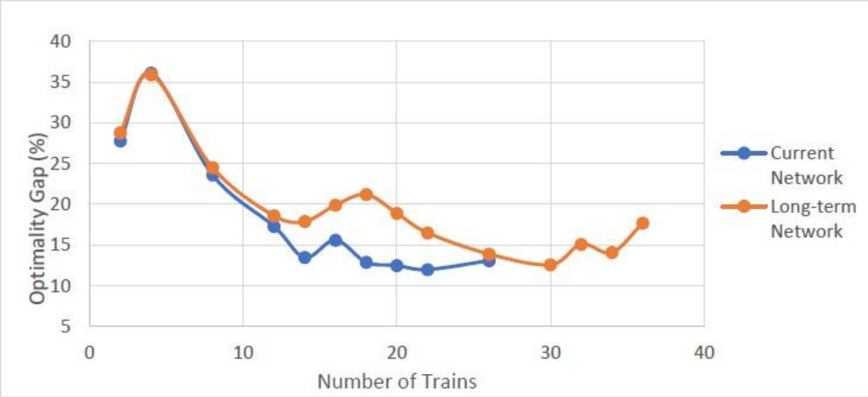

Figure 5-4: Effect of Travel Demand on the Optimality Gap.

The optimality gap decreases with the increasing number of iterations and then stabilizes. With an increase in the number of trains, the optimality gap decreases sharply at first and then with more trains, the change is not very significant. For any number of trains scheduled, the optimality gap for the long-term network appears to be higher than the one for the current network, and the difference tends to increase with an increase in the number of trains.

CONCLUSION

In this study, the train scheduling problem was formulated as a variable-based cumulative flow model for simultaneously routing and scheduling trains on a single-track railway line and applied on the Mombasa-Nairobi Railway line as a case study. By reformulating the infrastructure capacity using a vector of cumulative flow variables, the model enabled the decomposition of the original complex train routing and scheduling problem into a sequence of multiple single train optimization sub-problems to optimize the routes and passing times of each train at each station along the route. An open-source software FastTrain in which a Lagrangian relaxation solution framework combines an efficient time-dependent shortest path algorithm, and a priority-based implementing algorithm was used to solve the variable-based cumulative flow model, outputting feasible solutions with corresponding quality measures within reasonable time.

The presented model can be used as a reliable train scheduling tool for medium to large-scale networks as well as in railroad infrastructure and operations planning. It is useful to assess the impact on the train schedule due to increase in traffic demand. In the current network constructed with 33 stations, the average train travel time is generally higher than in the long-term network with 45 stations. The average train travel time also increases with increasing traffic demand in both networks. In the current network, optimal schedules could only be obtained for a maximum of 26 trains while the long-term network with more passing stations could take 36 trains. The forecast demand for year 2025 is 33 trains per day, which implies that the current short- term network could not support the demand and hence some passing stations should be opened before 2025. According to the feasibility study report, the traffic demand in the year 2035 is 51 trains per day, and therefore a capacity bottleneck is likely to occur in the long-term period. Therefore, a more detailed study on the ways to increase the capacity of this line can be carried out.

REFERENCES

1. B., S. "Optimal Scheduling of Trains on a single Line Track" Operations research, 1973; (72): 344-351.
2. Brännlund, U., P. O. Lindberg, A. Nou and J.-E. Nilsson "Railway timetabling using Lagrangian relaxation." Transportation science, 1998; 32(4): 358-369.
3. Cacchiani, V., A. Caprara and P. Toth "Scheduling extra freight trains on railway networks." Transportation Research Part B: Methodological, 2010; 44(2): 215-231.
4. Cacchiani, V., D. Huisman, M. Kidd, L. Kroon, P. Toth, L. Veelenturf and J. Wagenaar "An overview of recovery models and algorithms for real-time railway rescheduling." Transportation Research Part B: Methodological, 2014; 63: 15-37.
5. Caimi, G., L. Kroon and C. Liebchen "Models for railway timetable optimization: Applicability and applications in practice." Journal of Rail Transport Planning & Management, 2017; 6(4): 285-312.
6. Caprara, A., M. Fischetti and P. Toth "Modeling and solving the train timetabling problem." Operations research, 2002; 50(5): 851-861.
7. Caprara, A., M. Monaci, P. Toth and P. L. Guida "A Lagrangian heuristic algorithm for a real-world train timetabling problem." Discrete applied mathematics, 2006; 154(5): 738-753.
8. Carey, M. "Extending a train pathing model from one-way to two-way track." Transportation Research Part B: Methodological, 1994; 28(5): 395-400.
9. Carey, M. "A model and strategy for train pathing with choice of lines, platforms, and routes." Transportation Research Part B: Methodological, 1994; 28(5): 333-353.
10. Carey M, C. I. "Scheduling trains on a network of busy complex stations." Transportation Research Part B: Methodological, 2007; 2(41): 159-178.
11. Corman, F. and L. Meng "A review of online dynamic models and algorithms for railway traffic management." IEEE Transactions on Intelligent Transportation Systems, 2014;

16(3): 1274-1284.

12. Dorfman M J, M. J. "Scheduling trains on a railway network using a discrete event model of railway traffic." Transportation Research Part B: Methodological, 2004; 1(38): 81-98.
13. Higgins, A., L. Ferreira and E. Kozan "Modelling single line train operations." Transportation Research Record: Journal of the Transportation Research Board, 1995; 1489: 9-16.
14. Lee Y, C. C. "A heuristic for the train pathing and timetabling problem." Transportation Research Part B: Methodological, 2009; 8-9(43): 837-851.
15. Liu L, D. M. "a decomposition based hybrid heuristic algorithm for the joint passenger and freight train scheduling problem" Computers & Operations Research, 2017; (87): 165-182.
16. Meng, L. and X. Zhou Fast train: A computationally efficient train routing and scheduling engine for general rail networks. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2014.
17. Meng, L. and X. Zhou "Simultaneous train rerouting and rescheduling on an N- track network: A model reformulation with network-based cumulative flow variables." Transportation Research Part B: Methodological, 2014; 67: 208-234.
18. Pellegrini, P., G. Marlière and J. Rodriguez "Optimal train routing and scheduling for managing traffic perturbations in complex junctions." Transportation Research Part B: Methodological, 2014; 59: 58-80.
19. Zhou, W. and H. Teng "Simultaneous passenger train routing and timetabling using an efficient train-based Lagrangian relaxation decomposition." Transportation Research Part B: Methodological, 2016; 94: 409-439.