THE HONG KONG POLYTECHNIC UNIVERSITY

Department of Electrical Engineering

Hierarchical Real-time Train Control in DC Metro Systems

Wong Kwok Kam

A thesis submitted in partial fulfilment of the requirements for the

Degree of Doctor of Philosophy

June 2005

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_________________ (Signed)

Wong Kwok Kam (Name of student)

Synopsis

Metro systems or underground railways are the major means of mass transportation in metropolitan cities. To ensure the demand of passenger flow throughout a day, a reliable and regular metro service is highly desirable in life. Any delay or interruption of train service may bring a city to a standstill, which may induce a substantial economic loss and affect the public's daily activities. The main objective of this study is to develop a comprehensive train regulation and coordination system for metro systems, to optimise the train operation under the constraints of traffic conditions, tariff and computational demand.

To achieve highly efficient and flexible train operation and control, a hierarchical train regulation and control system, *HTRC*, is introduced in this study. A top-down approach is adopted in the *HTRC*. Such a decentralised control allows manageable problem complexity and reasonable computational demand for real-time applications. The operational instructions are directed to trains through the three layers of control, that is, central train controller (*CTC*), regional train controller (*RTC*) and on-board train-based controller (*TBC*). Run time and energy demand are allocated at different levels with respect to different operations constraints and requirements. Each level coordinates with the next level and enables the decision-making process within its own specific scope.

Of the three levels of train control, *TBC* is adopted to control individual train operation with respect to given run-time at the bottom level. Classical and heuristic approaches are introduced to locate optimal coasting point(s), with the aid of a single train simulator, according to specified inter-station run times. Classical methods are preferred for the search of single coasting point in a typical inter-station run. Heuristic approach to locate multiple-coasting points, however, is more applicable for a long inter-station run with extreme track geometry for the sake of energy saving.

RTC enables train movement coordination in a region to optimise train service in the second level of control. A region is defined as the section of track separating two successive substations, and the lengths of regions are fairly regular in a DC metro system. The optimal set of dwell times of trains at stations and run times for trains in successive inter-station runs (i.e., control actions) are devised by dynamic programming (DP) in this layer, under given traffic demand. Computational demand is further reduced with state grouping for real-time applications.

An unexpected short period of large peak power demand on substations may be induced when a number of trains are accelerating at the same time. The power utility usually imposes an additional cost for the peak power demand when it exceeds a specified threshold. *CTC* thereby decides the appropriate sequence of headway changes to trains in successive regions to meet the passengers flow demand throughout a day. The attained transition in headway minimises the peak power demand on the supply system. Adjustments of dwell and run times of trains are not conducted in the *CTC*. Given *CTC*'s relatively low complexity, exhaustive search and DP are employed in finding the optimal sequence of headway changes in this layer.

From the simulation results in this study, the proposed *HTRC* is capable of providing the necessary operational instructions to trains in three layers for different constraints and requirements. The *HTRC* has shown the potential benefits of on-line train coordination and control.

Publications

Journal Papers

1. Wong K.K. and Ho T.K., ‘Coast Control for Mass Rapid Transit Railways with Searching Methods’, IEE Proceedings – Electrical Power Application, pp.365-376, (Vol.151, No.3); May 2004.
2. Wong K.K. and Ho T.K., ‘Dynamic Coast Control of Train Movement with Genetic Algorithm’, International Journal of Systems Science, pp.835-846, (Vol.35, No.13-14); Oct 2004.

Conference Papers

1. Wong K.K. and Ho T.K., ‘Regulation of train service by coasting control in metro railway system’, The 9th International Conference on Enhancement and Promotion of Computational Methods in Engineering and Science, Aug 2003, University of Macau.
2. Wong K.K. and Ho T.K., ‘Coast Control of Train Movement with Genetic Algorithm’, The Congress on Evolutionary Computation 2003, Special Session on Evolutionary Computation for systems and Control Applications, Dec 2003, Canberra, Australia.
3. Wong K.K. and Ho T.K., ‘Dwell and Run Time Control in Metro System with Dynamic Programming’, IEE International Conference on Railway Engineering, Development into the 21st Century, Mar 2005, Hong Kong, P.R. China.

Acknowledgements

Special thanks are given to my supervisor, Dr. T.K. Ho, for his encouragement, patience and supervision through the work. His continuous guidance and countless hours of helpful discussion and advice on the progress of the project are deeply appreciated.

I would also like to express my gratitude to the industrial adviser, Dr. Siu, of the Kowloon-Canton Railway Corporation for sharing his valuable suggestions and experience with me.

Further, I would like to record my sincere thankfulness to my colleague, Mr. Dennis Tsang, for his disinterested exchange of know-how; and also to my family for the constant support I have received from them.

Finally, financial support from the Hong Kong Polytechnic University for carrying out this work is gratefully acknowledged.

Table of Contents

List of Figures

Chapter 8

Chapter 9

List of Tables

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Chapter 1
Background

1.1 Introduction

Electrified railway systems around the world [1] have undergone significant development in the last few decades. Metro systems or underground railways are the major means of mass transportation in many commercial and industrial cities nowadays because of environmental concerns and their cost-effective operation. The Mass Transit Railway (MTR) [2] in Hong Kong and the Mass Rapid Transit (MRT) [3] in Singapore are two typical examples. Since trains run on exclusive tracks, the service is usually free of congestion and therefore largely reliable. Indeed, a reliable, regular and safe metro service has become a part of life and is usually taken for granted by most people living and working in the cities. Any delay or interruption of service, however small, may bring a city to a standstill, which may incur significant financial loss and affect people's daily lives substantially.

Despite the fact that all metro systems must be financially well justified, with appropriate population concentrations spread along the line, most systems have to live off subsidy from the governments or local authorities. Some systems may not be well managed, but the daily operation and maintenance cost is a huge burden to bear, even for well managed systems. With all metro systems being electrified, one of the major expenses for train operation is the electricity bill. Therefore, as little as a few percentage of saving on the electricity means significant cost reduction. How to attain energy savings and maintain the service quality has become one of the topical issues among railway operators in recent years.

Train movement is governed by a large number of factors, such as track geometry, signalling, traction equipment characteristics, power supply system and speed restrictions [4-6], and co-ordinations among trains running on the same line are interactive and of multi-attributes. The common practice adopted for train operation adjustment is through variation of service headway, dwell-times at stations and inter-station run-times, to reduce energy consumption and maintain service at the same time. When taking all the track-related constraints, control variables and operational requirements into account, regulation and coordination of train operation becomes a complicated problem because of its non-linear and multi-dimensional nature. Further, the generation of multi-train operational instructions is very demanding and, therefore, the computational effort is the major concern for the operators in real-time applications.

This thesis presents a means to improve train control through advanced computational techniques and system design, observing the constraints of traffic conditions, tariff and computational capability in a DC metro system. A comprehensive train control system with a hierarchical structure has been introduced to divide the generation of trains' instructions into a small number of manageable units, in order to enhance the capability of the control. The aim of the control system is to strike an effective balance between the operational cost and the service quality through hierarchical train coordination with reasonable computational effort for real-time applications.

1.2 Train control

In practice, Automatic Train Control (*ATC*) and Automatic Train Regulation (*ATR*) [7,8] have been commonly adopted for adjustments to train services in modern railway systems, responding to the time-varying traffic conditions. To ensure reliable and safe train operation, *ATC/ATR* usually includes the following operational and functional components:

1. Automatic train protection (*ATP*) system – It provides trains with the safety signalling function to prevent collision. Adherence to speed restrictions and safe train separation are accomplished here.
2. Automatic train operation (*ATO*) system – It performs train movement between stations with speed regulation. In other words, it carries out the driver's job under normal traffic conditions and is superior to manual control, because the latter may be subject to variation in driving behaviour.
3. Automatic train supervision (*ATS*) system – It performs the functions of train monitoring, routing and regulating the train in conjunction with the timetable.

Since safety is the most important concern to passengers in train service quality, *ATP* is designed to take over the train operation and control from *ATO* and *ATS* in all conditions if *ATO* and *ATS* malfunction. In other words, a 'fail-safe' system [9] is necessary and critical in providing train service to passengers in railway operations.

With application of *ATC*(*ATR*) in modern railway operation, the decision-making process in multi-train operation is mainly performed by the *ATS*, and thus train control is centralised. As a result of relying on a single processing unit, the computational effort involved in multiple train control becomes very demanding, while quick decisions for train operation, which take all factors into consideration, cannot be fully accomplished in real-time applications. In addition, the bandwidth requirement of the communication link for bi-directional data transfer between *ATS* and the on-board *ATO* is relatively high.

To enhance the train control for real-time application, a method of hierarchical train regulation and control (*HTRC*) is proposed in this study to divide the decision-making process for multi-train operation into three levels, which are control on train, regional (i.e. a section of line) and system levels. Train service is adjusted at different levels to minimise the peak power demand and energy consumption, while respecting the operational constraints and requirements. A lower computational demand for train operational decision making is achieved by introducing several controllers. With *HTRC*, the scope of train operation and control at each level depends significantly on the power supply arrangement (i.e. AC or DC supplies) in a railway system.

The two upper levels of control in the *HTRC*, 'Central Train Controller' (*CTC*) and 'Regional Train Controller' (*RTC*) are designed to carry out the tasks of *ATS* at the system and regional levels respectively. *CTC* decides the appropriate sequence of headway changes in successive regions to meet the passenger flows and minimise the overall peak power demand on the supply system; *RTC* enables train movement coordination within a region through dwell-time and run-time control, within the constraints of the service headway imposed by *CTC*. The 'Train-Based Controller' (*TBC*) is designed to perform the function of *ATO* and to control individual train operation through coasting, at the train level, in accordance with the run-time given by the *RTC*.

The author of the thesis focuses on the design and development of these three controllers at the three different levels in the *HTRC*, to reach a balance between the power demand and service quality of train operation in a DC metro system for real-time application through hierarchical train coordination. Detailed descriptions of each control level and the traction power supply systems will be given.

1.3 Energy saving in railway operations

Electricity charges are one of the major expenses in railway operations. To achieve lower operations costs, train movement control and overall system design are two conceivable means to achieve an energy reduction.

1.3.1 Energy saving at stations

In daily railway operation, a significant proportion of the energy is used for facilities at stations. Air conditioning and ventilation systems account for major energy consumption in underground stations. To minimise energy demand and provide the passengers with a comfortable environment at stations, platform screen doors are installed on some metro systems for insulation purpose. Energy loss from air-conditioning is reduced by separating the track and platform when the train is not at the station.

Another possible approach used to save energy is the design of ‘hump’ station in some metro systems. With a ‘hump’ station, a track section of downhill slope and another of uphill slope are arranged at the ‘exit’ and ‘entrance’ of a station respectively. As maximum tractive effort is required for accelerating a train from a station, high current is thus always drawn from the power supply system. To alleviate the starting power demand for trains departing stations, a track section with downhill slope at the ‘exit’ of stations assists the train to accelerate from a standstill. Likewise, a smaller braking force is required with a track section of uphill slope at the approach of a station.

1.3.2 Energy saving with dwell-time and run-time control

Even though energy savings can be attained through better design of the facilities at stations, energy demand by train operation contributes the major proportion of the energy cost and a substantial reduction can be achieved through train coordination. Dwell-times at stations and inter-station run-times are the two major operational parameters that can be used to maintain train schedules in railway operations. Dwell-time control provides a simple approach to maintain headway regularity, regardless of other system constraints and parameters, such as traction equipment characteristics and signalling constraints. However, dwell-times cannot be extended or shortened without limitation, since the quality of train service may deteriorate as a result. The control envelope with dwell-time adjustment is therefore relatively small. From the viewpoint of passengers, a longer run-time is preferable to a longer station waiting time if ever a trip has to be lengthened, particularly during off-peak hours. In addition, train-doors have to be kept open as long as a train is at a station. Energy loss from the air-conditioning, either in the form of cooling or heating, through the open train-doors at stations accounts for a substantial proportion of the electricity bill.

Moving trains which weigh hundreds of tonnes, is the most energy-draining process in metro operations. Regenerative braking [10] and coasting [11] are the commonly used approaches to reduce energy consumption in practice. The former requires careful traffic regulation and train coordination to ensure that energy recovered from a braking train can be used to supply a motoring train nearby. Explicit dwell-time coordination for trains is thus essential to achieve energy recovery from regeneration. The latter allows simple and independent control in each train and is therefore more popular with the operators. The traction motors are turned off after the train has departed from a station and once its speed exceeds a specific value. The momentum will carry it through until the brake is required to stop the train at the next station. With different coasting speeds and positions, different travel times in an inter-station run are obtained. In other words, an energy saving can be obtained by inter-station run-time adjustment.

1.3.2.1 Regenerative braking

With the application of regenerative braking, energy exchange between trains can be improved near stations by means of dwell-time control, with one train departing and the other arriving on different platforms at the same station. However, the extent of energy saving is somewhat confined only by adjusting the dwell-time at stations rigidly. Further, the full advantage of regenerative braking can only be accomplished when an effective coordination of dwell and run time on trains is achieved.

To further achieve a high efficiency of energy recovery from train movements, energy losses through cables and rails should be minimised as much as possible by good traction power supply arrangements. A prefect energy recovery between trains could then be ensured if the energy released from braking trains is fully utilized by accelerating trains nearby. Takagi and Sone [12] reveal that a substantial peak power demand by train operations is the result, of even small time delays on train schedule, where regenerative braking is adopted. The railway equipment has to be enhanced, in practice, to cater for the sudden rise of regenerative current when there are a number of trains operating simultaneously in regenerative braking, unless the following measures can be taken, if efficient train coordination cannot be reached.

1. The terminal voltage of the driving system can be limited by a capacitor or similar energy storage system, and the surplus regenerative energy from braking trains can be fully absorbed by rheostatic braking if necessary.
2. Energy released from regenerative braking can be limited by controlling the conduction angle of the controlled rectifier driving system and the tap position of the transformer. The terminal voltage of the driving system changes with the firing angle of thyristors, and a large conduction angle usually implies more power can be taken from regeneration.

Because regenerative braking provides certain energy demands on the power supply system, efficient coordination and regulation among trains is needed to ensure the energy recovery between motoring and braking trains. Moreover, additional installation cost for regenerative operation and overcurrent protection are essential to assure safe and reliable operation.

1.3.2.2 Coast control

Coast control of train movement during inter-station runs provides some flexibility to manoeuvre between run-time and energy consumption. However, identifying the necessary starting points for coasting under the constraints of prevailing service conditions is no simple task because the train movement is influenced by a large number of factors, most of which are non-linear and inter-dependent. Even if coasting is adopted, the current practice in most metro systems is to start coasting a fixed distance away from the departed station [13]. The Singapore MRT system offers a typical example [3] – the maximum and average running speeds of train are 80 and 45 km/h respectively, in a typical inter-station run. The train accelerates from a station with maximum tractive effort and starts to coast after reaching the maximum speed. It then gradually slows down and the brake is used to bring the train to a stop at the next station.

To obtain a more flexible and efficient coast control of train movement, an audible system has been set up on the Hong Kong KCRC system [14] to alert train drivers when to start coasting during inter-station runs. The coasting speeds are recorded in the built-in lookup table, illustrated in Table 1.1. The table describes a set of optimised coasting speeds as a function of different passenger loadings for each inter-station run. It was shown that a flexible and efficient train regulation was achieved successfully, with a 3% energy reduction and a 10% saving on the maintenance cost of braking equipment, when coast control is adopted. Although the coast control is somewhat limited in this application, the result is very attractive and encouraging for the operators. To further enhance the coast control for real-time application, the coasting point(s) should be located according to the varying service demand but not the specific traffic conditions.

Table 1.1 – Coasting speed lookup table

Inter-station runs	Coasting speed* (km/h)			
	Loading 1	Loading 2	Loading 3	Loading 4
1	S_1	S_2	S_3	S_4
2	S_5	S_6	S_7	S_8
3	S_9	S_{10}	S_{11}	S_{12}
⋮	⋮	⋮	⋮	⋮

* The optimised coasting speeds are set as a function of different passenger loadings

1.4 Thesis structure

In order to ensure manageable problem complexity and reasonable computational demand for real-time railway operation, a hierarchical train coordination through headway, dwell-time and run-time control has been proposed in the thesis to attain a balance between the cost of operation and the service quality in metro systems.

The structure of this thesis is as follows:

To begin with, the hierarchical train regulation and control system (*HTRC*) is introduced in Chapter 2. The three layers of train control, system, region, and train levels, are explained, together with brief descriptions of the three train controllers: *CTC*, *RTC* and *TBC*. In order to represent the traffic flow at the three corresponding levels of control, time-based and event-based simulation models are adopted and they are explained in Chapter 3. Implications for the design of signalling systems, rolling stock and traction supply systems are also highlighted in this chapter.

Detailed discussions of each of the three levels of controllers are then covered in the subsequent chapters

In Chapters 4 and 5, a number of classical and heuristic methods are proposed to locate the optimal coasting point(s) at the train level of control (*TBC*). The performance of the methods is then compared to determine the most appropriate ones, in accordance with track layout characteristics and inter-station distances through a number of studies, with the aid of a single train simulator.

Chapters 6 and 7 look into the design of the *RTC*, for which a state-space traffic flow model has been developed, to investigate multi-train operation in a region. A number of studies to demonstrate the functions and to investigate the flexibility and performance of the controller are also discussed here.

The *CTC* is presented in detail in Chapters 8 and 9. An event-based traffic flow model has been established to represent the headway sequence. The controller is used to reduce the overall peak demand on the supply system and to maintain regular train operation at the system level. An evaluation of the *CTC* performance in multi-train operation in successive regions under different headway transitions is given. A number of studies are undertaken to explore the controller performance.

Finally, in Chapter 10, a summary of the work, as well as concluding remarks, are given, followed by discussions on possible further research.

Chapter 2
Hierarchical Train Control System

2.1 Hierarchy

A hierarchy is a system to rank and organise a number of manageable units [15]. A hierarchy usually consists of several distinct levels or layers to manage the overall system problem. In recent years, hierarchical methods have been widely adopted to enhance system efficiency in many applications, such as, traffic surveillance and control system [16], multi-media services [17], communication networks [18] and robotic systems [19]. Since the system control activity is divided between a number of layers to determine the necessary operational instructions, a relatively quick and efficient decision-making process is attained, when compared with a single centralised control system. Such a hierarchical structure a strong potential in train operation and control in real-time applications.

In addition, the concept of hierarchy is also applicable to searching and control methodologies in order to improve the flexibility in attaining a solution. For example, a hierarchical method was introduced in classical searching methods, linear programming (LP) [20] and least-squares algorithm [21], to achieve a quick convergence of solutions. Fuzzy logic control (FLC) [22] and genetic algorithms (GA) [23] with a hierarchical structure are also found in engineering applications.

Huang and Song [20] have proposed that optimised scheduling of generating units was achieved in power supply systems with two-stage linear programming within the constraints of the varying power demand. The proposed algorithm firstly determined the feasible solution (i.e., the number of units to be engaged) at the initial stage, and the optimised unit scheduling was then obtained at the subsequent stage. The study also showed that the computation effort was substantially reduced, by a ratio of 10.64, when compared with a simple LP.

Tang *et al.* [24] introduced an intelligent approach to attain the optimised numbers of fuzzy sets to enhance system performance. Hierarchical gene representation of GA (HGA) was adopted to perform this task. The gene is divided into two parts – the first part describes the possible numbers of fuzzy sets to be used, whereas the second part denotes the characteristics of membership functions. The number of fuzzy sets and membership functions are obtained through iterations and a highly flexible fuzzy logic control is thus attained with HGA. These successful applications demonstrate that the hierarchical system has the potential to provide flexible control in many applications, including that of train coordination.

2.2 Proposed hierarchical control

In most metro systems, centralised regulation and coordination of train operation are usually adopted [25,26], where the decision-making process of train service control is performed in a central control unit, as illustrated in Fig. 2.1. For example, London Underground first introduced [27] centralised computerisation for the control of lines, to replace the traditional control with ‘programme machine’ (i.e., a type of electro-mechanical device), in the early 1970’s. As a result of relying on a single processing unit, the computational effort of the multiple train decision-making process becomes very demanding, while quick decisions for train operation, which take all factors into consideration, cannot be fully accomplished for real-time applications.

In addition, the attained train schedule is only optimal with respect to a nominal schedule. Thereby, the train coordination is somewhat limited and not fully optimised with respect to the time-varying service demand.

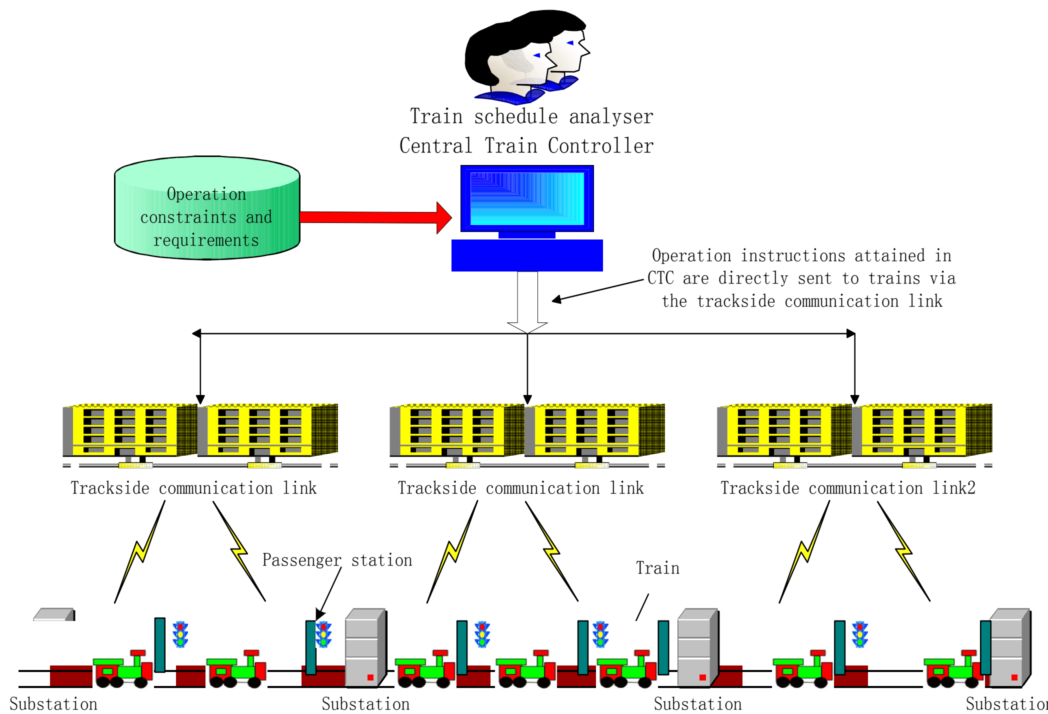


Fig. 2.1 Centralised train control system

To achieve a more efficient and flexible train operation and control for real-time applications, a hierarchical train regulation and control system, *HTRC*, is presented in this thesis and illustrated in Fig. 2.2. Under *HTRC*, a top-down approach is adopted where three layers of control, at the system, regional and train levels respectively, are integrated. The decision-making process of train operation is performed vertically by the central train controller (*CTC*), regional train controller (*RTC*) and on-board train-based controller (*TBC*). Run-times and power demand are adjusted at different levels with respect to different operational constraints and requirements. Each level coordinates with its adjacent levels and enables the decision-making process within its own specific scope. Brief descriptions of the controllers are given in the following sections. Moreover, with *HTRC*, the scope of control for *RTC* in one region is also an important factor, which depends on the traction power supply arrangement in a railway system. The main concern in this study is DC metro operations.

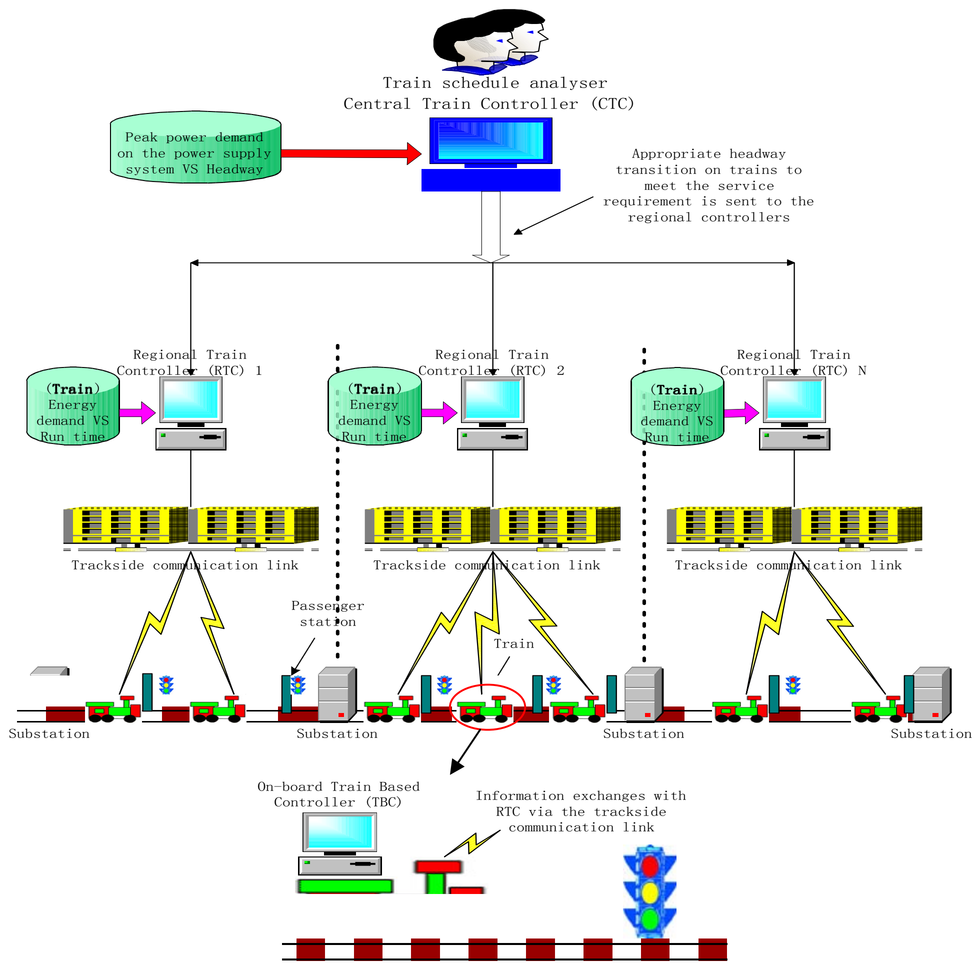


Fig. 2.2 Hierarchical train regulation and control system – *HTRC*

2.2.1 Hierarchical train control in DC traction power supply system

A typical power supply arrangement for a DC railway system is given in Fig. 2.3. The DC voltage is obtained from the local AC power transmission network through transformers and rectifiers, both located at substations. The traction power is delivered to the trains by the two nearby substations through either a third rail or an overhead wire. The running rails are usually used as the return conductor. With this feeding arrangement, the section of track between two adjacent substations is defined as a region (i.e., a section of line) in *HTRC*, which may cover a number of passenger stations. The number of feeder substations along the line hence implies the number of *RTC*s required.

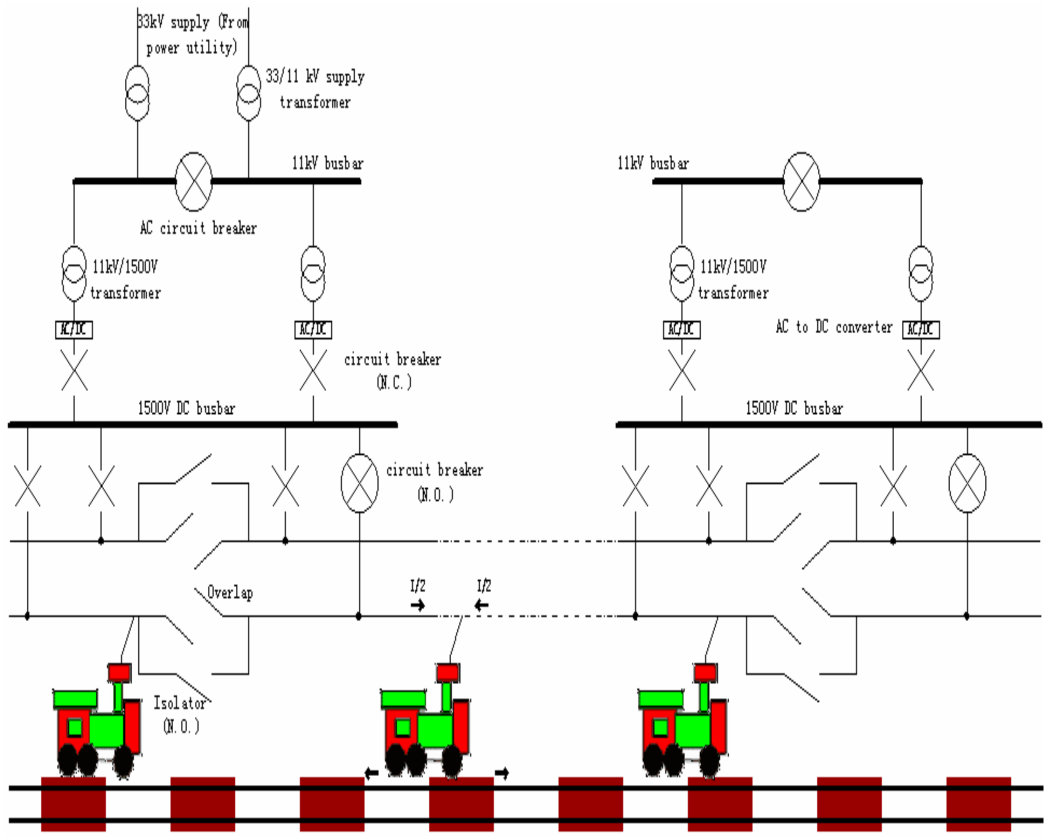


Fig. 2.3 Typical feeding arrangement for DC railway system

With the low voltage and high current supply typical for DC railway system, the substations are usually spaced quite closely to each other to limit the voltage drop in the transmission line, compared with an AC traction supply system. Detail descriptions of the modern DC traction supply systems and the scope of control using the *HTRC* with an AC traction supply system will be given in the next chapter.

2.2.2 Central train controller

The *CTC* is located at the system (line) level of the *HTRC*. The function of the *CTC* is to decide the appropriate sequence of headway changes in successive regions to meet service demand variations. The attained transitions in headway minimise the overall peak power demand on the supply system and help maintain regular train services.

Service headway is successively updated to meet the changing passenger demand from one time period to another, according to the daily predetermined timetable. During the transition between two levels of service headway (e.g., 120→240secs or 300→150secs), the traffic pattern is disrupted along the line. As a result, the peak demand on the supply system changes to meet the train service and it may either increase or decrease. In cases where a sudden and short period of large peak demand arises, an electricity tariff penalty for train operation is caused by the additional kVA demand with respect to the specified threshold imposed by the power supply utility. However, it is possible to reduce the peak demand by bringing the headway down (or up) in stages instead of in one step. For instance, 240sec→210sec→180sec→150sec, instead of 240sec→150sec. Therefore, it is desirable to investigate the possible headway-sequences at the transition between two levels of service headway.

To effectively monitor the peak power demand on the supply system and carry out headway changes in stages, a line is divided into a number of regions in the *HTRC*. The operational constraints consist of the number of regions and step changes available in headway at this level. Details on determining the appropriate sequence of headway changes in successive regions to meet the varying passenger flows will be discussed in Chapter 8.

2.2.3 Regional train controller

The *RTC* is located at the regional (section) level of the *HTRC* and the number of *RTC*s depends on the number of feeder substations in a railway line. *RTC* enables train movement coordination within its own region and acts as a platform to coordinate operational instructions for trains through *TBC*s. The *RTC* determines the appropriate set of dwell-times for trains at stations and the inter-station run-times for trains running in the region, according to the headway given by the *CTC*.

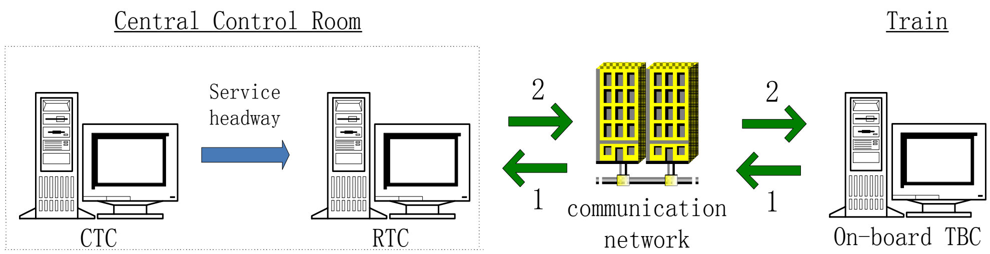
In cases where a few percent of change in headway are allowed with respect to the nominal value, it is possible to reduce energy consumption of train operation with a slight change in service regularity. *RTC* is capable of providing the optimised set of dwell and rum times of trains in a region within the limits of a small change in headway (e.g., a ±5% change in headway with respect to the nominal value). The operational constraints include the number of control steps available for dwell and run times, and the number of passenger stations in a region. In addition, some special rules are introduced in the *RTC* to recover service from disruption, when there are delays to the train service. To represent the train coordination within a region, a state-space traffic flow model has been developed. Details of the design and performance of the *RTC* will be given in Chapters 6 and 7.

2.2.4 Train-based controller

At the train level of the *HTRC*, *TBC* is adopted on board to control individual train movement in inter-station runs within the constraint of the run-times imposed by the *RTC*. To attain a flexible train movement control, coasting is the common practice to achieve a trade-off between the run-time and energy consumption. Energy saving is possible with coasting as the train spends less time in motoring mode when the run-time is allowed to extend.

To attain dynamic coast control between stations in the *TBC*, a number of classical and heuristic methods are proposed to locate the optimal coasting point(s). With the heuristic methods, a hierarchical genetic algorithm (HGA) is developed to determine the number of coasting point(s) required and its/their location(s) in an inter-station run. Details of the software development of the train controller and studies with different track layout characteristics and inter-station distances will be further described in Chapters 4 and 5.

Fig. 2.4 illustrates data exchange among the three levels of control in *HTRC*. To calculate the necessary train instructions at each level of control, a well-structured data representation, presented in Chapter 3, is vital for real-time applications.



1. Speed profiles in inter-station runs within a region.

2. Dwell times at stations and travel times in inter-station runs within a region.

Fig. 2.4 Data exchanges among the 3 levels

2.3 Advantages and disadvantages of the *HTRC*

The hierarchical train control system is preferred over the traditional centralised system because:

1. The hierarchical system can reduce the time taken for the decision-making process of train operation, by splitting the centralised control and computation into smaller tasks and solving them at successive levels, through system, regional and train control.
2. Since the decision-making process of train operation is streamlined in *HTRC*, the resolution of the time span in train control becomes smaller and smaller when 'going down' the layers. For example, the control cycle may be one hour or more at the system level, while a 10-minutes control-cycle could be adequate at the regional level. With the 3-layer control, a highly flexible supervision for train operation can thereby be accomplished.
3. The bandwidth requirement of the communication link in *HTRC* is not as high as that in a centralised system, since the amount of information transmitted

between layers is much smaller thanks to the ‘splitting’ of the decision-making process. A lower installation cost for the communication network is thus made possible.

Nevertheless, the hierarchical system is not free from drawbacks which include:

1. More controllers are involved in the system and hence their corresponding functions must be well defined and specified to avoid overlaps or gaps.
2. As more controllers are involved in *HTRC*, the total volume of data transmission among the 3 layers inflates and hence the risk of data loss increases inevitably. A higher demand on data error detection is therefore absolutely essential to enhance the reliability and integrity of data transmission.

Chapter 3
Railway System and Simulation

An electrified railway system consists of a number of subsystems, such as power supply, signalling and traction drives, which inevitably involve relatively high complexities and diversity. These subsystems are inter-linked and they interact frequently with each other when trains are running on a line. They usually have their own specific features in each railway system. With modern power electronic convertors, all combinations of AC or DC supplies and AC or DC traction machines are made possible to provide metro, suburban and mainline applications. Testing the performance of train activities against different aspects of railway operations can be achieved in a flexible and viable way by means of computer simulation. This chapter reviews the electric traction systems in common applications and various modelling techniques for simulation purposes.

3.1 Railway system and operation

Even though train operation can be optimised by some advanced techniques and system design under numerous operation-dependent and track-related constraints in railway operation, the maximum line capacity is still limited by a number of asset parameters in a railway system. Rolling stock performance, ratings of power supply facilities and signalling systems are the three important criteria in determining the maximum train service for passengers along a line, and these three factors are discussed in the following sections:

3.1.1 Signalling systems

Service headway is often referred to as one of the indicators for service quality in railway operation. A short headway implies a high train frequency for passengers but a higher energy consumption for train operation is incurred. Headway is limited by the signalling system which determines and monitors how close the trains can get to each other in a railway system. Fixed Block Signalling (FBS) and Moving Block Signalling (MBS) are the two commonly used systems around the world. FBS has been widely adopted in many railway systems over the past two centuries, in which the track is divided into a number of sections or blocks. Each section/block has its own track speed code to limit the train speed, and is occupied by not more than one train at a time [28,29]. A shorter block length in FBS can help to improve the line capacity, as more trains can run on a line with a smaller train separation.

The signalling system is designed to indicate the condition of route ahead and the details of implementation are different from one system to another. In the U.K., a signal indicates how far the section of track ahead is clear and the drivers regulate the train speed which is appropriate for the train running into the subsequent track sections. However, in the U.S., the driver does not know how far the line ahead is clear, since the signal only provides the allowable train speed for the drivers to proceed to the subsequent track sections. Further, the subsequent block (length of track section) is taken as the overlap distance in Americans' practice.

In British practices, headway distance of *n*-aspect signalling block system is described by:

$$H_d = T_L + O_d + S_d + \frac{n-1}{n-2} \times BK_d \tag{3.1}$$

where, T_L is the length of train; O_d is the overlap distance beyond a 'stop' signal; S_d is the sighting distance for driver to commence braking; and BK_d is the service braking distance for a train to stop from the maximum line speed to a standstill. Eqn. 3.1 shows that a higher number of aspects implies a smaller headway distance. To express the headway in terms of time, the following equation is given.

$$H_T = \frac{H_D}{V_{L-\max}} \tag{3.2}$$

where $V_{L-\max}$ is the maximum line speed.

With FBS, the minimum headway is described by the maximum permitted speed and the track section length. In practice, 3 or 4-aspect signalling schemes have been commonly adopted in many railway systems worldwide. For example, the 4-aspect scheme has found wide applications in U.K. [30]. An even higher aspect signalling system is employed for some high-speed operations, like the Tokaido line [31] in Japan and Altantic TGV line in France [32].

Moving Block Signalling (MBS) has been introduced to provide more room for headway reduction in a number of metro systems [33]. From Eqn 3.1, the length of track section is reduced to the minimum when $n \to \infty$ and,

$$\frac{n-1}{n-2} \approx 1, \quad O_d \text{ and } S_d = 0$$

Two successive trains are separated by a distance equivalent to the braking distance for the train behind and the length of the train ahead, together with a safety margin (*SM*). Eqn. 3.1 thereby becomes,

$$H_d = T_L + BK_d + SM \qquad (3.3)$$

The headway is then improved to the limit for the given operating speed and train characteristics, such as train length and braking rate. To provide an infinite positional resolution of the trains, MBS operations require continuous bi-directional communication links between trains and controllers, which can be distributed at track-side locations as well as being centralised. A higher frequency of train service for passengers is hence attained with MBS compared with FBS. However, most successful implementations of MBS systems are not exactly utilising the concepts in its original form [34-37], as the provision of an infinite positional resolution of trains is somewhat impractical.

3.1.2 Rolling stock

With significant developments in power conversion, control techniques and microprocessors over the past few decades, electronic DC and AC drives have become more popular in many railway operations because of their effectiveness [38]. DC motors are extensively used in variable-speed drives and they provide a high starting torque. A wide range of train speed control is applicable with DC chopper and rectifier control. An advantage of chopper control is that regenerative braking is possible for DC railway operations, which is impossible with conventional rheostatic control. Unfortunately, the use of choppers introduces odd harmonic currents of the fundamental chopping frequency into the DC traction supply system.

The magnitude of the harmonic currents depends on the train speed and position, as well as the configuration of the drive circuit.

DC drives include series-excited and separately-excited DC motor control. Chopper-fed controllers for separately excited DC motors have been widely adopted in railway applications [39]. Controlled rectifiers provide a variable DC voltage for chopper-fed DC drive from a fixed ac voltage at the feeder substations. With the characteristics of a separately-excited DC motor, the train's tractive effort can be regulated through chopper by: (1) armature voltage control, and (2) field weakening control. The train accelerates from a standstill to the base point speed through armature voltage control, at which the full voltage available is applied to the motor (i.e., constant torque region). In the constant power operation region, field weakening control is introduced to further increase train speed beyond the base point. Though DC drives are usually simpler and less expensive, compared with AC drives, DC motors are still not suitable for high speed drive because of voltage drop in brushes and commutators and hence less efficient operation.

With the advent of 3-phase induction motor and high power rated GTO device, high power drives are made possible with variable-frequency control for AC railway operations. Thanks to the elimination of commutator and brush in the AC motor, AC drives have become more competitive as compared with DC drives, since the need for maintenance is substantially reduced. In addition, the AC induction motor comes in smaller size, lighter weight and is less expensive.

AC drive enables a wide range of speed control by changing the supply frequency, voltage and current with the application of power converters and pulse width modulation inverters. The torque-speed characteristic of the induction motor varies with its control. To alter the speed of the train, there are three control regions. In the first region, the train speed can be varied by constant flux control with constant slip frequency (i.e., constant torque); the induction motor is operated at constant stator current by reducing flux in the second region; while in the third region, the speed of the motor is controlled by frequency at a reduced stator current [40].

3.1.3 Traction power system

To provide electrical energy for trains moving on a line through overhead conductor and rails, AC and DC traction power supply systems are the two commonly used approaches and they have been well established in most cities over the past few decades. The AC electrification railway system is more applicable in mainline services, which usually cover a distance of over 30 kilometres. A high supply voltage of 15 kV at $16\,{}^{2}\!/_{3}$ Hz, 25 and 50 kV at 50 or 60 Hz overhead has been adopted in all AC railway systems [41]. DC railway system is usually adopted for metro and suburban services with a shorter travelling distance in cities. Nominal supply voltages of 750, 1500 or 3000V are usually used in DC railway systems.

3.1.3.1 AC traction supply system

High voltage of AC traction supply system has been adopted for many railway operations. Electrical power for train movement is obtained from the feeder substations through the overhead line conductor (i.e., catenary), while the traction operating voltage at industrial frequency is usually obtained by stepping-down transformers and then rectifiers at substations, from 66kV, 132kV or 275kV distribution network of the local power supply company. With AC traction power supply, fewer feeder substations are needed because of high voltage and low current operation. The overhead lines are also smaller and lighter. The capital investment and maintenance cost can therefore be reduced.

132/25 kV high-voltage at industrial frequency is the commonly used power supply arrangement in most modern AC railway systems [42]. The overview of the AC electrification power supply system is illustrated in Fig. 3.1. The power supply for trains is normally taken from the 132 kV HV system through a 3-phase to 1-phase transformer, which converts 132 kV to 25 kV. Fig. 3.1 also depicts that phase 'Y-B' and 'R-Y' provide energy to trains running in section A and B of the track respectively. Phase 'B-R' acts as a 'hot-standby' power supply, and delivers the power to trains by closing the circuit breaker (N.O.) in case power failure arises either in section A or B of the track. Since the power for trains is taken from a different phase of the 3 phase supply in each section of the track, a phase break is required to isolate one phase from another. With *HTRC*, a region is defined as the section of track between two 'neutral sections'.

With the high voltage and low current supply of the AC traction systems, the power loss through the transmission network is lower, compared with DC systems. The typical distance between substations may go up to 40 kilometres. It can even be increased to 100~150 kilometres with a 50 kV autotransformer feeding scheme. A section of 183km of the South East TGV high speed line in France [43] is an example. Direct feeding, booster transformer (BT) and autotransformer schemes are the three common feeding arrangements in AC railway operations.

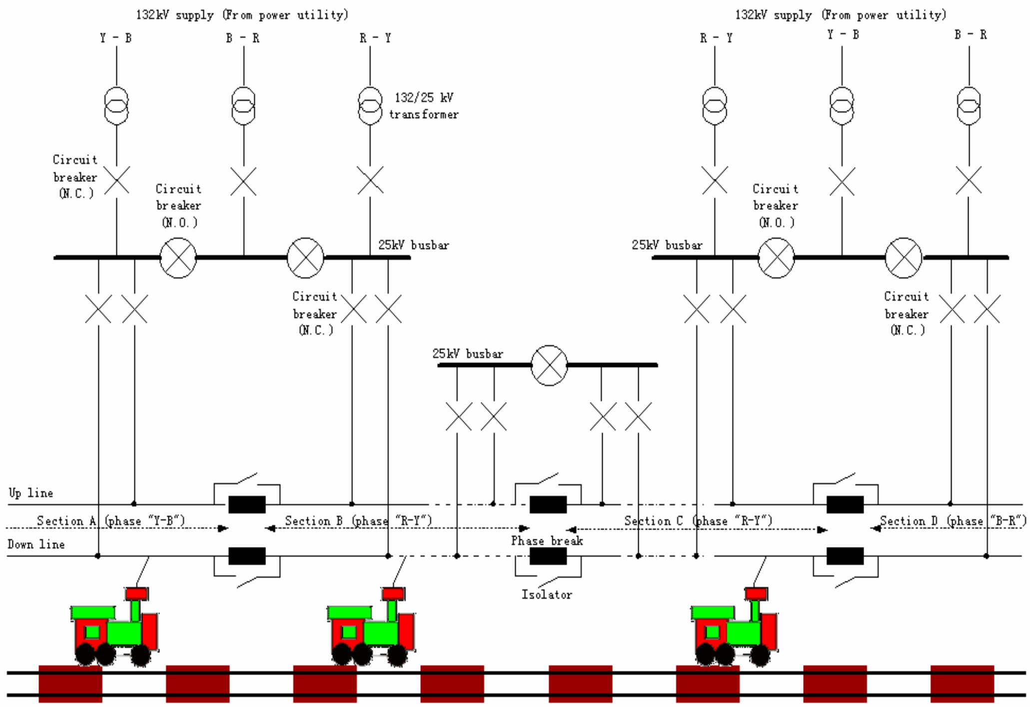


Fig. 3.1   Typical feeding arrangement for AC railway system

1. Direct feeding

Direct feeding is the most straightforward approach to provide power for train operation. The overhead contact wire and rails are connected to the secondary winding of the 132/25kV step-down transformer. Rail-to-earth leakage current is the major drawback with the direct feeding supply arrangement because of the rail-to-earth impedance. The leakage current increases the ground potential which may cause a current flow in any conductor, committed to the track. Since a portion of the current returns to the feeder substations outside the rail, the current flowing through the rails is not equal to that in the contact wire. An imbalance current often leads to substantial electromagnetic interference to the track-side communication circuitry.

2. Booster transformer (BT)

The primary winding of the BT is connected to the overhead contact wire in series at 25 kV; while the secondary one is linked with the return conductor. The typical spacing between two BTs is 3km [1]. Fig. 3.2 illustrates the BT traction supply arrangement.

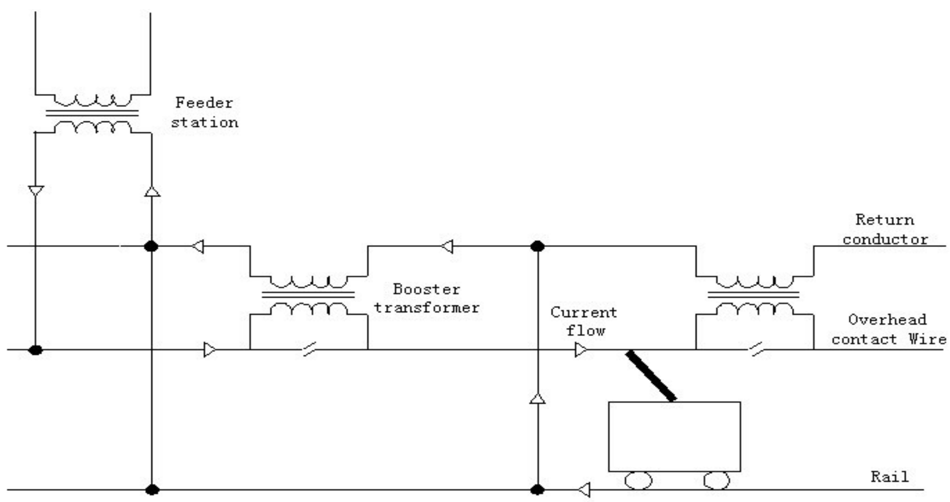


Fig. 3.2 Booster transformer traction supply scheme

BTs reduce electromagnetic interference from a railway line [44] by forcing the current flowing through the return conductors to become the same as that in the overhead contact wire since the winding ratio of BTs is 1:1. However, a major drawback is that the impedance of the booster transformer connected with the conductors along the line is greatly increased, which induces a higher energy loss. As a result, substation separation is preferably shorter in BT feeding, compared with direct feeding, because of the extra voltage drop in the line.

3. Autotransformer (AT)

With the autotransformer supply scheme [45] shown in Fig. 3.3, a high transmission voltage of 50kV is provided to a railway line. The traction power for trains is obtained from the power circuit between 25kV overhead contact wire and rail through the split winding ratio of the AT. The rail and earth currents are retuned to the feeder substations at 50kV through the two adjacent autotransformers.

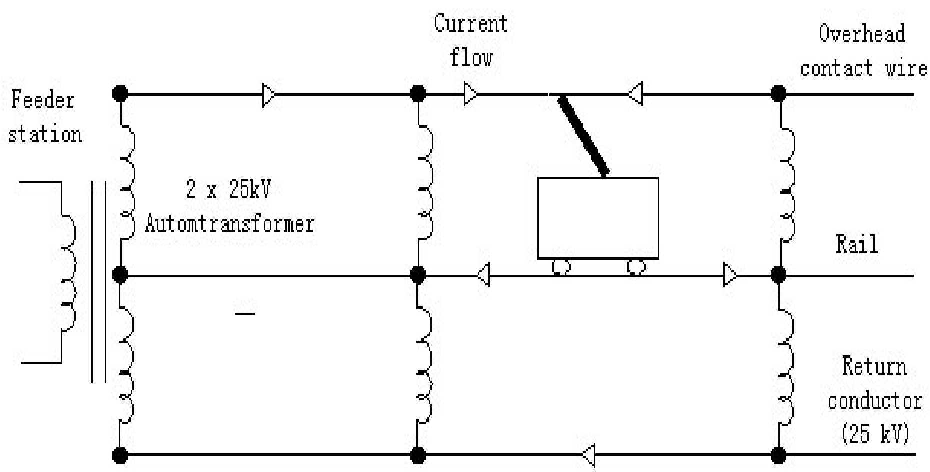


Fig. 3.3 50kV autotransformer traction supply scheme

The advantage of AT traction supply system is that the voltage drop on the contact wire is reduced along the line with a high voltage and low current operation, when compared with that in a 25kV supply system. With lower line current, longer separation between two feeder substations is allowed. Further, thanks to the voltage balance arrangement with AT, the electromagnetic interference on the track-side communication cable is alleviated. In addition, simple arrangement of the contact wire is possible since the contact wire and return conductor no longer need to be sectioned. Nevertheless, more circuit breakers are required in AT feeding.

3.1.3.2 DC traction supply system

A typical power supply arrangement for DC railway system was briefly described in Section 2.2.1. The DC power for trains is obtained at the local 132kV, 66kV or 33kV of the power utility. With reference to Fig. 2.3, the supply voltage is stepped down through a transformer and then rectified at the feeder substations. Each feeder substation covers a section of track and the length of section is usually short because of the lower distribution voltage level. The number of feeder substations to maintain train service depends on the volume of traffic and the track topology.

The traction current of trains is obtained from the two adjacent feeder substations through either a third rail [1] or an overhead cable, where the running rails serve as the return conductor. London Underground [46], where a fourth rail scheme is employed, uses the two rails for track circuits and the other two to provide energy for train movement (i.e., one feed and one return).

With the low-voltage and high-current operation of the DC railway system, the feeder substations are usually spaced at regular but short distances to maintain an acceptable voltage level along the line. The distance between substations is determined by the traction loads and the maximum permissible voltage drop, a 15 to 30% drop being allowed in practice. A separation of 5 to 6 km and 8 to 13 km between substations is the most appropriate for a 750 V and 1500V DC railway system respectively [47].

With proper filtering in the rectifier substations and traction equipments, DC railways are usually less susceptible to electromagnetic interference. However, the rail-to-earth leakage current leads to corrosion on the rails. The amount of leakage current depends on soil properties and earthing structure in the surroundings while the leakage rail-to-earth impedance is usually less than 1 Ω in a typical track circuit. The other drawback of the DC railways is the large voltage drop along the contact wire. A low power transmission efficiency is attained and thus more substations are required to sustain the train service.

To investigate the performance of railway operations with many aspects of power supply system design, train movement and rolling stock performance, different approaches and scales of modelling techniques are required to represent train activities. Two different simulation approaches, which provide the evaluation tool for the three levels of the *HTRC*, are described in the following section.

3.2 Railway system simulation

With the rapid development of computer hardware and software over the past few decades, simulation has been widely applied to study system performance in many applications [48] because of its cost effectiveness. Various modelling techniques have also been adopted for simulation of the railway system and its operation. The advantages of computer simulation, for both existing and new systems, include:

1. It is the only conceivable means to carry out the system performance study, when the system has not yet been built.
2. It is possible to carry out 'what-if' scenario studies with uncertain inputs whereas this is not allowed and indeed not possible in real life situations due to the safety constraints.

3. It provides a quick and flexible approach for the operators to estimate system performance with the given operational constraints, such as tractive effort, inter-station run-time and track topology.

Time-based and event-based models [49] are the two commonly used approaches to represent the system dynamics in software simulations. Time-based models represent a series of repetitive activities which are updated at each regular time step. This mechanism is more appropriate for modelling continuous and sometimes repetitive behaviour. Event-based models identify all the necessary events of the system and describe the order in which these events are allowed to happen. Event-based models are suitable for modelling interactive and concurrent systems.

To describe the dynamic behaviour of train operation in a railway system, railway simulation can be divided into the following three categories:

1. Train movement calculator [50] – it determines the position and speed of trains at each time interval for inter-station runs, in accordance with the track topology and tractive effort applied. The status of tractive effort on the trains depends on the last known train voltage and current.
2. Power network calculator [51] – it calculates the voltage and current for each train in the rail network with respect to their positions and operational modes, according to the given power feeding schemes.
3. Time scheduling calculator [52] – it provides a timetable for trains to serve passengers in a railway system. Train regulation and coordination, e.g., automatic route setting, are required.

In general, time-based models provide an evaluation tool for the train movement and power flow calculation in inter-station runs. Event-based model, however, is more likely to be used in scheduling applications to represent the trains' operation, such as traffic control at junctions [53].

3.2.1 Time-based model

Time-based models proceed by effecting calculations in regular time increments. To investigate the train operation between stations in a railway system, time is considered constant and the train movement is evaluated in each interval, in accordance with the track geometry and tractive effort. The train behaviour at that time is updated simultaneously and the system evolves continuously with time. The speed and distance profile of a train on an inter-station run is thus derived.

Since the train activity is continuously updated in regular time steps, it is easier to design and develop the models with a time-based approach. However, a time-based model requires not only a high computational effort to determine the detailed performance of train movement [54], but also a vast amount of memory in hardware. Computational demand can be reduced with a longer update time interval, but the level of detail of train operation will be compromised.

Since the detail of train performance (i.e., energy consumption and run-time) is essential at the train level of control in the *HTRC*, a time-based simulation model has been developed to provide the evaluation tool for the coast control effected by the *TBC*s. Details of the single train movement and the numerous operational constraints will be further discussed in the corresponding chapters.

3.2.2 Event-based model

With event-based models, the passage of time is considered at irregular intervals and the simulation proceeds from one event to another. The updates of the train's movement are not carried out synchronously. For a railway simulation, the events are linked to each other according to interactions between trains through signalling, power system and other system characteristics, where one event happens as a consequence of a prior event. As a result, the progress of the trains is denoted by a chain of events.

Since the details of the train operation between events are not taken into account in an event-based model and an event is only triggered when it is allowed to happen, the computational effort can be reduced substantially. Nevertheless, each event needs to be verified and the train movement is re-processed if certain conditions are found to be invalid. Great care is therefore needed in the development of event-based models. In general, event-based models involve the processes of data retrieval, matching and addition. System dynamics in the links between events are devised by the previously calculated performance values with additional control actions.

Since a large amount of data is essential for train operation and computational time is an important concern at the higher levels of control in the *HTRC*, two event-based models have been established to represent the traffic of multi-train operation in the *RTC* and *CTC* respectively. A state-space traffic flow model to represent the regional train coordination will be described explicitly in Chapters 6 and 7, whereas the methods to provide a view of the sequences of changes of headway will be explained in Chapters 8 and 9.

3.3 Data referencing

Since a large amount of data needs to be organised to represent the dynamic behaviour of train at each level of control in *HTRC*, a well-structured data representation is important for data retrieval and matching in real-time applications.

Key data involved in the three levels of control in the *HTRC* are described as follows:

A. *Train level of control – TBC*

System and operational constraints

1. Track topology – gradients, curvatures, speed restrictions and station locations.
2. Train characteristics (electrical part) – tractive effort and brake rate.
3. Train characteristics (mechanical part) – length, weight (including passenger loads) and frictional drag (varies with train speed).

A distance from the departed station for train to start coasting is the control variable.

B. *Regional level of control – RTC*

System and operational constraints

1. Nominal service headway in a region.
2. Nominal dwell and run times of trains in a region.
3. Number of stations/inter-station runs in a region.
4. Energy consumption with the corresponding run-time of a train in a successful

inter-station run.

5. Possible range of variation of train run-times with respect to the nominal schedule.

Dwell-times of trains at stations and run-times for trains in inter-station runs are the two control variables.

C. *System level of control – CTC*

System constraints

1. Number of regions along the line.
2. Run-times in each region with the nominal service headway.

The number of step changes of the headway and the corresponding values to change from one nominal headway to another are the control variables.

The above data are represented and organised in the following manner:

1. User-defined data type – data can be defined by users through an interface prior to simulation. For example, simulation time step, frictional drag, train's length and weight are often referred to in this data format in the train level of control.

2. Sequential data type – data is described sequentially in a lookup table and the size of the table depends on the level of detail in the application. For example, track is divided into a number of sections, and each section of track indicates the characteristics of gradient, curvature and speed restriction.

3. Linked data type – data is linked together in a hierarchical structure or evolves

in a tree format. For example, at the regional level of control, an event describes a set of arrival times of trains at stations, and the transitions from one event to another are linked together when they are allowed to happen.

With an efficient data representation, a relatively shorter simulation time for the decision-making process of train operation is obtained and further development or modification can be more conveniently incorporated.

Chapter 4
Train-Based Controller

This chapter focuses on the formulation of control actions to adjust an individual train's movement at the train level in order to meet inter-station run-time constraints imposed by the *RTC* in the *HTRC*. In general, inter-station run-time and energy consumption are at the two ends of a tug-of-war, where energy consumption is usually reduced at the expense of run-time, except for some extreme track topologies，and vice versa. The train-based controller, *TBC*, is designed for the purpose of real-time control of this trade-off. A set of dwell-times of trains at stations and inter-station run-times for trains within a region is determined by the *RTC* for the given service headway. Dwell-time control can be achieved for trains by simply lengthening and shortening the waiting times at stations, while run-time control actions for the next inter-station run are forwarded to trains when they stop at the stations each time. To meet the specified inter-station run-times on trains, coast control is the means to provide flexible and efficient control of individual trains.

4.1 Introduction

4.1.1 Methodology

Coast control [55,56] is an effective approach to achieve a compromise between run-time and energy consumption of train movement in inter-station runs by turning off the traction motors at a certain point (i.e., coasting point) if time is allowed to extend. Energy reduction is attained at the expense of run-time as the train spends less time in motoring mode. Another advantage of coast control is the reduction of maintenance cost of braking equipment, evidenced by Chui *et al.* [14]. Therefore, the capability to identify the starting point for coasting according to the current run-time provides the necessary flexibility for train operation.

Where coasting has been adopted, the current practice on most metro systems is to start coasting a fixed distance away from the last (departed) station, as in the MRT system in Singapore. The coasting points are pre-determined and only optimal with respect to a nominal operational condition, but not the current service demand which varies throughout the day. The extent of energy reduction is therefore somewhat limited and the advantage of coasting can only be fully achieved when coasting points are determined in real time, taking into account the imminent needs of the train service. A dynamic coast control is thereby preferable, so as to provide flexible train control according to the time-varying traffic conditions.

Train movement is governed by a large number of factors, such as track geometry, signalling, traction equipment characteristics, power supply and speed restrictions. Some of them are position-dependent whilst the others are speed-dependent. As the coasting control alters the speed profile of the train at a particular position, formulation of an analytical model to connect the coasting points and their corresponding run-time and energy consumption and then applying appropriate optimisation techniques is very much impractical, if not entirely impossible, because of non-linearities in the traction equipment characteristics and interactions between trains through power and signalling systems. Further consideration of uncertainties, like human behaviour and equipment delay, only makes matters more complicated.

Having ruled out an analytical approach, search methods are the potential candidates to attain the optimal coast points with respect to the real-time operational conditions. This chapter describes the design and implementation of the *TBC* with coast control, which is integrated into the on-board computer in trains for real-time applications, to attain a balance of run-time and energy demand of train movement with the constraints of inter-station run-times devised by the *RTC*. The solution space of coast control with single or multiple coasting point(s) under numerous operational constraints is presented. Golden, Fibonacci and Gradient searches are introduced for single coasting point control, while the Nelder and Mead method is applied for multiple coasting point control. For the purpose of demonstration, a maximum of two coasting points is used for the multiple coasting point control in this study.

A genetic algorithm (GA) method is adopted to determine the necessary number of coasting point(s) and the corresponding location(s) in an inter-station run with a hierarchical gene structure. In addition, a fast mutation scheme is introduced to improve the trade-off between computation time and the quality of the solution for real-time applications. The feasibility and performance of coast control with different search methods and track topologies will be thoroughly investigated through extensive studies, with the aid of a time-based single train simulator.

4.1.2 Literature reviews

In recent years, a number of advanced techniques have been introduced to improve individual train operation in terms of run-time and energy consumption. A fuzzy logic control (FLC) system [50] has been adopted to determine the necessary operational modes (i.e., motoring, coasting and braking) of trains in inter-station runs, according to the operational constraints and requirements, within which it allows safe and accurate train speed control. Simulation results reveal that the number of changes of notches to regulate the train speed is minimised with FLC and hence the ride comfort is improved in comparison with the traditional PID control. In addition, the tear-and-wear of the traction drives is reduced with less 'switching-on/off' the drive systems.

To achieve the optimal train control under the constraints of inter-station run-times, a GA-based method [57] was proposed to synthesise a coasting lookup table. The lookup table provides the locations where coasting, motoring or braking should commence. This study is primarily designed to minimise energy consumption in an inter-station run. With this coast control, a more flexible and efficient train operation can be obtained when compared with the fixed-point coast control. The result also shows that energy consumption obtained by coast control is even better than that achieved by fuzzy control. However, the suitability for real-time control and hence robustness to changing service demand were not discussed in the study.

An expert system approach of coast control has also been established [58,59] in which the loading effect and train voltage variations are both taken into account. In this application, the expert system was adopted as an advisory system to provide the coasting solution for train service control. However, the advantage of dynamic coast control cannot be fully taken as the coast control action is limited by the development of the built-in knowledge base. A change of knowledge base is required if different system conditions and operational requirements are expected.

The three studies show that a highly flexible and efficient train control is achieved, with either FLC or coast control, to meet the run-time requirement in inter-station runs. However, the feasibility and performance of FLC and coast control for real-time railway applications are not discussed. Further investigation on the capability of the controls with different track topologies is also recommended.

4.2 Problem formulation

4.2.1 Coast control

Train movement in an inter-station run includes three operation modes, motoring, coasting and braking. Run-time and energy consumption required for this simple pattern depend on the relative proportions of these modes of operation. Fig. 4.1 illustrates a typical inter-station run and it consists of 4 phases: (1) acceleration from a complete stop to maximum permissible speed; (2) maintaining the speed as close to the maximum permissible speed as possible; (3) coasting; and (4) deceleration to a complete stop by the application of the brake. Phases (1) and (4) are indispensable in an inter-station run and they depend on the train’s traction equipment characteristics. In a short inter-station run, phase (2) may not exist and coasting may start once the train sustains a certain speed. On the other hand, coasting may bring the train speed down to such an extent that re-motoring is necessary to take the train to the next station (i.e., repeated phases (1) and (3)). The location where coasting commences may thus drastically change the speed profile of the inter-station run, leading to a wide range of possible combinations of run-times and energy consumption.

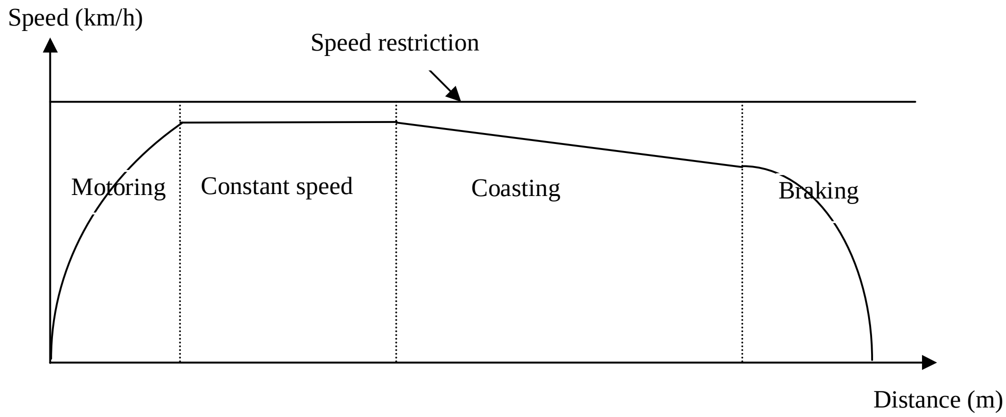


Fig. 4.1 Speed profile of a simple inter-station run

A typical flat-out inter-station run sees a train accelerating to maximum speed from a station, maintaining the speed as accurately as possible until it is necessary to brake to a halt for the next station. While the train is travelling very close to the maximum permissible speed throughout the trip, the running time is the shortest but the energy consumption is the highest. A flat-out run is always necessary during rush-hours and recovery of train service from disturbance, whereas it also serves as a reference to the operators for relaxation of the train schedule during off-peak hours.

When coasting is allowed, the supply to the traction motors is turned off once the train accelerates above a certain speed. The momentum of the train then carries it through and the brake is still needed to bring the train to a stop at the next station. Inter-station run-time is longer but energy saving is possible as the train spends less time on motoring. The longer run-times can be easily absorbed during off-peak hours when the train frequency is lower. In fact, a longer run-time is preferable to a longer station waiting time because of the possible additional energy consumption when the train-doors have to be kept open at stations.

Fig. 4.2 shows the speed profile of a flat-out run between two stations, as well as the trajectories for four different coasting point assignments. It is evident that different coasting points alter the speed profile significantly. One of the speed profiles even indicates re-motoring. Fig. 4.3 illustrates the run-time differences for four coasting points. Generally, the run-time is extended when coasting is allowed to start sooner. The run-time variation resulting from coast control provides headway regulation and possible energy reduction for the operators. Depending on the traction drive system and track geometry, an energy saving of 30% can be attained with only a 5% increase in run-time [60].

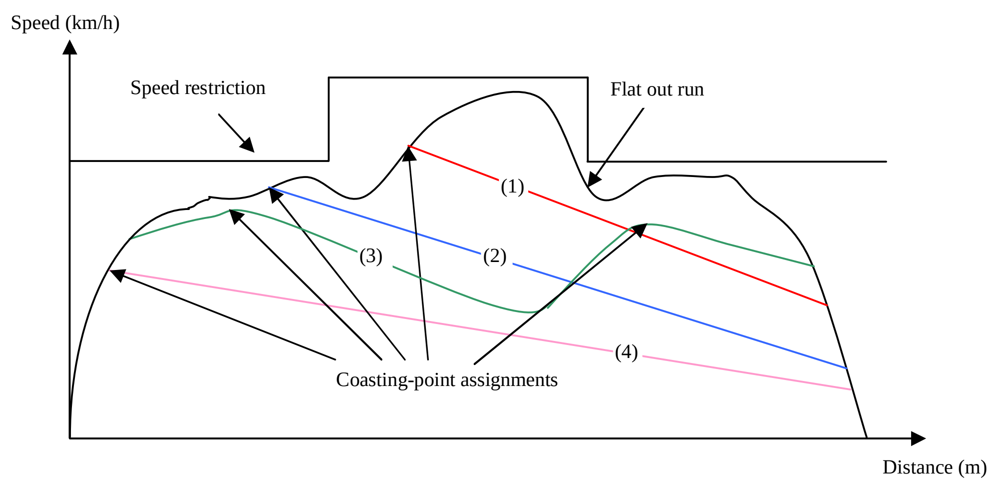


Fig. 4.2 Speed profiles of flat-out run and some possible coasting-points

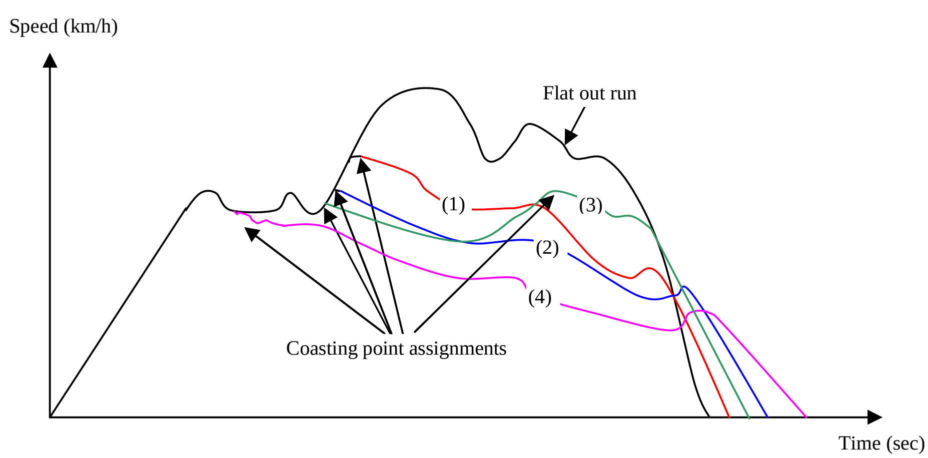


Fig. 4.3 Run-time extensions with some possible coasting points

4.2.2 Solution space

Theoretically, any point between the two stations is a possible coasting point. Fig. 4.4 illustrates that the solution set X may contain all points between A and D. Although a certain distance resolution is imposed so that X is a finite set, the solution space can be further reduced and confined by some of the train's operational constraints. A number of subsets of X can be excluded from the space to make the searching process more feasible.

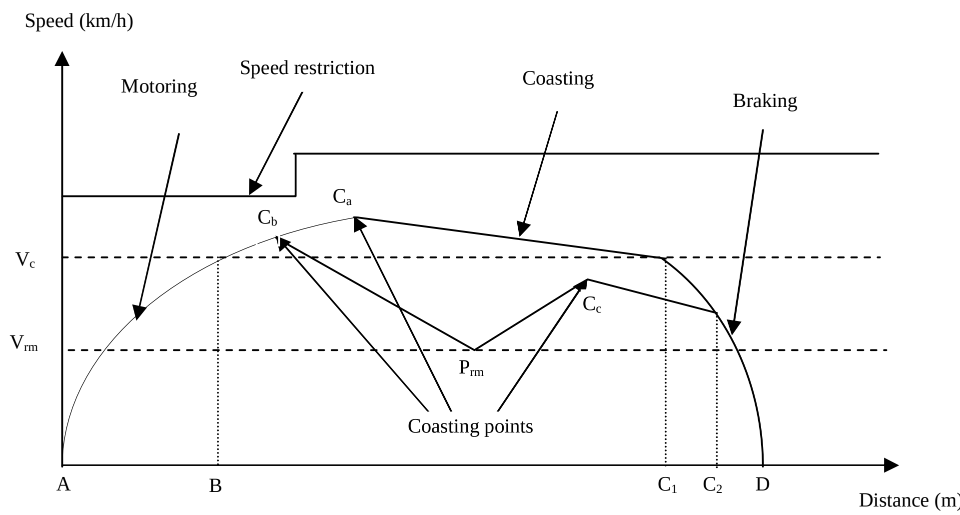


Fig. 4.4 The range of possible coasting point

If a train starts coasting at a low speed, it is very likely that it will re-motor before it reaches the next station. The unnecessary turn-on/off of the traction drives is not desirable in practice because it will hasten the wear-and-tear of both the electrical and mechanical components. Re-motoring should thus be kept to a minim if possible. In addition, energy reduction is not guaranteed with excessive re-motoring. To ensure that the train has sufficient momentum prior to coasting, a train is allowed to start coasting only when it reaches a minimum coasting speed V_c. In other words, coasting is prohibited from points A to B, as indicated in Fig. 4.4, and the set of the possible points between A and B, denoted as Ψ, can be excluded from X. The size of Ψ is thus determined by V_c, which is a pre-defined system parameter.

When the inter-station distance is so short that only one coasting point is needed, run-time decreases and energy consumption increases monotonically as the coasting point shifts from the starting station to the next. The necessary coasting point to trade off run-time and energy consumption can be attained by simple optimisation techniques, except for extreme track geometry and speed restrictions, because there are no local optima clouding the global one.

At the other end of the speed profile, after a train has entered the braking region, it cannot start coasting or it will overshoot the station and miss the stopping mark completely, unless excessive braking is applied, which may over-strain the braking system and, more importantly, cause passenger discomfort. Therefore, at the approach to a station, there is a distance, between points C_1 and D, or C_2 and D, where the set of points Ω can also be discounted for the searching process. Given a service braking rate, Fig 4.4 shows that under two different coasting trajectories, individual braking points C_1 and C_2 are required to lead the train to stop at a station. Hence, point C_x is not a fixed parameter and largely depends on the train's trajectory before the train starts braking, where the coasting point plays a decisive part. During the search, point C may be defined as a specific point on the braking profile for simplicity. Alternatively, it can be made a variable and estimated by train movement simulation according to the most up-to-date location of the coasting point in the search.

Inter-station distances vary within the same railway line and multiple coasting points may be required for longer inter-station runs. However, there are no specific rules to determine the number of coasting points which inevitably turns the solution space multi-dimensional. Fig. 4.5 shows the solution space of a typical 2-coasting point control. Given the speed profile of various 2-coasting point assignments, the cost can be obtained by the train simulator. From Fig. 4.4, V_{rm} is the parameter that allows multiple coasting point identification. When the train speed falls below this specific value from coasting, it is allowed to re-motor to ensure sufficient momentum to go on. The locations of P_{rm} (i.e., at which the train starts to accelerate again) and C_c (i.e., at which the train starts to coast again), change dynamically with the location of the last coasting point, C_b. It should be noted, when multiple coasting points are allowed, that the energy consumption required for train operation with a given inter-station distance may not be lower than that with a single coasting point as the train spends more time in motoring mode. V_{rm} is not used to eliminate the solution space in the searching process as it just ensures the next coasting point is located for the purpose of multiple coasting point identification. The solution space for the next coasting point varies with the location of the previous coasting point and the location of the first coasting point inevitably affects that of the second and so on.

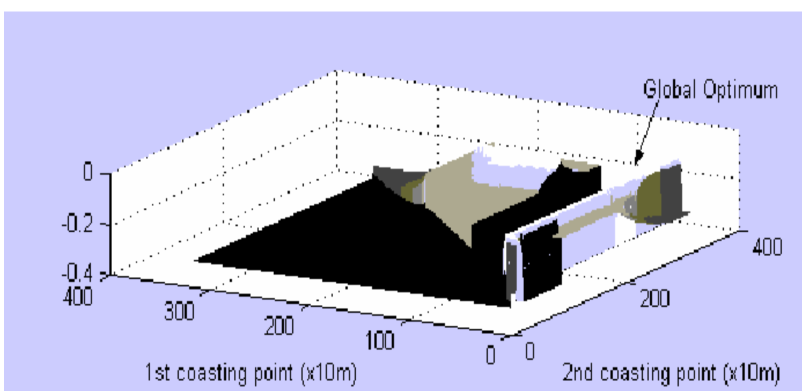


Fig. 4.5 A typical solution space with 2 coasting points

From the viewpoint of real-life application, there is a wide range of locations to start coasting(s) and each will produce different run-time and energy consumption. In other words, given the required run-time, locating the required coasting point(s) quickly is the essence of this searching problem.

4.2.3 Objective function

To evaluate how close the chosen coasting point(s) is/are to achieving the desired run-time and energy consumption in inter-station runs, an objective function is necessary and this is quantified in Eqn. 4.1. The objective function enables the adjustment of the relative weights for the two conflicting factors, energy consumption and run-time. A smaller value '*F*' implies the solution is closer to the desired requirement. The relative effect of energy consumption on '*F*' is ignored when the energy consumption provided by the solution is lower than the desired value. In other words, $\frac{E_g - E_D}{E_D}$ is set to zero when $E_D > E_g$. Since the run-time may be either above or below the desired values in a particular run, the absolute sign is in place to nullify the polarity effect. Other definitions of *F* are equally valid if other considerations are taken into account.

$$F = W_T \times \left| \frac{T_g - T_D}{T_D} \right| + W_E \times \mathrm{sgn}\left(\frac{E_g - E_D}{E_D} \right) \tag{4.1}$$

where $W_T + W_E = 1$

W_T is the weighting factor for run-time

W_E is the weighting factor for energy consumption

T_D is the desired run-time

E_D is the desired energy consumption

T_g is the run-time achieved by the updated solution

E_g is the energy consumption attained by the updated solution

4.3 Search methods

To conduct real-time coast control in the *TBC* for a specified run-time constraint imposed by the *RTC*, classical search methods and heuristic approaches are both adopted to look for the optimal solution efficiently on the solution space. The search problem can be simply divided into two categories – single and multi dimensional searches. Three classical methods and one heuristic approach are introduced to identify the coasting point(s) in inter-station runs.

In general, there are two major approaches – direct and indirect searches. With the direct search or numerical methods, the coasting solution is obtained in a step-wise manner and the value of the cost function *F* is improved at each step. In addition, the direct search methods do not require an explicit evaluation of any partial derivatives of the function, but rely on values of the cost function *F*, obtained in previous iterations. These methods basically use the cost function values to obtain numerical approximations to the derivatives of the cost function. Dichotomous search, Golden section search, Fibonacci search and Simplex method are examples of direct search methods.

Indirect or analytic searching methods, on the other hand, attempt to reach the necessary coast point by calculation, without test or guess. It is based on the analysis of the special properties of the cost function *F* at the position of the extremum. In the simplest case, the tangent plane at the optimum is horizontal, which implies the first partial derivatives of the cost function exists [61], and it can be defined as follows:

$$\left.\frac{\partial F}{\partial x}\right|_{x=x_i} = 0 \tag{4.2}$$

where x_i is the optimum coasting point under the specified operational constraints for coasting control to achieve service regulation. The gradient method is one of the examples of the indirect search and it selects the search direction $\boldsymbol{d_i}$ using the polarity of the value of partial derivatives of the cost function $\boldsymbol{F}$ with respect to the independent variables $\boldsymbol{x}$, and the information gained from previous iterations.

Simulated Annealing (SA), Genetic Algorithms (GA), Tabu Search (TS) and Simulated Evolution (SE) are well-known heuristic approaches [62]. They solve problems by means of a trial-and-error approach with certain rules-of-thumb or guidelines when an analytic approach is impractical. Heuristic methods often have an intuitive justification. One of the main differences between classical and heuristic methods is that a mathematical model is essential in classical methods whilst it is not necessary for heuristic approaches. Moreover, classical methods always provide the best solution but heuristic approaches obtain a good, rather than the best, solution satisfying the defined constraints. Nonetheless, a common feature of these two approaches is their iterative nature.

4.3.1 Classical methods

4.3.1.1 One-dimensional search

Since the distance between stations is rather short in metro systems (i.e., the solution space is small), a single coasting point is usually adopted for service regulation as there is not enough room to accommodate multiple coasting points. In general, run-time decreases and energy demand increases when the coasting point is further away from the starting station. Hence, this single-variable problem can be simply solved by classical optimisation methods.

With one-dimensional search in coasting control, the Bi-section (Golden and Fibonacci) and Gradient-based searching methods are deemed to be appropriate. With the Gradient-based methods, the cost function F has one variable (i.e., coasting point) only and the inter-station run-time is also a function of the coasting point. However, the functional form of F may not be well defined. In other words, the continuity or differentiability of the function are not guaranteed. Hence, a point-by-point evaluation of F on the solution space is required. However, the applicability of the Bi-section method is limited to unimodal functions (i.e., the function has only one global optimum point).

Golden section search

With this algorithm on coast control, the costs of two initial coasting points are determined and then used for further search for a new coasting point. These two coasting points are obtained from either end on the solution space with the spacing of a golden ratio. The basic idea of the Golden section method [63,64] is that the solution space is divided into two unequal parts, whereas the ratio of the larger of the two segments to the total length of the interval should be the same as the ratio of the smaller to the larger segment.

Assume the solution space consists of a length z which is composed of two segments z_1 and z_2, as shown in Fig. 4.6, the Golden section requires that

$$\frac{z_1}{z} = \frac{z_2}{z_1} \tag{4.3}$$

$$z = z_1 + z_2 \tag{4.4}$$

Eqn. 4.3 gives

$$z_1^{\ 2} = z \times z_2 \tag{4.5}$$

By substituting for z from Eqns. 4.4 into 4.5 and normalizing with z_2^2, the following equation is obtained.

$$\left(\frac{z_1}{z_2}\right)^2 + \left(\frac{z_1}{z_2}\right) = 1 \tag{4.6}$$

This quadratic Eqn. 4.6 can be solved for the ratio z_1/z_2. The positive root is

$$\left(\frac{z_1}{z_2}\right) = 0.618033989 \tag{4.7}$$

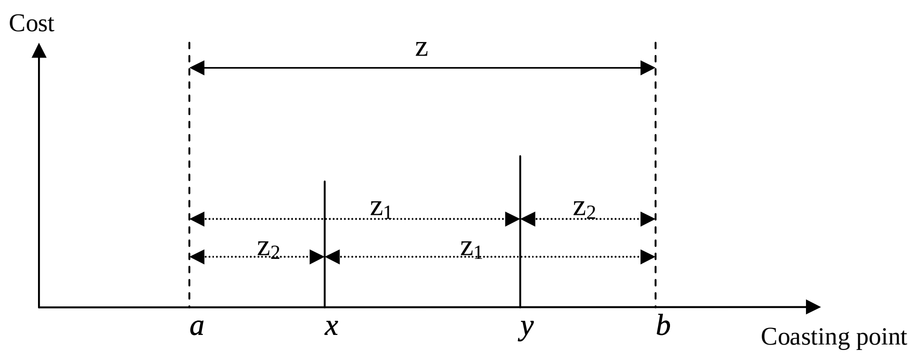


Fig. 4.6 Golden section search

Now assume the fitter coasting point is located between points a and b in a search, the searching process with the Golden section method is listed as follows:

1. Two initial coasting points, x and y, are placed with the 'golden ratio' spacing (i.e., 0.618) from either end on the solution space between a and b, as shown in Fig. 4.6, the solution space z will then be reduced to a fraction of 0.618.
2. Assume $F(x)$ is smaller than $F(y)$, y replaces b and the new solution space z_1 becomes (a, y).
3. The process is repeated and the new solution space z_1 is further reduced by the golden ratio until the obtained coasting point satisfies the expected run-time requirement of train operation.

The last estimate is placed in the middle of the latest confined solution space.

Fibonacci search

The concept of Fibonacci search [64,65] is very similar to the Golden section search. The main difference is that the reduction ratio on the solution space at each iteration is fixed at 0.618 with Golden search, whilst the reduction ratio is predetermined and optimized according to the given number of iterations in a Fibonacci search. The arrangement of the search point (i.e., coasting point) within the new search interval is shown in Fig. 4.7.

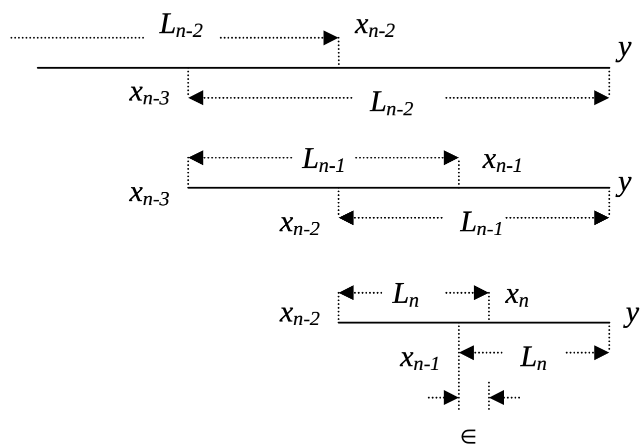


Fig. 4.7 Sequence of uncertainty intervals in a Fibonacci search

To help illustrate how the reduction ratio at each iteration is devised, the iterations are shown in reverse order where x_n and x_{n-1} are the last pair of estimates. Point 'y' is one end of the solution space and the successive new coasting points obtained (i.e., x_{n-2}, x_{n-1} and x_n) in iterations are assumed to be a fitter solution for the sake of simplicity. Hence, the new solution space becomes (x_{n-3}, y), (x_{n-2}, y) and (x_{n-1}, y) through the iterations. The searching process repeats until the obtained coasting point satisfies the expected run-time requirement of the train operation.

From Fig. 4.7, the interval of uncertainty L_{n-1} is (x_{n-2}, y) and the final search interval is defined as,

$$L_n = \frac{L_{n-1} + \varepsilon}{2} \tag{4.8}$$

L_n is the length of the interval of uncertainty after the nth iteration and ε represents the smallest distance by which two evaluations may be separated and still be distinguished from one another. The symmetry requirement for the search interval is that

$$L_{n-2} = L_{n-1} + L_n \tag{4.9}$$

Combining Eqns. 4.8 and 4.9,

$$L_{n-2} = 3L_n - \varepsilon \tag{4.10}$$

It is possible to work backwards to determine the required size for any intermediate interval of uncertainty.

$$L_{n-3} = 5L_n - 2\varepsilon$$

$$L_{n-4} = 8L_n - 3\varepsilon$$

It can then be generalised as

$$L_{n-k} = F_{k+1}L_n - F_{k-1}\varepsilon \tag{4.11}$$

The coefficients F_{k+1} and F_{k-1} can be obtained by

$$F_{k+1} = F_k + F_{k-1} \qquad k = 1,2,3,\ldots n \tag{4.12}$$

and $F_0 = F_1 = 1$.

The fibonacci method provides a specific reduction ratio for the solution space at each iteration, and the interval of uncertainty can be used to plan the evaluation spacing if the maximum number of iterations is fixed in advance. In general, Fibonacci search method retains one of the two ends of the interval from the previous iteration and therefore requires only one new estimate of the coasting point in the next iteration.

Gradient based search

The Gradient-based method [66] must adopt a certain derivative to enable the search. In this application, the derivative should relate the control variable (i.e., coasting point) to the consequence (i.e., run-time). Hence, the gradient required is given in Eqn. 4.13. The two initial coasting points are chosen randomly and the gradient is obtained with these two points. The search direction of the updated coasting point depends on the polarity and the magnitude of the gradient,

$$Gradient = \frac{\Delta\ Run\ Time}{\Delta\ Coasting\ location} \tag{4.13}$$

The length of the next step can then be calculated by,

$$Step\ length = \left\{Gradient^{-1} \times \left[Run\ time_{Flat-out} - Run\ time_{Expected}\right]\right\} \tag{4.14}$$

The new coasting point can be obtained by the following equation

$$New\ coasting\ location = Old\ coasting\ location + Step\ length \tag{4.15}$$

In general, the step size varies with the difference between the run-time obtained by the latest iteration and the expected one. Therefore, the Gradient method is likely to require a smaller number of iterations to achieve the same level of solution quality, compared with other methods. However, the drawback of this algorithm is that the step size and search direction cannot be defined in the searching process when there is no change in run-time obtained from two successive iterations (i.e., the gradient is zero) and the searching process will be terminated. For example, Fig. 4.8 shows that the train is forced to operate in coasting between points ***a*** and ***b*** even in a flat-out run because the train speed is confined by the lower speed restriction. The speed profile of the train behind point ***b*** is then likely to be the same and so is the run-time, even if the coasting point is chosen at any location between points ***a*** and ***b***.

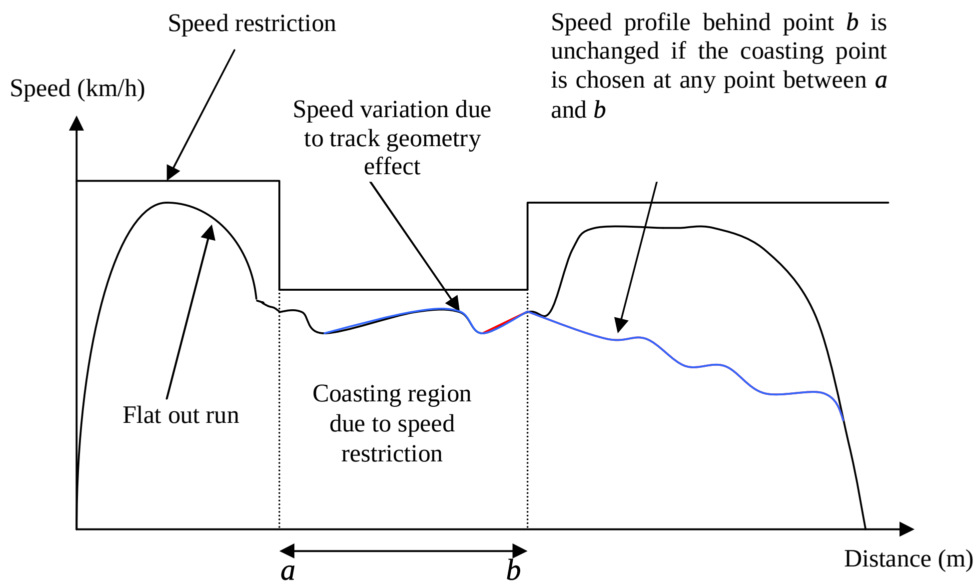


Fig. 4.8 Coasting assignments without adjustment on run-time

4.3.1.2 Multi-dimensional search

Nelder and Mead method

This algorithm is an extension of the simplex method for the purpose of multi-dimensional search. A set of (n+1) mutually equidistant points in an n-dimensional space is known as a regular simplex. Thus, in a two-dimensional problem, the simplex is an equilateral triangle and in a 3-dimensional space, it is a regular simplex tetrahedron. The idea is to compare the return value of the cost function at the (n+1) vertices of the simplex and move the simplex towards the optimum point during the iterative process. The vertices of the simplex represent the multiple coasting points on coasting control and the vertex is in a two-dimensional form (i.e., a pair of coasting point) in this application for the sake of simplicity. The original simplex method maintains a regular simplex at each stage. Nelder and Mead [67] proposed several modifications to the method, which allows the simplices to become non-regular. The result is very robust and the algorithm is extremely powerful, provided that the number of variables does not exceed five or six as the shape of the simplex becomes complicated and, hence, the iterative process is slowed down.

In order to locate multiple coasting points, three basic operations, reflection, expansion and contraction, are applied to reshape and resize the simplex. The simplex takes on a new shape and/or size when a vertex is replaced by a better one with respect to the three factors α, γ and β corresponding to the three operations. The details of the application of this algorithm are depicted as follows:

1. Select initial guess of three pairs of coasting points and evaluate their cost with these points.

   i.e., $F_1=F(x_1)$, $F_2=F(x_2)$ and $F_3=F(x_3)$

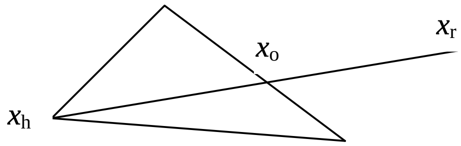
   where x_i represents one pair of coasting points and F indicates the cost of the solution.

2. Arrange cost in ascending order F_s, F_g and F_h with the corresponding pairs x_s, x_g and x_h, and x_s is the fittest pair. The lower the cost value, the better is the quality of the solution.

3. Find the mid-point of x_g and x_s and let it be x_o, then calculate its cost value $F(x_o)=F_o$.

4. Move away from x_h in x_o to obtain the reflection point x_r and its cost value

$F(x_r) = F_r$ as shown in the following diagram.



Simplex generation by reflection

If α>0 is the reflection factor, x_r can be calculated as follows,

$$x_r - x_o = \alpha(x_o - x_h)$$

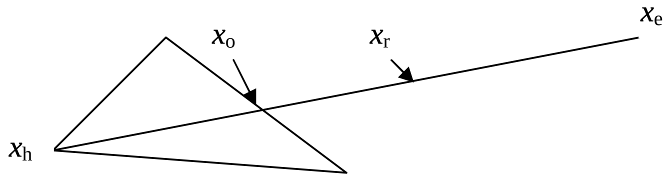
$$\text{i.e., } x_r = (1+\alpha)x_o - \alpha x_h \tag{4.16}$$

5. Compare F_r with F_s

a. if F_r is smaller than F_s, the lowest cost value is obtained. The direction from x_o to x_r is a good one to move along. Therefore, a further expansion in this direction to find x_e and evaluate its cost F_e=$F(x_e)$. The operation of expanding the simplex is illustrated as follows. If γ>1 is the expansion factor, x_e can be defined as follows,

$$x_e - x_o = \alpha(x_r - x_o)$$

$$\text{i.e., } x_e = \gamma x_r + (1-\gamma)x_o \tag{4.17}$$



Simplex generation by expansion

i. if F_e is smaller than F_s, replace x_h by x_e and test the (n+1) points of the simplex for convergence to the minimum. The process terminates if the expected cost value is achieved, otherwise return to step 2.

ii. if F_e is larger than F_s, it is evident that x_e has moved too far in the direction of x_o to x_r. Therefore, replace x_h by x_r which has an improvement from step 5a. The process terminates if the expected cost value is attained, otherwise return to step 2.

b. if F_r is larger than F_s but lower than F_g, x_r has an improvement on the two worst points of the simplex. Thus, replace x_h by x_r and then test for convergence, otherwise return to step 2.

c. if F_r is larger than F_s and F_g, it is evident that the search direction is not valid and go to step 6.

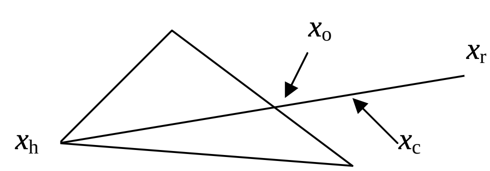
6. Compare F_r with F_h.

a. if F_r is smaller than F_h, replace x_h by x_r and F_h by F_r. x_c can then be obtained by contraction

$$\textit{i.e.,} \quad x_c - x_o = \beta(x_r - x_o)$$

$$x_c = \beta x_r + (1-\beta)x_o \tag{4.18}$$

where $\beta(0<\beta<1)$ is the contraction factor.

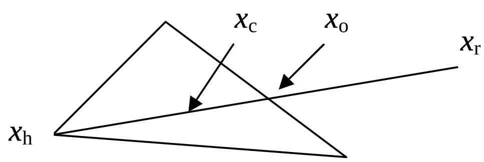


Simplex generation by contraction

b. if F_r is larger than F_h, the move goes too far in the direction x_h to x_o. Therefore, x_c can be found by contraction.

$$\textit{i.e.,}\ x_c - x_o = \beta(x_h - x_o)$$

$$x_c = \beta x_h + (1-\beta)x_o$$



Simplex generation by contraction

7. Compare F_c with F_h.

   a. if F_c is smaller than F_h, replace x_h by x_c and F_h by F_c. Check for convergence and otherwise return to step 2.

   b. if F_c is larger than F_h, it is proved that a new point, which is smaller than F_h, cannot be found simply by reflection, expansion and contraction method. Thus, go to step 8.

8. Reduce the size of the simplex by halving the distance of each point of the simplex from x_s which has the lowest function value. x_i can be replaced by

$$x_s + \frac{1}{2}(x_i - x_s) = \frac{1}{2}(x_i + x_s) \tag{4.19}$$

Then calculate F_i for i=1,2, …, (n+1), test for convergence and return to step 2 if the expected cost cannot be reached.

The search process will repeat until the new vertex satisfies the expected run-time requirement in an inter-station run. There are no specific rules to assign the factors of expansion, reflection and contraction. Nevertheless, these three factors cannot be too small because a fast convergence may not be attained, nor can it be too large because the generated solution (i.e., a pair of coasting points) may be out of the boundary of the solution space. Details of the setting of α, γ and β are given in Appendix A.

4.3.2 Heuristic methods

Genetic algorithms (GAs) have already found many successful applications in management science and scheduling problems [68-70]. Encouraging results have also emerged from the attempts to deploy GA on railway-related problems, such as power system monitoring [71], signalling design [72] and conflict-area traffic control [73,74]. In addition, a preliminary attempt of applying GA in coasting control has shown promising results [57], where the number of coasting points was pre-determined. GA may not be able to pinpoint the optimal solution, but it can present a near-optimal solution whose cost can be improved upon if more time is given for further evolution. A GA-based coast control therefore has a potential to conduct on-line train regulation in inter-station runs.

Genetic algorithm (GA) [75,76] is an evolutionary algorithm that models biological processes to optimise a cost function. It is applicable for solving one and multi dimensional search problems. It allows a population composed of many individuals (solution of problem) to evolve to a generation that optimises the cost. There are two basic steps to have the chromosomes evolved through successive generations, selection and replacement. The former is to decide which chromosomes in the generation are deemed to be fit to produce off-spring whilst the latter is to allow the chromosomes with the worst cost to vanish in order to make room for the better off-spring to compete and survive. A new generation thus consists of the surviving and the reproduced chromosomes in the population. An individual with better cost may be generated through the natural evolution process by mutation, crossover or other possible evolution methods.

4.3.2.1 Essential components of GA

Using GA, every possible solution (i.e., chromosome) in a problem is considered as an individual. Each chromosome contains a set of genes which describes the required information for the problem. The cost function is the selection criterion to determine the fittest chromosomes for further evolution. The following are the important components of GA:

1. Encoding of the possible solutions or chromosome structure

   This mechanism is used to convert solutions into a specific format for implementation in GA. Chromosome evolution can be enhanced by an efficient representation of the solution.

2. Initial population

   An initial population of solutions is defined as the reference for the further chromosome reproduction. Initial generation is often selected randomly from the population. In general, there is no specific rule on the size of the initial generation. However, the space of solutions can be limited by a certain constraint in a particular search problem. The constraint is normally dependent on the nature of the problem. The number of evolutions can be reduced if a better initial generation is used.

3. Cost function

A cost function is used to determine how good the solution is. It gives out a number, which alone is meaningless unless it is compared with the numbers from other solutions. The cost function plays an important role in GA since it indicates how good the solution is. Therefore, a meaningful, sensible and well-structured cost function is essential.

4. Selecting solutions for producing new solutions

Chromosomes are ranked in an order within the current generation based on their return cost values. Fitter chromosomes will have a better chance of being selected for further evolution. In this application, the fitter chromosome results in a lower cost value. However, producing offspring continuously will increase the size of the population and it will in turn decrease the search efficiency. Therefore, chromosomes with a higher cost value are to be deleted to make room for the offspring.

5. Operators to create new solutions

Once the fitter solutions are selected, off-spring can be created by crossover, mutation and other methods. This process is called evolution and it may lead to better and worse off-spring. Crossover normally takes two parental chromosomes and creates off-spring with a mixture of both parents' genetic characteristics. A common form of crossover is to allow one cut-off point at the parents so that each parent chromosome is divided into two segments, which are then swapped across to form two new offsprings. For example, when the two parents are in binary format:

Parent 1: 11010|00110 bin Parent 2: 01011|10100 bin

Crossover takes place at the specific points, as indicated by the vertical bar, and yields two new off-springs:

Off-spring 1: 11010|10100 bin Off-spring 2: 01011|00110 bin

Mutation is a process for a single gene to change its characteristics and it requires two parameters, the number of bits for mutation and which binary bits in the chromosome are to be reversed. These variations make some individuals more capable to survive and reproduce. Nevertheless, the number of crossover points/mutated bits to be adopted and their corresponding locations in a chromosome depend on applications.

The basic flow of GA is illustrated in Fig. 4.9. At the beginning, an initial population is generated. The fitter individuals have a better chance of evolving. Offspring are then created by crossover and mutation. The search process repeats until the latest solutions satisfy the desired conditions or result in the maximum number of generations.

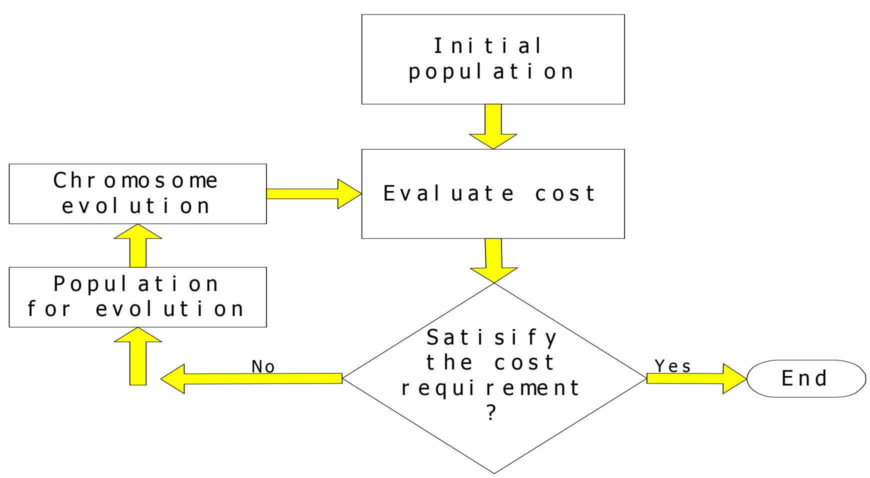


Fig. 4.9 Flow of GA

4.3.2.2 Coast control with GA

In GA-based coast control, a coasting point can be represented in binary, octal, decimal and hexadecimal format for ease of evolution. To ensure the new offspring are within the boundaries of the solution space, a gene in a chromosome is defined in binary format to represent the relative position to start coasting between stations in this application. Resolution on the coasting point representation depends on the number of binary bits used. In addition, the number of bits used for the coasting point representation is directly related to the distance between stations, which is in general over a kilometre. Further, a train may still travel at a few metres per second even when it is just slowing down on the approach to a station. Thereby, the resolution on the coasting point representation up to a metre is sufficient.

To further enhance the flexibility of the coast control with GA, a hierarchical structure of gene representation (i.e., coasting points) is introduced to determine the number of coasting point(s) and its(their) corresponding location(s).

4.3.2.3 Single-coasting-point control with simple GA

Single coasting point control is applied to meet the service demand in this application. The location of the coasting point and its distance from the departed station is encoded in binary form. The number of binary bits required is determined by the inter-station distance. A single gene is embedded in a chromosome in this application and a gene represents a single coasting location in an inter-station run. Table 4.1 shows an example of the chromosome representation of single coasting-point control.

Table 4.1 – Chromosome representation of single coasting point

Inter-station distance (m)	Number of bits required	Chromosome representation of coasting point
1200	11	00111110100 (500m)

4.3.2.4 Multiple-coasting-point control with simple GA

Chromosome representation for multiple coasting points is similar to that for the single coasting point control, but a larger number of binary bits is required to represent the two coasting locations in this application. For the sake of simplicity, two coasting points are assumed (i.e., two genes) in the following descriptions and they are integrated in a single chromosome. Two types of chromosome representation of multiple coasting point control are proposed, as shown in Table 4.2.

The absolute distance of the locations of the first (Gene 1) and second (Gene 2) coasting points from the departed station are encoded. Genes 1 and 2 then form a chromosome as the coasting solution. Alternatively, the separation between the first and second coasting points can be used to represent Gene 2 with the relative distance representation.

Table 4.2 – Chromosome representation of absolute and relative coasting point

	Absolute distance representation	Relative distance representation
1st coasting point (Gene 1)	011001000000 (1600m)	011001000000 (1600m)
2nd coasting point (Gene 2)	100111000100 (2500m)	001110000100 (2500-1600=900m)
Chromosome	011001000000 | 100111000100	011001000000 | 001110000100

4.3.2.5 Hierarchical genetic algorithm (HGA)

In the previous sections, the number of coasting point(s) required, either one or two, is fixed prior to the simulation. The capability to select the number of coasting points is not available in a simple GA. From the viewpoint of applications, a fixed number of coasting points does not necessarily provide a flexible train control. In general, single and multiple coasting points are also possible for train service control. In practice, the operators usually determine the necessary number of coasting point(s) to maintain the train schedule according to their experience. A dynamic coasting control according to the system conditions and operational requirements is preferable. A flexibility of up to two coasting points is allowed in order to explore the feasibility of dynamic coasting control with HGA.

HGA approach [77] is adopted here to represent both the number and location(s) of coasting point(s) in a chromosome. HGA can provide the coast control information in a hierarchical manner according to the current traffic conditions. Chromosome representation of HGA is similar to the multiple coasting point control. Genes 1 and 2 represent the locations of the two possible coasting points. An additional bit, Gene 3, is introduced to identify the number of coasting points required, as shown in Table 4.3. HGA allows two coasting points when this single bit is ‘1’, and a single coasting point otherwise. In other words, Gene 2 is easily called for by this multiple coasting point control identifier. To further enhance the coast control of train operation, it is always possible to use two or more bits to identify more coasting points in an inter-station run when necessary.

Table 4.3 – Chromosome representation of HGA

1st coasting point (Gene 1)	011001000000 (1600m)
2nd coasting point (Gene 2)	100111000100 (2500m)
Multiple coasting point control identifier (Gene 3)	0/1
Chromosome	011001000000 | 100111000100 | 0/1

4.3.2.6 Minimum-Allele-Reserve-Keeper (MARK)

Crossover and mutation are the two commonly used genetic operators to evolute new offsprings from parent(s) in GA. The role of crossover is to combine genetic characteristics from different individuals in the population. Since crossover proceeds by recombining information from parents, offspring only contains the information that was present in the parents. Crossover may lead to premature convergence easily as it never creates new information to the offspring, if the solution is trapped at the local optimum already. Further, GA increases the effort of search for the optimal solution with crossover when GA starts to approach the optimum of the search space.

Mutation is the other general genetic operator for evolution in GA and it creates offspring by performing a random modification on an individual. Although the chance for the offspring to escape from the local optimum is improved in mutation, because it leads to more open space for evolution, the likelihood of the search to move away from the solution is higher when offspring happen to be closer to the solution. In other words, the classic mutation is sometimes counter-productive when the GA begins to reach a near-optimal solution in evolution.

To meet the demand of the real time application in train level of control, a new mutation scheme in GA, Minimum-Allele-Reserve-Keeper (MARK) [78], is introduced as a genetic operator. With MARK operation, a minimum reserve (MARK rate) of each binary value at the same bit positions must be kept within the population. In other words, the chromosomes of each generation have a minimum number of '0' and '1' at each bit position. Since MARK makes minimal disturbances to the population and while providing modification on an individual like classic mutation, a less destructive change in the gene can be accomplished in evolution. Hence, MARK avoids excessive bit-inversions and it also prevents the offspring from going into local optima.

With MARK, the minimum number of bit '1' and bit '0' in the population are governed by a rate α. A_i and B_i are the ratio of bit '1' and bit '0' at column '*i*' in the mating pool respectively. The operation of MARK is illustrated in Fig. 4.10, in which the size of the population for evolution is set at 10 and the rate α of 0.2 is supposedly to be attained at each column in this example. Fig. 4.10 shows that the ratio of bit '0' in column 3 (i.e., B_3=0.1) and bit '1' in column 1 (i.e., A_1=0) are less than 0.2 prior to evolution. A single bit '1' in column 3 and two bit '0' in column 1 are thus randomly chosen and mutated respectively to achieve the rate of α=0.2.

Chromosome

	0	1	1	0
	0	0	1	1
	0	1	1	0
	0	1	1	1
	0	0	1	0
	0	0	0	1
	0	1	1	1
	0	0	1	1
	0	0	1	1
	0	1	1	1
A_i	0	0.5	0.9	0.7
B_i	1	0.5	0.1	0.3

A_1

B_3

Evolution

Chromosome

	0	1	1	0
	0	0	1	1
	1	1	1	0
	0	1	1	1
	0	0	1	0
	0	0	0	1
	0	1	0	1
	0	0	1	1
	1	0	1	1
	0	1	1	1
A_i	0.2	0.5	0.8	0.7
B_i	0.8	0.5	0.2	0.3

Fig. 4.10 Mark operation

4.4 Software implementation

4.4.1 Train simulator with coast control

To conduct coast control with different searching methods, a single train simulator [79] is an important tool for cost evaluation in the coasting-point identification process. A time-based train movement simulator has been implemented for this purpose. The principal loop in the simulator is the incrementing time, and detailed descriptions of train movement such as speed, distance and operation mode are devised through successive simulation time steps.

At the beginning of each update period, it is assumed that the position and speed of the train are known. The movement simulator examines these new positions and speeds with respect to track-based data, and determines the possible train modes (motoring, coasting and braking) for the next update period. Once the train mode is established, the performance of the train can be calculated, taking into account track details, train speed and position. This requires a representation of track gradient and curvature, motor characteristics and train loading. Finally, the speed and position of the train is updated and then used as the initial values for the next time update. The structure of the single train simulator is given in Fig. 4.11.

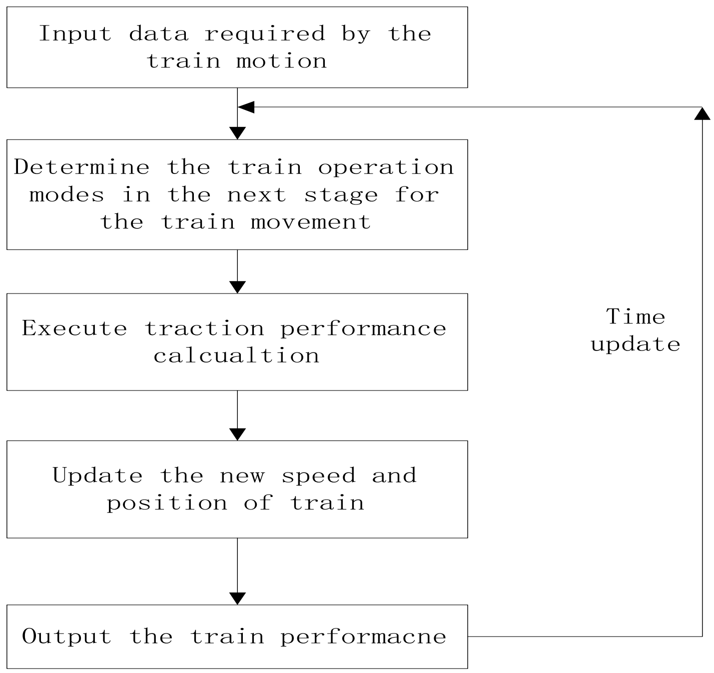


Fig. 4.11 Single train simulator

Once the train performance with a 'flat-out' speed against position profile is obtained from the train simulator, the coast control module starts to seek necessary coasting point(s). New coasting point(s) is produced if the train output performance does not satisfy the run-time requirement given by ***RTC***. The same process repeats until the new coasting point(s) satisfies the inter-station run-time requirements. Since the simulation time step is an important factor in the train movement calculation, the location(s) of the coasting point(s) is/are inevitably affected if the time step of the train movement is changed. In practice, the simulation time step of one second for train movement calculation is sufficient to cope with this operation. The structure of the module is shown in Fig. 4.12.

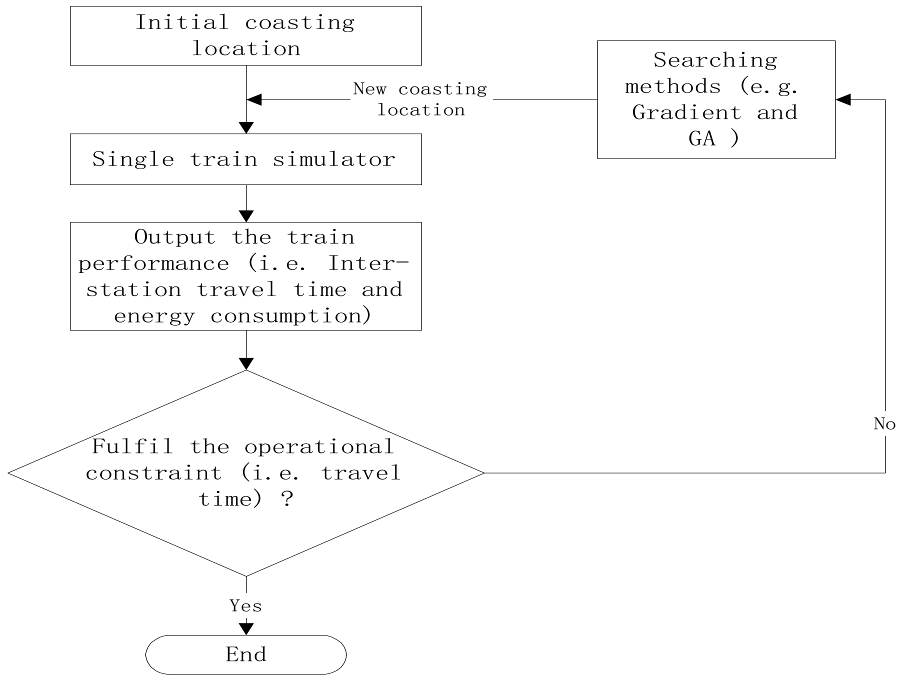


Fig. 4.12 Coast control module

4.4.2 Accuracy and data storage

With the single train simulator, accuracy of train performance modelling depends on the simulation time step between two activities. A smaller time interval implies better detail in the descriptions of the train movement. However, a longer computational time and much larger size of data storage are required.

4.5 Practical considerations

In practice, the 'coast control' system can be taken as an independent control tool for the regulation of an individual train's movement at train-level of the *HTRC*. The system can be integrated into the train-borne computers to perform the Automatic Train Operation (ATO) task in inter-station runs, directly interfacing with the *RTC*.

The *TBC* is not safety critical and its actions are monitored by the other systems, like the Automatic Train Protection (ATP). The allowed run-time between two stations is forwarded to the *TBC* when the train stops at start station of the inter-station run. The *TBC* then determines an appropriate speed profile for the next inter-station run in line with the track topology. Two sets of input are required for coast control: static and dynamic data. The former consists of track topology and traction equipment characteristics which are loaded onto the train-borne computer in advance; whilst the latter contains traffic conditions and operational requirements which are obtained from the *RTC*.

As repeated train movement calculation is needed to identify the coasting command during the course of the search, a fast microprocessor platform is necessary to speed up the search process for on-line implementation as there are only 30 seconds or less to find out the coasting solution once a train stops at a station. Selection of searching algorithms is of course a key concern from this viewpoint. The advantages and limitations of both classical and heuristic methods have been stated. The C++ language has the potential to serve as the tool to develop the train controller as it provides a relatively short computation time when compared with other high-level languages like Basic and Pascal, and with certain software development kits (e.g., Borland C++ or Microsoft visual C++), a ‘user-friendly’ interface can also be easily attained. Further, duplicated and hot standby hardware and software are expected to enhance the system availability and reliability.

Chapter 5
Performance Analysis of Train Based Controller

This chapter presents studies on the feasibility and flexibility of coast control to regulate the run-time of a train in an inter-station run in real-time railway applications. As stated in Chapter 4, such coast control allows for individual train control by asking each train to meet certain inter-station run-time requirements, while the run-times are derived from the *RTC* at the regional level of the *HTRC*. To evaluate the train movement with the chosen coasting points, a single train simulator has been developed.

Various classical approaches and a GA-based method are employed to locate the necessary coasting point(s) for inter-station runs and the quality of the solutions is compared, in terms of the trade-off between computational time, the optimality of the run time and energy use, for real-time control. As a quick solution is important in real-time train control, identifying the coasting point(s) to achieve a run-time as close to the required one as possible with minimum computational time is desirable. Further, the appropriate algorithms with respect to the track topology characteristics and inter-station distances are also investigated.

5.1 Simulation setup

A number of tests are carried out to study the performance of the search methods on coasting point identification for various track topologies and inter-station distances. The track topology and train information are obtained from the West Rail line in Hong Kong.

With the aid of the train simulator, details of the train movement throughout the inter-station run are shown. The simulator, equipped with a full set of user interfaces, is integrated with the search methods. The input interfaces, as illustrated in Fig. 5.1, 5.2 and 5.3, allow the data on track topology, train information, traction equipment characteristics and selection of inter-station run to be adapted for the train movement calculation. The input and output interfaces for coast control are incorporated and shown in Fig. 5.4, where the coasting point(s) is/are obtained to alter the train’s speed profile with given operational constraints, and this determines inter-station run-time and energy consumption.

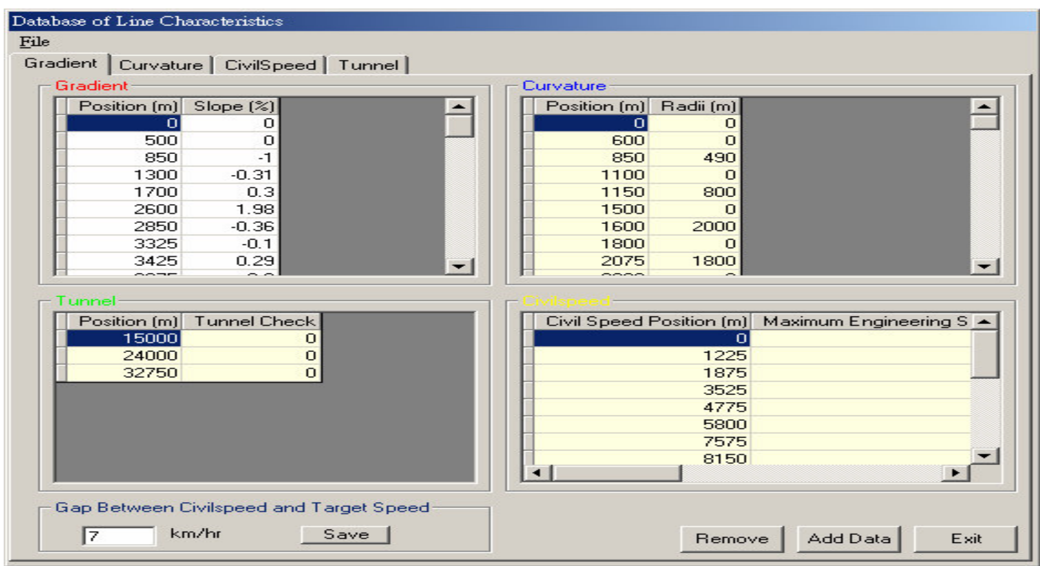


Fig. 5.1 Track topology

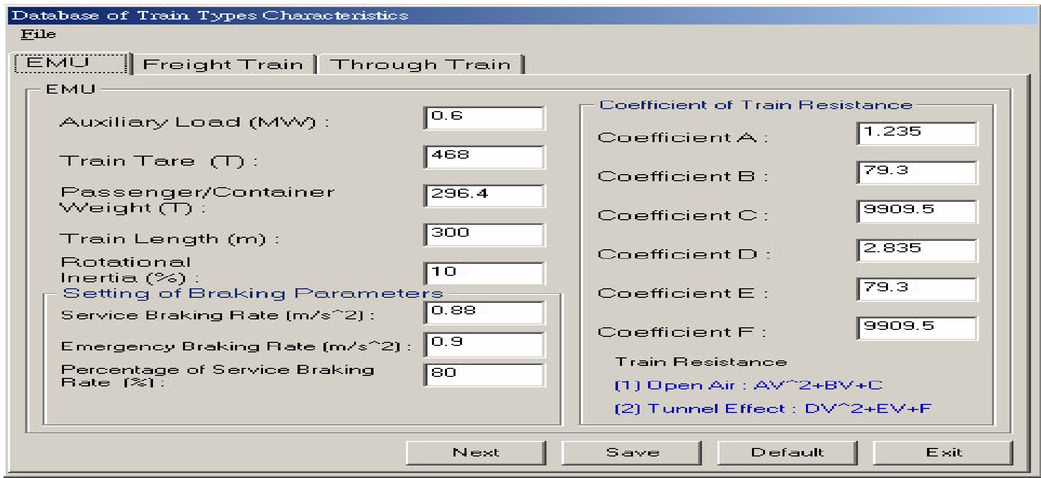


Fig. 5.2 Train information

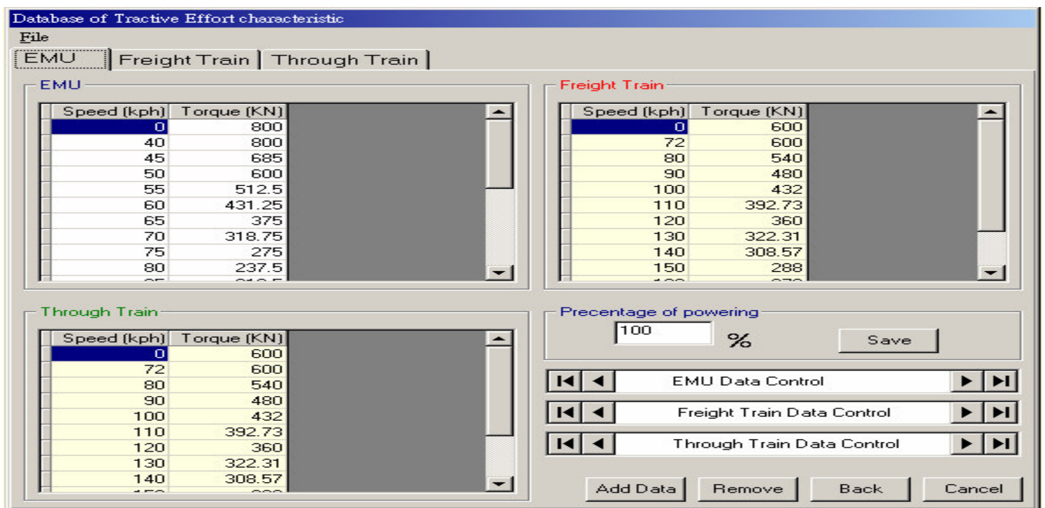


Fig. 5.3 Tractive equipment characteristics

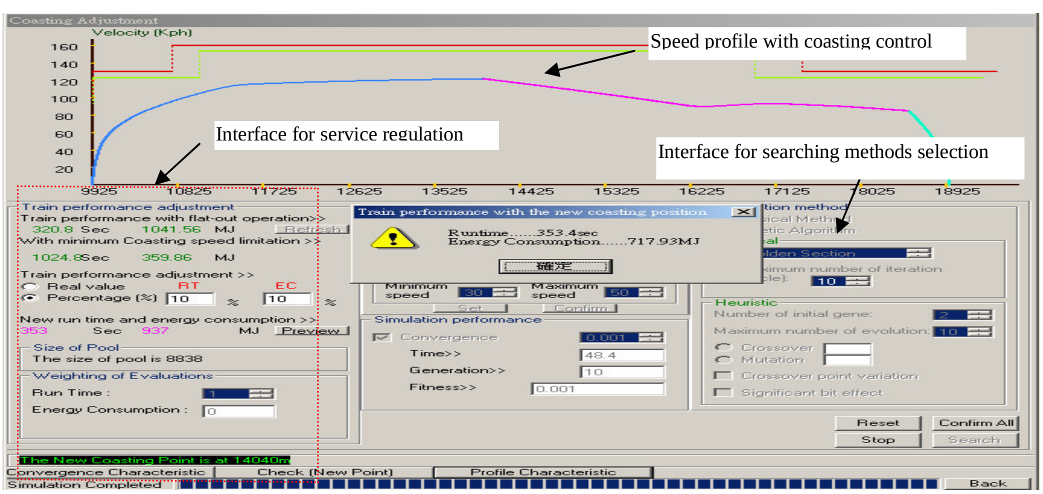


Fig. 5.4 Output interface with coast control performance

The maximum (i.e., operation with minimum coasting speed) and minimum (i.e., flat-out run) run-times for each inter-station run are obtained from simulation and they are taken as reference for the run-time regulation. This range is considered as one of the operational constraints in the regional level of control. The maximum line speed is different in each case study. The train is not allowed to operate in coasting mode until it exceeds the minimum coasting speed of 45km/h. A total of 6 case studies have been carried out to demonstrate the performance of various methods of searching for the coasting points.

As mentioned in Chapter 4, coast control can be divided into two categories – single and multiple coasting point control. Single-coasting-point searches are undertaken with the Golden search, Fibonacci, Gradient method and GA in the first case study. With GA, resolution on the coasting-point representation depends on the number of binary bits used, which is directly related to the distance between stations, where the resolution of one metre is predetermined. Further analysis on the performance of single-point search methods in accordance with the track topologies is given in the second case study. Two specific inter-station runs are chosen for investigation in the two studies – one of a distance of 1.1km, another 9km.

To illustrate better the application of multiple coasting point control, a longer inter-station run is preferable over a short inter-station run. Two multi-dimensional search methods, Nelder and Mead and GA, are applied. The advantages and limitations of the two techniques are highlighted in the third case study.

Three further case studies using GA-based coast control were carried out. The first shows the identification of single and multiple coasting point control with a simple GA method, while the second illustrates a hierarchical gene structure to incorporate the decision regarding the number of coasting point(s). Lastly, a quick mutation scheme, MARK, is used to reduce the computational time. The operational conditions and the related GA-based parameters are stated in the studies.

The simulations were performed on IBM-compatible PC with Pentium III CPU in all tests. A time-step of 0.2 sec is employed in the simulator as a compromise between reasonable computation time and adequate resolution of the solution. A cost target, obtained from the cost function defined in Eqn. 4.1, of 0.01 and a maximum number of iterations of 20 were set for 10 tests in the first three case studies. To further elaborate GA-based coast control, a cost target of 0 and a maximum number of iterations of 100 were adopted for the last three case studies.

5.2 Case studies

5.2.1 Single coasting point

This study revealed the performance of various search methods to achieve single coasting point control to fulfil a specific run-time requirement, which is given in terms of a run-time extension compared to the flat-out runs. To put the focus of the search on run-time regulation, the weighting factor W_T is set at 1 in the cost function. Short and long inter-station runs with different track topologies are chosen for investigation. These two inter-station operational conditions are given in Tables 5.1

and 5.2. In order to simply explore the basic function of coast control, the track gradient effect on the train movement calculation is neglected in this application.

Table 5.1 – Short inter-station operational conditions

Short inter-station run	
Inter-station distance	1.1 km
Max. line speed	80 km/h
Min. coasting speed	45 km/h
gradient effect on track	✕
Flat-out operation	81.8 sec and 181.37 MJ
Run-time extension	8% (i.e., 88 sec)
Weighting factor	W_T =1, W_E=0

Table 5.2 – Long inter-station operational conditions

Long inter-station run	
Inter-station distance	9 km
Max. line speed	130 km/h
Min. coasting speed	45km/h
gradient effect on track	✕
Flat-out operation	309 sec and 1004.6 MJ
Run-time extension	12% (i.e., 346sec)
Weighting factor	W_T =1, W_E=0

Results:

Table 5.3 – Average number of iterations with no track effect

Search methods	Average number of iterations	Run-time with the corresponding coasting solution (sec)
Short inter-station run		
Golden*	4	88.4
Fibonacci*	4	88.4
Gradient	2.8	87.8 ~ 88.2
GA	5.6	89
Long inter-station run		
Golden*	4	344.6
Fibonacci*	4	344.6
Gradient	3.4	345.6 ~ 347.4
GA	11.8	346 ~ 349

*Average is not applicable as the number of iterations is the same for the 10 tests
** The cost value of 0.01 is required in the tests

Discussions:

The run-time requirements for a short and long inter-station run, as well as the desired cost value of 0.01 are achieved. The results are summarised in Table 5.3. As each iteration requires a simulation run with the single train simulator and each simulation run takes a similar CPU time, the number of iterations is therefore a convenient time unit for performance indication. With Gradient search and GA-based method, the initial solutions are randomly chosen in each test and the number of iterations required to attain the cost of 0.01 is not the same in the 10 tests. Thus, the average number of iterations must be found for performance comparison. The initial solutions, however, are the same in each of the 10 tests with Golden search and Fibonacci and they are taken from the two ends of the solution space, Ψ excluding from X (see Section 4.2.2). The ratio of reduction of solution space in each iteration is evaluated prior to the simulation, and the number of iterations required is therefore the same in all tests.

From the simulation results, the classical and heuristic methods provide an acceptable solution with a reasonable average number of iterations. In general, the classical searching methods result in a smaller average number of iterations. The capability of GA to achieve a fitter solution is reduced when the search is approaching the desired solution. Since the coasting solution with GA is in binary representation of distance within the genes and the bits carry binary-weighted significance on the distance according to their locations in the gene, some of the useful genetic characteristics may be discarded when the mutation (reversion) of a bit is assigned to the most significant ones when the search is approaching the required solution. The randomly assigned initial population within GA also accounts for the higher total number of iterations.

The simulation results also reveal that the average number of iterations is lowest with the Gradient method. The step size and search direction of a new coasting solution with the Gradient method depend on the difference between the current and previous run-times with their corresponding coasting solutions, but not the solution space. In other words, the step size between the current and new coasting solution becomes larger when the difference between run-time of the current coasting location and the expected one increases. On the other hand, the solution space with the Golden section and Fibonacci search is reduced with a specific ratio. The number of iterations in all tests with these two methods is always the same.

5.2.2 Track topologies

This study is to examine the relationship between coast control and track topology. The inter-station operational conditions are the same as in the previous study and the corresponding track topology characteristics are shown in Tables 5.4 and 5.5. The related flat-out operation and desired run-time requirement are listed in Tables 5.6 and 5.7. Again, the four search methods with single coasting point search are applied.

Table 5.4 – Track topology of the short inter-station run

Section (m)	Slope (%)
0 ~ 150	0.3
150 ~ 600	1.08
600 ~ 1100	0.3

Table 5.5 – Track topology of the long inter-station run

Section (m)	Slope (%)
0 ~ 125	0
125 ~ 1325	0.76
1325 ~ 6225	0.8
6225 ~ 6975	0.76
6975 ~ 7775	0
7775 ~ 8475	0.2
8475 ~ 8675	0.11
8675 ~ 9025	0

Table 5.6 – Short and long inter-station condition with positive slope effect

Short inter-station run	
Slope effect	Table 5.4 with positive value
Flat-out operation	82.4 sec and 233.75 MJ
Run-time extension	7% (i.e., 88 sec)
Long inter-station run	
Slope effect	Table 5.5 with positive value
Flat-out operation	334.6 sec and 1198.23 MJ
Run-time extension	3.5% (i.e., 346 sec)

Table 5.7 – Short and long inter-station condition with negative slope effect

Short inter-station run	
Slope effect	Table 5.4 with negative value
Flat-out operation	80 sec and 169.9 MJ
Run-time extension	10% (i.e., 88 sec)
Long inter-station run	
Slope effect	Table 5.5 with negative value
Flat-out operation	292.6 sec and 668.1 MJ
Run-time extension	18.2% (i.e., 346 sec)

Results:

Table 5.8 – Average number of iterations with positive track slope

Search methods	Average number of iterations	Run-time with the corresponding coasting solution (sec)
Short inter-station run with positive slope (i.e., uphill)		
Golden*	6	88.4
Fibonacci*	6	88.6
Gradient	3.2	87.8 ~ 88.4
GA	10.3	86 ~ 89
Long inter-station run with positive slope (i.e., uphill)		
Golden*	6	345.8
Fibonacci*	6	345.6
Gradient	4.2	346.2 ~ 347.4
GA	6.9	347

*Average is not applicable as the number of iterations is the same of 10 tests.

Table 5.9 – Average number of iterations with negative track slope

Search methods	Average number of iterations	Run-time with the corresponding coasting solution (sec)
Short inter-station run with negative slope (i.e., downhill)		
Golden*	6	87.8
Fibonacci*	6	88
Gradient	3.5	87.8 ~ 88.2
GA**	5	87 ~ 88
Long inter-station run with negative slope (i.e., downhill)		
Golden*	9	346.2
Fibonacci*	10	346.8
Gradient	6.3	344.8 ~ 347
GA**	6.3	346 ~ 350

*Average is not applicable as the number of iterations is the same of 10 tests.
**The required cost of 0.01 cannot be achieved at the maximum number of iterations.

Discussions:

The simulation results with extreme track geometries are shown in Tables 5.8 and 5.9. Again, both the classical and heuristic methods provide the desired solution within a reasonable average number of iterations. The average number of iterations with classical method is also smaller than that with the heuristic method. From Tables 5.3, 5.8 and 5.9, it is obvious that the average number of iterations with the Golden section, Fibonacci and Gradient method increase when the track topology is taken into account because the train movement is affected by the track topology and the run-time is not necessarily monotonically related to the corresponding coasting location. With GA, the required cost may not be achieved even with the maximum number of iterations.

Further, it is worth noting that the average number of iterations with the track topology with positive slopes to achieve the same cost is smaller than that with the negative one. The gradient force with the track topology of positive slope and the other train resistances are all opposing to the train movement and hence slows down the train speed. Run-time thus decreases monotonically if the coasting point is shifted from the starting station to the next and hence the search is more significantly uni-directional. Nevertheless, the gradient force with the track topology of negative slopes is against the other train resistances and it thus leads to lower energy consumption with the same operational constraints in an inter-station run. In addition, a clear-cut difference between the current and a new coasting solution cannot be easily obtained and thus more iterations are needed.

5.2.3 Multiple coasting points

This study was undertaken to explore the performance of multi coasting point control with different track topologies and inter-station distances. The inter-station operational conditions and track topologies remain the same as illustrated in the previous two studies. Two multi-dimensional search methods, Nelder and Mead method and simple GA, have been applied. The advantages and limitations of multiple coasting point control are investigated in the comparison with single coasting point control.

Results:

Table 5.10 – Average number of iterations with multiple coasting point control in the short inter-station run

Search methods	Average number of iterations	Run-time with the corresponding coasting solution (sec)
Short inter-station run with no slope		
Nelder & Mead	13.67	86.2 ~ 87.6
GA	6.3	88 ~ 90
Short inter-station run with positive slope (i.e., uphill)		
Nelder & Mead	7.17	87.2 ~ 88
GA	5.56	88 ~ 89
Short inter-station run with negative slope (i.e., downhill)		
Nelder & Mead	9.8	86.6 ~ 87.2
GA	6	87 ~ 88

Table 5.11 – Average number of iterations with multiple coasting point control in the long inter-station run

Search methods	Average number of iterations	Run-time with the corresponding coasting solution (sec)
Long inter-station run with no slope		
Nelder & Mead	9.6	344.2 ~ 348.4
GA	8.2	346 ~ 348
Long inter-station run with positive slope (i.e., uphill)		
Nelder & Mead	12.75	344.6 ~ 346.8
GA	12.1	347 ~ 348
Long inter-station run with negative slope (i.e., downhill)		
Nelder & Mead	8.5	343 ~ 348.4
GA	9.8	346 ~ 350

Remark: The required cost of 0.01 may not be attained at the maximum number of iterations in both methods.

Table 5.12 – Comparison of the train movement performance with single and multiple coasting point control in the short inter-station run

Inter-station distance	9 km	
Max. line speed	160 km/h	
Run-time extension	35% run-time of the flat-out run	
Energy consumption of train movement with the corresponding coasting solution (MJ)	Single point	Multi point
	1089.9 (see Fig. 5.5)	1306.7 (see Fig. 5.6)
Search method	Golden	Nelder and Mead

Table 5.13 – Comparison of the train movement performance with single and multiple coasting point control in a very long inter-station run

Inter-station distance	30 km	
Max. line speed	160 km/h	
Run-time extension	30% run-time of the flat-out run	
Energy consumption of train movement with the corresponding coasting solution (MJ)	Single point	Multi point
	2675.8 (see Fig. 5.7)	2535 (see Fig. 5.8)
Search method	Golden search	Nelder and Mead

Discussions:

As shown by the results summarised in Tables 5.10 and 5.11, GA provides a lower average number of iterations and a fitter solution than the Nelder and Mead method in general. The search performance with the Nelder and Mead method is limited by the three operation factors: reflection, expansion and contraction, when resizing and reshaping the simplex during the iterative process. With a large size of operation factors, the new solution at the next iteration, which evolves from the previous solution, will be further apart from its preceding vicinity. A new solution is therefore more likely to be trapped out from the desired solution with Nelder and Mead methods when the search is approaching the optimum, if the three operation factors are set to larger values. Nevertheless, the capability of finding the optimum solution is improved in GA if the resolution of the solution gets smaller.

Even though these two methods obtain the solution within the maximum number of iterations, the energy consumption of the corresponding train movement is roughly 20% higher than that with the single coasting point for the same run-time extension in a long inter-station run. The results from the Golden and Nelder and Mead methods are taken out for comparison between the single and multiple coasting point control. The results are summarised in Table 5.12 and Figs. 5.5 and 5.6. A train spends more time at high speed with the single coasting point, whilst it has to accelerate more at low speed with multiple coasting points and power consumption is thus higher. Lower energy consumption can be accomplished with single coasting point control.

In addition, the second or further coasting points are only necessary when a train operates in motoring again to recover the momentum of the train movement from a low speed level because of the track topology. Therefore, in a short inter-station run, the algorithms usually produce a single coasting solution even though they are designed to search for multiple coasting points (i.e., the second coasting point is located within the first coasting) since there is not enough space to accommodate more coasting points.

Despite that the application of multiple coasting point control is not the most desirable in term of energy consumption, it is still one of the possible measures to provide a broader solution space. To further explore the relationship between the inter-station distance and number of coasting points, a test on a very long inter-station run was carried out to identify the necessary coasting solution with the Golden and Nelder and Mead methods to achieve the operational conditions listed in Table 5.13. The results are illustrated in Figs. 5.7 and 5.8. The simulation results show that the energy consumption with multiple coasting point control is 5 ~ 6% less than that of single coasting point control. Hence, adoption of multiple coasting point control is preferable in a very long inter-station distance run. Although there are no specific rules to identify the number of coasting points in an inter-station run, it has been shown that the inter-station distance and track topology are the two key factors in the application of coast control for train service regulation.

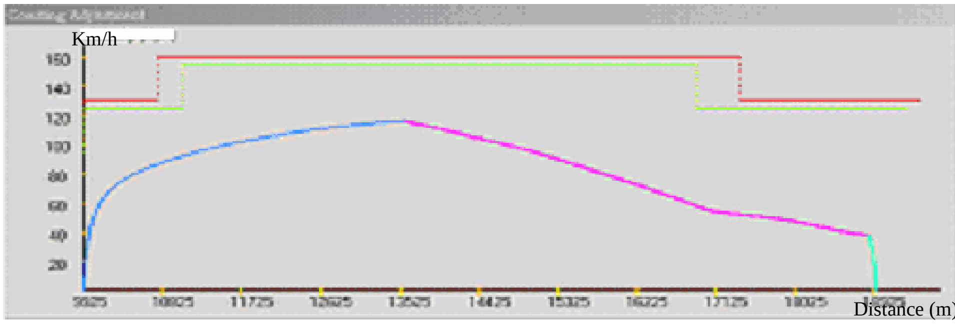


Fig. 5.5 Train speed profile with single coast control

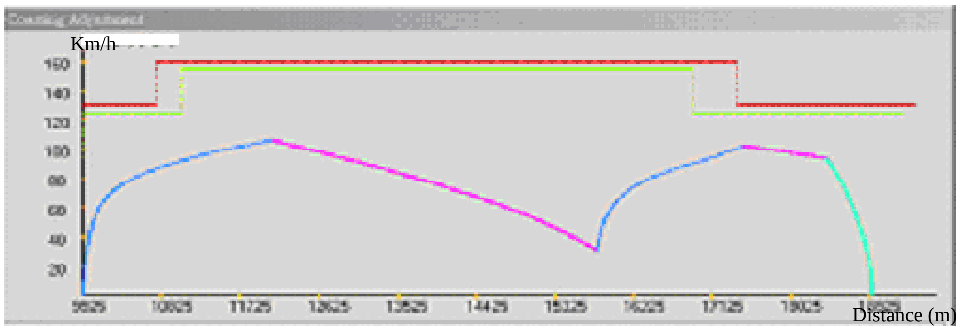


Fig. 5.6 Train speed profile with multiple coast control

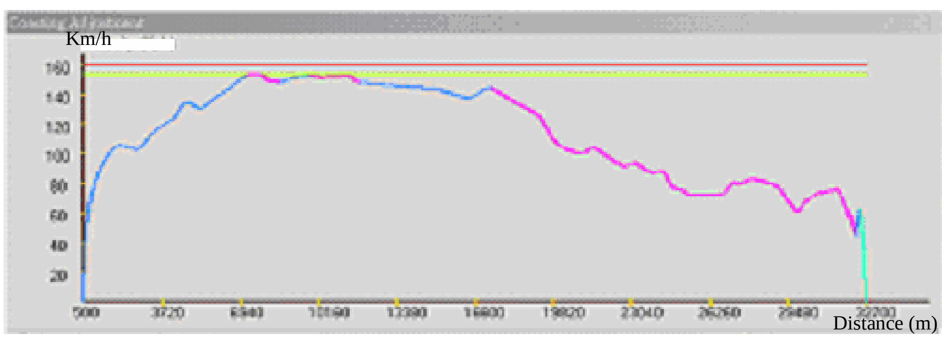


Fig. 5.7 Train speed profile with single coasting point in a very long inter-station run

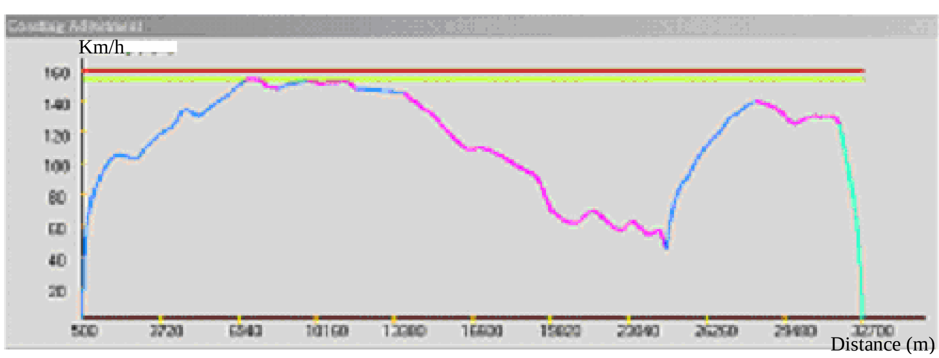


Fig. 5.8 Train speed profile with multiple coasting points in a very long inter-station run

5.2.4 Simple GA-based coast control

This experiment was used to investigate the application of single and multiple coasting point control of train operation with a simple GA. The number of coasting point(s) required was predetermined by the operators. The inter-station run-time is extended by 10% compared to the flat-out run. Two cases are considered here with different track topologies. The operational conditions and the GA-based parameters for the simulation are listed in Table 5.14.

Table 5.14 – Operational conditions and GA parameters

Operational conditions		
	Case I	Case II
Inter-station distance	9km	
Run-time extension	10% more than that in flat-out run	
Weighting factor	W_T=1, W_E=0	
Track	Downhill slopes	Uphill slopes
GA parameters*		
	One-fixed coasting point	Two-fixed coasting points
Population size	12	12
Crossover rate	1	1
Mutation	No	10%
MARK	40%	No

*The GA parameters given here are proven to provide the best solutions with extensive study.

Results:

Table 5.15 – Inter-station runs with uphill and downhill slopes

	Single coasting point		Multiple coasting points	
Case	I	II	I	II
Cost	0.0015	0.0186	0.0079	0.0008

* The computation time is below 10 seconds in all tests.

Discussions:

With the aid of the train simulator, the situations were modelled and the results are summarised in Table 5.15. It can be seen that a simple GA-based coast control provides near optimal solutions in both cases at the end of the maximum number of iterations. In addition, it is possible that the GA-based control is able to find the optimal solution, if a larger number of generations is allowed. A single coasting point control is more applicable for an inter-station run with downhill slopes as a train tends to maintain its speed during coasting and hence it favours one coasting point. Nevertheless, a train loses speed quickly during coasting with uphill slopes and it usually needs re-motoring and then another coasting point is required. With simple GA based coast control, the number of coasting point is assumed to be fixed and a flexible coast control cannot be attained. In the next study, HGA is introduced in order to obtain the number of coasting points for the regulation of train schedule in accordance with track topologies and inter-station distances.

5.2.5 Hierarchical genetic algorithm

This experiment was used to investigate the application of GA with a hierarchical gene structure to provide the appropriate number of coasting point(s) in a search. A 3.2km long inter-station run is chosen here while the other operational requirements and GA-based parameters for simulation are given in Table 5.16. The stated GA parameters are shown to provide the best solutions, base on the experience of 100 tests that have been carried out with different initial generations.

Table 5.16 – Operational conditions and GA parameters

Operational conditions			
	Case I	Case II	
Inter-station distance	3.2km		
Run-time extension	30% more than that in flat-out run		
Energy consumption	30% less than that in flat-out run		
Weighting factor	W_T=0.5, W_E=0.5		
Track	Downhill slopes	Uphill slopes	
GA parameters*			
	One-fixed coasting point	Two-fixed coasting point	HGA
Population size	12	12	12
Crossover point	1	1	1
Mutation	No	10%	10%
MARK	40%	No	40%

*The GA parameters given here are proven to provide the best solutions with extensive study.

Results:

Table 5.17 – Percentages of coasting point selection

	1-coasting point	2-coasting point
Case I	70%	30%
Case II	28%	72%

* The computation time is within 50 seconds in all tests.

Discussions:

Different inter-station runs with uphill and downhill tracks have been examined to obtain the number and location(s) of coasting point(s) with HGA. Figs. 5.9 and 5.10 show that a near-optimal solution is attained with HGA in both cases and an even lower cost value can be achieved when a higher maximum number of generations is allowed. Simulation results also reveal that the cost of the solution attained from the HGA is lower, when compared to the two fixed coasting-points in Case I and the one fixed coasting-point in Case II respectively. However, the situation is reversed when compared with one-fixed coasting point control in Case I and two-fixed coasting point control in Case II.

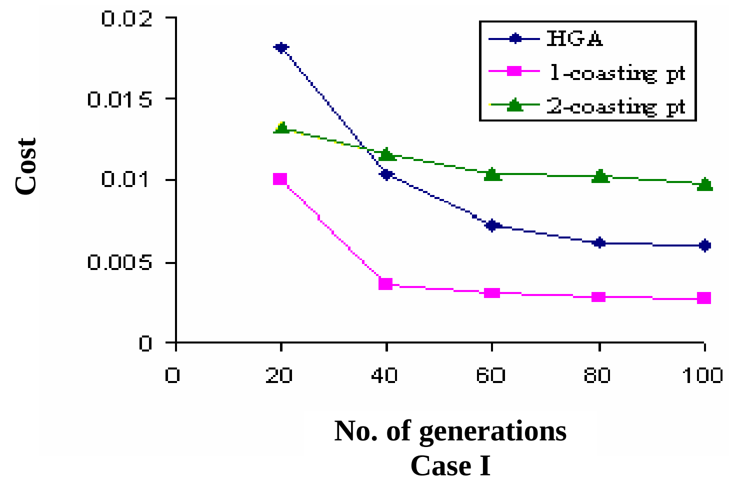


Fig. 5.9 Average cost of an inter-station run with downhill slopes

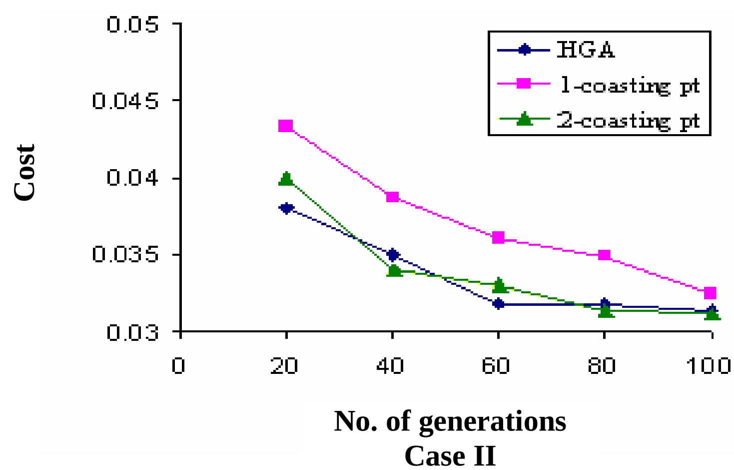


Fig. 5.10 Average cost of an inter-station run with uphill slopes

The study illustrates that one coasting point is preferred in Case I because of the downhill track, whilst the track in Case II mainly consists of uphill slopes and hence re-motoring and further coasting points are necessary. Table 5.17 summarises the percentages of coasting point selection (i.e. 1 or 2 points) in 100 tests, with HGA in both cases. HGA selects the correct number of coasting points in more than 70% of the cases, in which the necessary number of coasting point(s) provides a lower cost value. Even though HGA may not provide solutions as good as those obtained from GA with a fixed number of coasting points in certain traffic conditions, the difference, however, is very small. Moreover, HGA provides an all-purpose approach to determine the appropriate number of coasting points dynamically, according to a wide range of operational conditions. A high flexibility of train control can be achieved.

5.2.6 Minimum-Allele-Reserve-Keeper

This study investigates the performance of the MARK operator in GA. The inter-station conditions and operational requirements remain as in the last case study. The track topology characteristics are shown in Table 5.18. The evolution methods adopted in GA and their parameters are depicted in Table 5.19.

Table 5.18 – Track topology

Section (m)	Slopes (%)
0 ~ 850	0
850 ~1300	-1
1300 ~ 1700	-0.31
1700 ~ 2600	0.3
2600 ~ 2850	1.98
2850 ~ 3200	-0.36

Table 5.19 – Evolution method in GA

Genetic operators	Parameter
Crossover points	2
mutation	0 ~ 40%
MARK	0 ~ 40%

Discussions:

Fig. 5.11 illustrates the cost distribution of the resulting chromosomes with different rates for mutation and MARK. The darker area implies a lower cost value. The simulation result shows that the cost is gradually improved when the percentage of mutation adopted in evolution increases, even when no MARK is introduced.

Similarly, the cost improvement can also be obtained when a higher percentage of MARK rate is given, if mutation is not applied in the test. With MARK, the cost value is even lower than that from mutation.

Tests have also been undertaken to investigate the improvement in chromosome cost when both MARK and mutation are allowed in evolution. Fig. 5.11 reveals that the introduction of MARK produces significantly fitter solution with the same number of generations when mutation manages a gradual improvement on cost. The 20%~40% MARK rate and 0~10% mutation lead to a lower cost in this application. With MARK, the chance for offspring to escape from the optimal solution is reduced since the number of bits to be mutated is limited. Hence, a lower cost can then be achieved with MARK.

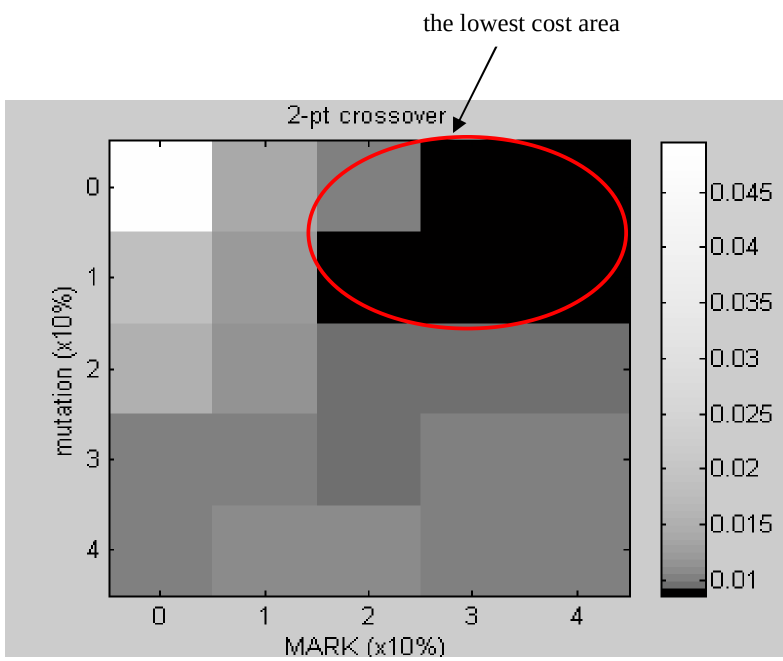


Fig. 5.11 Cost values with different mutation and MARK rates

5.3 Concluding remarks

Classical and heuristic approaches to identify the necessary coasting point(s) for inter-station run-time regulation in metro operations have been presented. The results reveal that the proposed search methods are applicable in real-time coasting control. Inter-station distance and characteristics of the track topology are the two important factors in choosing search methods for coasting control.

It is also shown that the average number of iterations is smaller with the Gradient based method in single coasting point control. Its main drawback is that the search process terminates in the Gradient based method if the slope of the cost function vanishes in the search. Golden and Fibonacci methods, however, are more robust than the Gradient based methods as they only depend on the size of the solution space. An additional advantage of the Fibonacci method is the reduction scale on the solution space in each iteration is maximised with the given number of iterations in a search.

It can also be deduced that the heuristic approach, GA, offers a lower average number of iterations and a fitter solution with multiple coasting point control when compared with Nelder and Mead method. Energy demand in a very long inter-station run is slightly lower with multiple coasting points. However, there is usually not enough room to accommodate multiple coasting points if the distance between stations is short. Single coasting point control is therefore preferable in metro operations.

With GA-based control, a flexible train control cannot be attained with a fixed number of coasting point(s) even though it may be able to perform well in certain operational conditions, but not all. The application of HGA has therefore been adopted to obtain the number and locations of coasting points according to traffic conditions. Although the HGA does not guarantee a fitter solution in certain traffic conditions, when compared to GA with fixed number of coasting points, it provides a generic approach to determine the appropriate number of coasting points according to the characteristics of track geometry and inter-station distances. A high flexibility of train movement control can be achieved.

Further, MARK has been introduced to speed up the search and it has been successfully incorporated in the HGA to provide solutions of lower cost and meet the demand of this real-time application. Intelligent generation evolution, like the basic concepts of genetic engineering, by the combination of mutation and crossover, gene inversion, gene duplication, gene deletion and other techniques, is a potential direction to further speed up the search with GA.

The railway system consists of numerous sub-systems which require stringent real-time monitoring and control because of the demanding safety standards. The results show that the search methods are capable of providing reasonably good and fast coasting solution(s) for flexible online train scheduling control with the aid of a train simulator according to the operational requirements in all case studies. In practice, dynamic coasting control has not yet been commonly applied in run-time regulation. It can be integrated in the on-board Automatic Train Operation (ATO) system and the coasting point(s) for the next inter-station run can be obtained when a train stops at a station. Even though it may not be the solution for all systems, it certainly offers an alternative to the operators. In addition, dynamic coasting control is more flexible in the regulation of train schedules as it adapts to the current train service demand and the additional advantage is that energy savings can be achieved. From the application viewpoint, the search of coasting points within multiple inter-station runs for a specific overall run-time is a challenging proposition. The search problem becomes multi-dimensional and the solution space is large.

Chapter 6
Regional Train Controller

Even though an effective train control for inter-station runs is obtained with the train-based controller, the balance between run-time and energy consumption is merely focusing on individual trains, but not taking multi-train operation into account. The advantage of train coordination is thus somewhat confined. To meet the variations in demand for train service, dwell-time and run-time coordination between trains are the two commonly used control approaches in real life. The current practice on dwell-time and run-time control relies on pre-determined settings. The solution is therefore only optimal with respect to a nominal operational condition, but not necessarily the current train service demand. In order to achieve flexible and efficient train control responding to time-varying traffic conditions, a more sophisticated train service management system at the regional level of control is required, to adjust the dwell-time and run-time for multi-train operation.

6.1 Introduction

6.1.1 Methodology

Since a railway line is sectioned into regions in *HTRC*, the formulation of the trains' instructions is divided into a number of manageable tasks for implementation. The computational effort to control the train operations is reduced, thanks to the introduction of a number of regional train controllers (*RTC*), compared to centralised control. As a train usually takes 10 minutes or more to run across a region, which may consist of three or more stations in metro systems, a relatively longer computational time is allowed for the *RTC* to evaluate train operation in each control cycle, compared with that in the *TBC*. With the lower computational demand, the dynamic programming (DP) [80] technique has the potential to perform appropriate dwell-time and run-time control at this level.

DP divides the multi-stage decision process into a series of single-stage problems, and the optimal solution is guaranteed under the given traffic conditions and operational constraints. As computation time is extremely critical for real-time applications and since it escalates near exponentially with increasing complexity of control actions, state grouping is introduced in DP to reduce the computational demand while not jeopardising the solution optimality. At the regional level of control, an event-based traffic flow model is employed to represent the multi-train operation, with dwell-time and run-time control, as calculation of every detail of the trains' movement is not required. A number of studies are then conducted to explore the *RTC*'s performance.

6.1.2 Literature reviews

Over the past few decades, numerous studies on multi-train control have been carried out to improve the performance of train operation either by classical searches or heuristic methods. Ho *et al*. [81] developed an event-based traffic flow model to represent the traffic control at railway junctions. The traffic controller assigns the optimised train distribution at junctions by DP to impose minimum total weighted delay on trains. In the model, conflict resolution is treated as a multistage process in which each stage (i.e., event) allows one train to pass through the junction and is characterised by the number of trains left in the conflict area. Each stage has a set of possible states which denote the distribution of trains with different speeds and positions in the conflict area. This study is a good example to demonstrate the establishment of the event-based model in the application of traffic control with DP.

Chang *et al*. [82] proposed a fuzzy logic train controller to optimise the dwell-time schedule at successive stations, according to the traffic conditions and operation demands in DC railway system. An event-based model provides the evaluation tool to conduct the dwell-time control in this application. The tested line is over 6.6 km with 6 passenger stations. An event is defined as a train departing at a station with dwell-time adjustment. The provision of dwell-time at stations is a compromise between the service regularity and energy consumption of train operation with regenerative braking. However, the impact on the train operation with run-time regulation in inter-station runs was not considered in this study.

The two examples show that event-based models are commonly adopted in railway simulation and they provide pictorial views of traffic flow. Given the traffic flow model, regulation of train operation is then attained either by classical or heuristic methods. To achieve highly flexible train coordination, multi-train control through dwell-time, run-time and a combination of the two is proposed in the regional level of control. Train coordination is connected through a series of event, and DP is adopted to seek for the optimal solution.

6.2 Problem formulation

6.2.1 State definition for traffic flow model

To investigate the traffic flow of trains at regional level of *HTRC*, an event-based model is appropriate as details of train operation are not required. With the application of DP, a state-space traffic flow model consists of a number of possible states and each stage allows one inter-station run for trains. To establish an event-based traffic flow model, the links between events are the transformation between states in the state-space model with DP. The characteristics of train movement and the definitions of the states in the model are described in the following sections.

6.2.1.1 Nominal schedule

In a simple inter-station run between two stations, the train behaviour can be described as follows:

$$AT_x{}^i - AT_{x-1}{}^i = NDW_{x-1}{}^i + NRT_{x-1,x}{}^i \tag{6.1}$$

where $AT_x{}^i$ and $AT_{x-1}{}^i$ are the arrival times of the train i at two successive stations x and x-1 respectively; while $NDW_{x-1}{}^i$ and $NRT_{x-1,x}{}^i$ are the corresponding nominal dwell-time of the train at station x-1 and nominal run-time for the train between the two stations. Inter-station run-time depends upon the exact train movement between stations.

In reality, there are a number of trains running on the same line and they run in a sequential order, as shown in Fig. 6.1, with their separations governed by the headway. Each train has its own arrival time at stations. The state of this traffic flow model at a particular stage k can be defined as the set of arrival times of trains at successive stations. For instance,

$$State:[AT]_k = \lfloor AT_x{}^{i+k},\ AT_{x-1}{}^{i+k+1},\ AT_{x-2}{}^{i+k+2},\ \ldots\ ,\ AT_{x-n}{}^{i+k+n} \rfloor_k \qquad (6.2)$$

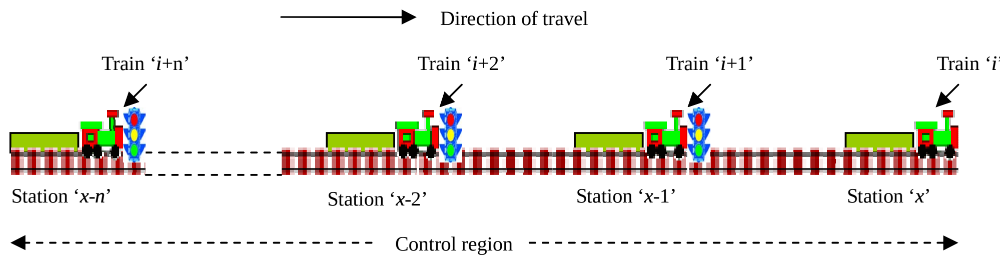


Fig. 6.1 Traffic conditions

The dynamic behaviour of multi-train operation with respect to the nominal schedule between two states is therefore summarised in the following equation:

$$[AT]_k - [AT]_{k-1} = [NDW]_{k-1,k} + [NRT]_{k-1,k} \qquad (6.3)$$

where $[AT]_k$ and $[AT]_{k-1}$ are the set of arrival times of trains at successive stations at stage k and k-1 respectively. The train movement activities between stages (i.e., stage-to-stage transformation) are described in Eqn. 6.3. $[NDW]_{k-1,k}$ and $[NRT]_{k-1,k}$ are the corresponding nominal dwell-times for trains at stations and nominal inter-station run-times of trains in successive stations between stages k-1 and k respectively.

$$[NDW]_{k-1,k} = \lfloor NDW_x{}^{i+k-1},\ NDW_{x-1}{}^{i+k},\ NDW_{x-2}{}^{i+k+1},\ \ldots,\ NDW_{x-n}{}^{i+k+n-1} \rfloor_{k-1,k} \qquad (6.4)$$

$$[NRT]_{k-1,k} = \lfloor NRT_{x-1,x}{}^{i+k},\ NRT_{x-2,x-1}{}^{i+k+1},\ \ldots,\ NRT_{x-n,x-n+1}{}^{i+k+n-1} \rfloor_{k-1,k} \qquad (6.5)$$

In order to maintain the train service required by the demand, regulation of dwell-time at stations and run-time in successive inter-station runs with respect to the nominal schedule are the two viable control actions.

6.2.1.2 Dwell-time and run-time control

To maintain the train service with dwell-time control, the dwell-time of trains at stations can be either extended or reduced with respect to the nominal schedule. Eqn. 6.3 then becomes,

$$[AT]_k - [AT]_{k-1} = [NDW]_{k-1,k} + [RDW]_{k-1,k} + [NRT]_{k-1,k} \tag{6.6}$$

where $[RDW]_{k-1,k}$ is the set of dwell-time extensions or reductions of trains at successive stations, with respect to the nominal dwell-time schedule.

$$[RDW]_{k-1,k} = \left[RDW_x{}^{i+k-1},\ RDW_{x-1}{}^{i+k},\ RDW_{x-2}{}^{i+k+1},\quad \ldots,\quad RDW_{x-n}{}^{i+k+n-1}\right]_{k-1,k} \tag{6.7}$$

Since the nominal train schedule (i.e., $[NDW]_{k-1,k}$ and $[NRT]_{k-1,k}$) is constant, $[RDW]_{k-1,k}$ is the possible control variable to maintain the train schedule. Similarly, when run-time control is adopted, the dynamic behaviour of the trains' operation can be defined as:

$$[AT]_k - [AT]_{k-1} = [NDW]_{k-1,k} + [NRT]_{k-1,k} + [RRT]_{k-1,k} \tag{6.8}$$

where $[RRT]_{k-1,k}$ is the set of possible run-time extensions or reductions for trains in successive inter-station runs, with respect to the nominal run-time schedule.

$$[RRT]_{k-1,k} = \left[RRT_{x-1,x}{}^{i+k},\ RRT_{x-2,x-1}{}^{i+k+1},\quad \ldots,\quad RRT_{x-n,x-n+1}{}^{i+k+n-1}\right]_{k-1,k} \tag{6.9}$$

To further enhance the flexibility of train control, both $[RDW]_{k-1,k}$ and $[RRT]_{k-1,k}$ can be introduced into Eqn. 6.3 and such mixed control is described by:

$$[AT]_k - [AT]_{k-1} = [NDW]_{k-1,k} + [RDW]_{k-1,k} + [NRT]_{k-1,k} + [RRT]_{k-1,k} \tag{6.10}$$

With the above definitions, the state space traffic flow model is now established, while the dynamic behaviour of multi-train movement along the line, as well as dwell-time or run-time control, are represented properly.

6.2.2 Traffic model and solution space

With reference to Fig. 6.2, a train is assumed to be at station '0' and it will reach station '4' through three intermediate stops (i.e., stations '1', '2' and '3'). The numbered states represent the corresponding arrival-times of the trains at successive stations; while the line connecting the states indicates the sum of the dwell-time at the current station and the possible inter-station run-times, with the corresponding energy consumption of the trains to the next station. For instance, the run-time and energy consumption between station '0' and '1' are represented by E_{01} and T_{01} respectively in the diagram.

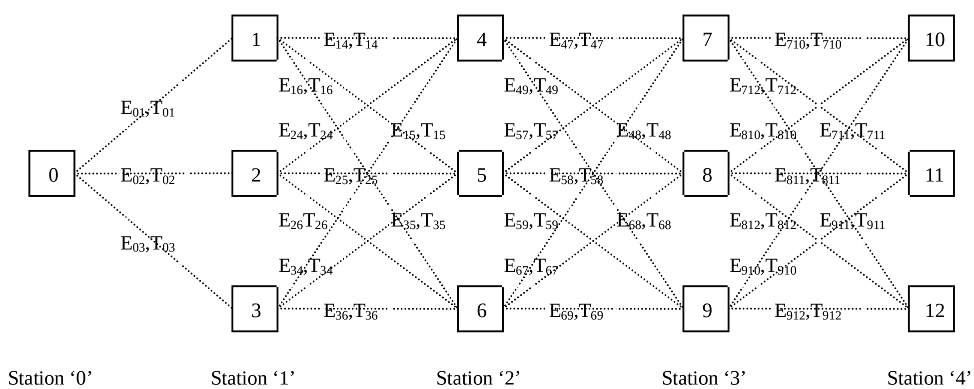


Fig. 6.2 Possible inter-station run-time combinations to meet the total run-time and energy consumption of a single train in the whole journey

To meet the necessary operation demands, it is worth noting that the total travelling time (i.e., dwell-times and run-times through three intermediate stops) and energy consumption of the whole journey from station '0' to '4' may vary, depending on how the operator chooses the corresponding dwell-times at stations and run-times in successive inter-station runs. Dwell-time does not affect energy consumption of individual trains, but only the total travelling time, whereas run-time affects both. There are a large number of dwell-time and run-time combinations for the train to reach station '4' through three intermediate stops and each will produce different total run-time and energy consumption. In other words, given the required total run-time or energy consumption for the whole journey, as set by the service schedule, determining the appropriate dwell-times at stations and run-times in successive inter-station runs is the essence of the problem. In addition, a high number of stations implies escalating sets of dwell-time and run-time combinations to achieve the specific operational requirement.

With multi-train operation and control, train movement is interactive and inter-dependent. Each train carries its own corresponding dwell-time and run-time schedule which is different from that of the others. The size of the solution space in the state diagram inevitably inflates and the problem is extensive and multi-dimensional. The objective of *RTC* is to find the optimum solution with respect to a particular criterion, such as a specified run-time or minimum energy consumption.

6.3 Dynamic programming

Dynamic programming (DP) has been established since the 1950s [83,84]. It converts a multistage decision process into a sequence of single-stage problems. DP examines only a small subset of possible decision sequences and, hence, computational demand can be vastly reduced. The subset is guaranteed, under the right conditions, to contain the optimal solution.

The following characteristics are common to most of the applications of dynamic programming:

1. The problem can be divided into stages with a decision required at each stage.
2. Each stage has a finite number of associated states.
3. The decision chosen at any stage describes how the state at the current stage is transformed into the state at the next stage.
4. Given the current state, the optimal decision for each of the remaining stages must not depend on previously reached or chosen decisions.

6.3.1 State-space traffic flow model for optimisation

Train scheduling is a multistage decision process in which each stage allows one inter-station run for trains and is characterised by the arrival time schedule of the trains at successive stations. Dwell-time adjustment at stations and run-time regulation in inter-station runs with respect to the nominal schedule are the two possible control mechanisms to meet the operational demands.

To optimise train operation with dynamic programming, the following assumptions have been made,

1. No bypass operation through sidings.
2. The speed and distance profiles for the nominal run-time schedule are the same for all trains. It hence implies no mixed traffic.

6.3.2 State formulation

6.3.2.1 Initialisation

The state is defined as the arrival times of trains at successive stations. For example, with the given traffic condition as shown in Table 6.1, in which there are four stations, a state at the initial stage '0' can be deduced as:

Assume train 1 departs at station '0' when time = 0 sec;

Arrival time of train 1 at station '3', AT_3^1 = (135+25+121+25+90) sec

= 396 sec

With 120 sec headway, the arrival time of train 2 at station '2', AT_2^2

= (120+135+25+121) = 401 sec

Arrival time of train 3 at station '1', AT_1^3 = (120×2 +135) = 375 sec

Arrival time of train 4 at station '0', AT_0^4 = (120×3-25) = 335 sec

An initial state $[396, 401, 375, 335]_0$ is obtained.

Table 6.1 – Traffic conditions

Headway	120 sec		
Nominal dwell-time at stations	25 sec		
Time extension of train service	5 %		
Nominal run-times	Inter-station run (sec)		
	0-1	1-2	2-3
	135	121	90
Control steps in run-time	0 or 7 sec	0 or 6 sec	0 or 5 sec

* Dwell-time control is not introduced in this application.

6.3.2.2 State evolution

Given the corresponding run-time extensions (i.e., 7, 6 and 5 sec) with respect to the nominal schedule in successive inter-station runs, and no dwell-time control is introduced, a new state at the stage ‘1’ can be calculated by Eqn. 6.8,

$$\begin{bmatrix} AT_3{}^2 \\ AT_2{}^3 \\ AT_1{}^4 \end{bmatrix}_1 = \begin{bmatrix} NDW_2{}^2 \\ NDW_1{}^3 \\ NDW_0{}^4 \end{bmatrix}_{0,1} + \begin{bmatrix} NRT_{2,3}{}^2 \\ NRT_{1,2}{}^3 \\ NRT_{0,1}{}^4 \end{bmatrix}_{0,1} + \begin{bmatrix} RRT_{2,3}{}^2 \\ RRT_{1,2}{}^3 \\ RRT_{0,1}{}^4 \end{bmatrix}_{0,1} + \begin{bmatrix} AT_2{}^2 \\ AT_1{}^3 \\ AT_0{}^4 \end{bmatrix}_0$$

$$= \begin{bmatrix} 25 \\ 25 \\ 25 \end{bmatrix}_{0,1} + \begin{bmatrix} 90 \\ 121 \\ 135 \end{bmatrix}_{0,1} + \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}_{0,1} + \begin{bmatrix} 401 \\ 375 \\ 335 \end{bmatrix}_0 = \begin{bmatrix} 521 \\ 527 \\ 502 \end{bmatrix}_1 \text{ sec}$$

With 120 sec headway, the arrival time of train 5 at station ‘0’, AT_0^5

$= 120 \times 4 - 25 = 455$ sec

A possible state, $[521, 527, 502, 455]_1$, at the stage ‘1’ is attained.

Excluding the initial stage, the number of possible states at a particular stage in the state-space traffic model depends on: (1) the number of possible states in the previous stage; and (2) the number of possible control actions, τ , that each state can take. They can be calculated by Eqns. 6.11 and 6.12 respectively.

$$\textit{Number of states at stage } (k+1) = \textit{Number of states at stage } k \times \tau \tag{6.11}$$

$$\tau = (\textit{Number of steps in control variable})^{\textit{Number of stations}-1} \tag{6.12}$$

A larger number of states is therefore obtained in later stages and the number of stages increases with the number of stations in the region.

6.3.3 Operation with complete state-space diagram

To describe the details of the sequence of decisions made to optimise train operation under the given operation demands with dynamic programming, the following definitions are stated:

1. $T(j) = \{T(1), \ldots, T(n)\}$ is the set of possible stage transformations at stage *j* = 1, … , *n*, where *n* is the number of inter-station runs in a region.

2. $x^g(j)$ is the g^{th} state in stage *j*, where $g = 1, \ldots, r_j$. r_j is the maximum number of states in the stage *j*.

3. $x(j) = \left\{x^1(j), \ldots, x^g(j), \ldots, x^{r_j}(j)\right\}$ is the set of possible states in stage *j*.

4. $\hat{x}(j) = \left\{\hat{x}^1(j), \ldots, \hat{x}^{h_j}(j)\right\}$ is the set of possible states in stage *j* after grouping. h_j is the maximum number of states in stage *j* after grouping. $h_j = r_j$ when no states are available to be combined through grouping. The states are numbered again when grouping is achieved.

5. $F\left(x^g(j)\right) = \left\{F\left(x^g(j)\right)_1, \ldots, F\left(x^g(j)\right)_{t_g}\right\}$ is the set of stage-to-stage cost(s) of reaching the given state '$x^g(j)$' in stage *j* from the state(s) in stage *j*-1. t_g is the number of possible states in stage '*j*-1' to reach '$x^g(j)$', where $t_g < r_{j-1}$. Cost function evaluation requires calculation of train movement under given operational constraints and requirements.

6. $M\left(x(j)\right) = \left\{\min\left(F\left(x^g(j)\right)_i + M\left(x^k(j-1)\right)\right)\right\}$ such that $x^k(j-1)$ reaches $x^g(j)$ with cost $F\left(x^g(j)\right)_i, \mathit{for}\ 1 \le g \le r_j)$. $M\left(x^k(j-1)\right)$ is the minimum cost to reach $x^k(j-1)$ from the initial stage. $F\left(x^g(j)\right)_i$ is the minimum cost to reach $x^g(j)$ from $x^k(j-1)$ and $1 \le i \le t_g$.

7. $M = \left\{M\left(x(1)\right), \ldots, M\left(x(n)\right)\right\}$ is a set of the sets of the minimum costs of reaching all possible states in successive stages.

8. $X = \left\{\left\{x^a(1), \ldots, x^b(n)\right\}; \quad \ldots\ldots; \quad \left\{x^y(1), \ldots, x^z(n)\right\}\right\}$ is a set of all the possible ordered sets of states which denote the sequence from the initial stage to the final stage in successive stage transformations (i.e., All the possible paths to

reach the final stage), where $x^a(1), x^y(1) \in x(1)$ and $x^b(n), x^z(n) \in x(n)$.

9. $x^\bullet$ is one of the elements of X.

10. $x^* = \{x^*(1), \ldots, x^*(n)\}$ represents the optimal path from the initial stage 1 to the final stage *n* under given operational requirements, where $x^* \in X$.

11. $c^s(j)$ is a set of individual dwell-time and/or run-time adjustments to the trains with respect to the corresponding nominal schedule at stage *j*, where $1 \le s \le f_j$

f_j is the maximum number of control actions at stage *j*.

For example, with the given traffic condition as shown in Table 6.1, where there are three inter-station runs.

Assume *s* = 3 and *j* = 2, Eqn. 6.9 (where, *k* = *j*, *x* = *n*= 3) then becomes

$c^3(2) = [RRT]_{1,2} = \lfloor RRT_{2,3}{}^{i+2}, RRT_{1,2}{}^{i+3}, RRT_{0,1}{}^{i+4} \rfloor_{1,2}$

where $\lfloor RRT_{2,3}{}^{i+2} \rfloor_{1,2} = 5$ or 0; $\lfloor RRT_{1,2}{}^{i+3} \rfloor_{1,2} = 6$ or 0; $\lfloor RRT_{0,1}{}^{i+4} \rfloor_{1,2} = 7$ or 0 sec

$\because c^3(2) = [RRT]_{1,2} \in c(2)$, and

$$c(2) = \{c^1(2), \quad \ldots, \quad c^8(2)\}$$

$$= \{(5,6,7), (5,6,0), (5,0,7), (5,0,0), (0,6,7), (0,6,0), (0,0,7), (0,0,0)\}$$

$\therefore c^3(2) = (5,0,7)$ sec

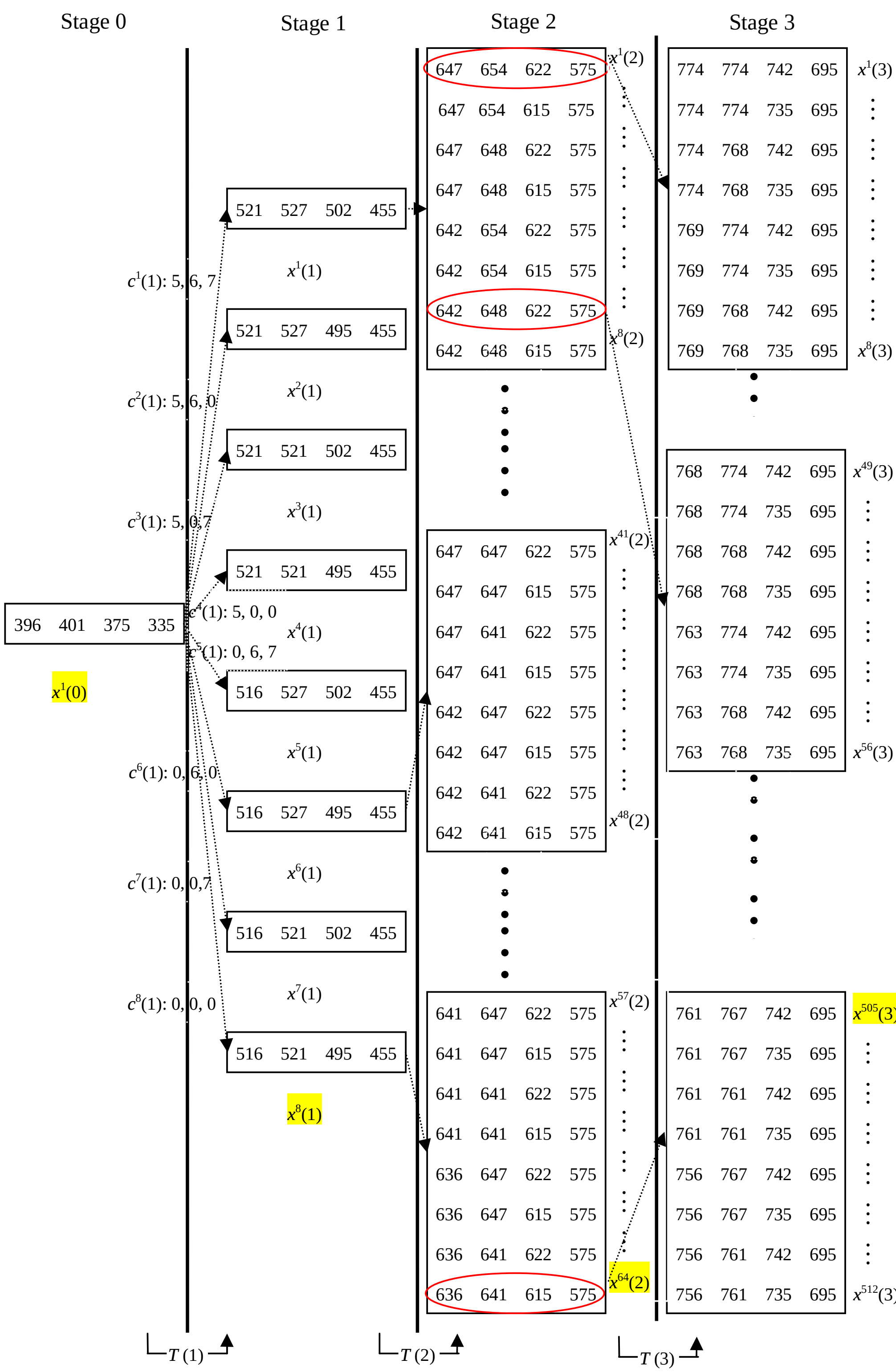
12. $c(j) = \{c^1(j), \ldots, c^{f_j}(j)\}$ is the set of control actions leading to each element in $x(j)$ from the previous stage '*j*-1'.

13. $C = \{\{c^p(1), \ldots, c^q(n)\}; \quad \ldots\ldots; \quad \{c^w(1), \ldots, c^v(n)\}\}$ is a set of all the possible ordered sets of control actions made to attain the final stage in successive stage transformations, where $c^p(1), c^w(1) \in c(1)$ and $c^q(n), c^v(n) \in c(n)$.

14. $c^\bullet$ is one of the elements of C.

15. $c^* = \{c^*(1), \ldots, c^*(n)\}$ represents the set of the optimal control actions to attain each element in x^* in successive stages, where $c^* \in C$.

Fig. 6.3 shows a simple example of the state-space traffic flow model to represent train operation with all combination of states in successive stages, in which four trains are travelling through 4 successive stations with different arrival times. '7 or (/) 0', '6 / 0' and '5 / 0' sec(s) are the three sets of two run-time extension levels in the three inter-station runs '0-1', '1-2' and '2-3' respectively, and a new train is fed into the line at station 1 with 120 sec headway. The initial state is the same as derived in Section 6.3.2.1. The traffic condition is summarised in Table 6.1. In the traffic flow model, each stage has a number of states representing the possible arrival times of trains at successive stations. A stage transformation allows one corresponding inter-station run for trains and the details of the transformation are given in Fig. 6.4. The number of stage transformations increases with the number of stations. In this example, there are a total of 512 (i.e., $2^3 \times 2^3 \times 2^3 = 8^3$) combinations of the arrival times of trains at the final stage, even with only two run-time extension levels in each of the three inter-station runs.



$c^m(k)$: x, y, z are the corresponding run-time extensions with respect to the nominal schedule in successive inter-station runs from stage k-1 to k

AT_0	AT_1	AT_2	AT_3	indicates the corresponding arrival times of trains at successive stations

Fig. 6.3 Stage-to-stage transformation with complete state-space diagram

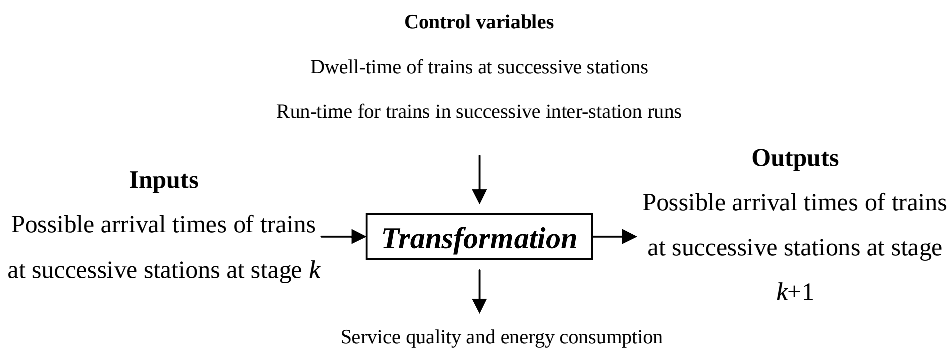


Fig. 6.4 Stage transformation

With the given traffic conditions, the sequence of transformations in successive stages and all the possible sets of control actions in C, obtained to reach the final stage in the successive stage transformations, are expressed in Eqns 6.13 and 6.14 respectively:

$$T(j) = \{T(1), T(2), T(3)\} \tag{6.13}$$

$$\begin{aligned}
C = \{&\{c^1(1),\, c^1(2),\, c^1(3)\};\ \{c^1(1), c^1(2), c^2(3)\};\quad \ldots\ldots\ldots;\quad \{c^8(1),\, c^{64}(2),\, c^{512}(3)\}\} \\
= \{&\{(5,6,7),(5,6,7),(5,6,7)\};\ \{(5,6,7),(5,6,7),(5,6,0)\};\quad \ldots;\quad \{(5,6,7),(5,6,7),(0,0,0)\}; \\
&\{(5,6,7),(5,6,0),(5,6,7)\};\ \{(5,6,7),(5,6,0),(5,6,0)\};\quad \ldots;\quad \{(5,6,7),(5,6,0),(0,0,0)\}; \\
&\{(5,6,7),(5,0,7),(5,6,7)\};\ \{(5,6,7),(5,0,7),(5,6,0)\};\quad \ldots;\quad \{(5,6,7),(5,0,7),(0,0,0)\}; \\
&\qquad\vdots \\
&\{(5,6,0),(5,6,7),(5,6,7)\};\ \{(5,6,0),(5,6,7),(5,6,0)\};\quad \ldots;\quad \{(5,6,7),(5,6,7),(0,0,0)\}; \\
&\{(5,6,0),(5,6,0),(5,6,7)\};\ \{(5,6,0),(5,6,0),(5,6,0)\};\quad \ldots;\quad \{(5,6,0),(5,6,0),(0,0,0)\}; \\
&\qquad\vdots \\
&\qquad\vdots \\
&\{(0,0,0),(5,6,7),(5,6,7)\};\ \{(0,0,0),(5,6,7),(5,6,0)\};\quad \ldots;\quad \{(0,0,0),(5,6,7),(0,0,0)\}; \\
&\{(0,0,0),(5,6,0),(5,6,7)\};\ \{(0,0,0),(5,6,0),(5,6,0)\};\quad \ldots;\quad \{(0,0,0),(5,6,0),(0,0,0)\}; \\
&\qquad\vdots \\
&\{(0,0,0),(0,0,0),(5,6,7)\};\ \{(0,0,0),(0,0,0),(5,6,0)\};\quad \ldots;\quad \{(0,0,0),(0,0,0),(0,0,0)\}\}
\end{aligned} \tag{6.14}$$

Given combinations of the control actions in C, the possible sets of states $x^{\bullet}$ made in successive stages with the corresponding control actions $c^{\bullet}$ in C can then be defined as:

$$
\begin{aligned}
c^{\bullet} &= \{(5,6,7),(5,6,7),(5,6,7)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{1}(2), x^{1}(3)\} \\
c^{\bullet} &= \{(5,6,7),(5,6,7),(5,6,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{1}(2), x^{2}(3)\} \\
&\vdots \\
c^{\bullet} &= \{(5,6,7),(5,6,7),(0,0,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{1}(2), x^{8}(3)\} \\
c^{\bullet} &= \{(5,6,7),(5,6,0),(5,6,7)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{2}(2), x^{9}(3)\} \\
c^{\bullet} &= \{(5,6,7),(5,6,0),(5,6,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{2}(2), x^{10}(3)\} \\
&\vdots \\
c^{\bullet} &= \{(5,6,7),(5,6,0),(0,0,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{1}(1), x^{2}(2), x^{16}(3)\} \\
&\vdots \\
&\vdots \\
c^{\bullet} &= \{(5,6,0),(5,6,7),(5,6,7)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{2}(1), x^{9}(2), x^{65}(3)\} \\
c^{\bullet} &= \{(5,6,0),(5,6,7),(5,6,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{2}(1), x^{9}(2), x^{66}(3)\} \\
&\vdots \\
c^{\bullet} &= \{(5,6,0),(5,6,7),(0,0,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{2}(1), x^{9}(2), x^{72}(3)\} \\
&\vdots \\
&\vdots \\
c^{\bullet} &= \{(0,0,0),(0,0,0),(5,6,7)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{8}(1), x^{64}(2), x^{505}(3)\} \\
c^{\bullet} &= \{(0,0,0),(0,0,0),(5,6,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{8}(1), x^{64}(2), x^{506}(3)\} \\
&\vdots \\
c^{\bullet} &= \{(0,0,0),(0,0,0),(0,0,0)\} \quad \rightarrow \quad x^{\bullet} = \{x^{1}(0), x^{8}(1), x^{64}(2), x^{512}(3)\}
\end{aligned}
$$

For example, when the control actions $c^{\bullet} = \{(0,0,0),(0,0,0),(5,6,7)\}$ are obtained in successive stage transformations and train 1 is assumed to depart at station '0' when time = 0 sec. The corresponding states made in successive stages are:

$$x^{\bullet} = \{x^{1}(0),\ x^{8}(1),\ x^{64}(2),\ x^{505}(3)\}$$

where,

$x^1(0) = \{396, 401, 375, 335\}$ is the initial state (see Section 6.3.2.1)

$x^8(1) = x^1(0) + [NRT]_{0,1} + [RRT]_{0,1} + [NDW]_{0,1}$

$$= \begin{Bmatrix} AT_3{}^2 \\ AT_2{}^3 \\ AT_1{}^4 \\ AT_0{}^5 \end{Bmatrix} + \begin{Bmatrix} 401 & + & 90 & + & 0 & + & 25 \\ 375 & + & 121 & + & 0 & + & 25 \\ 335 & + & 135 & + & 0 & + & 25 \\ & 120 & \times & 4 & & - & 25 \end{Bmatrix} = \begin{Bmatrix} 516 \\ 521 \\ 495 \\ 455 \end{Bmatrix}$$

$x^{64}(2) = x^8(1) + [NRT]_{1,2} + [RRT]_{1,2} + [NDW]_{1,2}$

$$= \begin{Bmatrix} AT_3{}^3 \\ AT_2{}^4 \\ AT_1{}^5 \\ AT_0{}^6 \end{Bmatrix} + \begin{Bmatrix} 521 & + & 90 & + & 0 & + & 25 \\ 495 & + & 121 & + & 0 & + & 25 \\ 455 & + & 135 & + & 0 & + & 25 \\ & 120 & \times & 5 & & - & 25 \end{Bmatrix} = \begin{Bmatrix} 636 \\ 641 \\ 615 \\ 575 \end{Bmatrix}$$

$x^{505}(3) = x^{64}(2) + [NRT]_{2,3} + [RRT]_{2,3} + [NDW]_{2,3}$

$$= \begin{Bmatrix} AT_3{}^4 \\ AT_2{}^5 \\ AT_1{}^6 \\ AT_0{}^7 \end{Bmatrix} + \begin{Bmatrix} 641 & + & 90 & + & 5 & + & 25 \\ 615 & + & 121 & + & 6 & + & 25 \\ 575 & + & 135 & + & 7 & + & 25 \\ & 120 & \times & 6 & & - & 25 \end{Bmatrix} = \begin{Bmatrix} 761 \\ 767 \\ 742 \\ 695 \end{Bmatrix}$$

and the given states are highlighted in Fig. 6.3.

With reference to the complete state-space diagram in Fig. 6.3, the total number of possible states at stages 1, 2 and 3 are 8, 64 and 512 respectively. State $x^1(1)$ and $x^5(1)$, in which they have two different sets of arrival-times of trains at stations (i.e., $[521, 527, 502, 455]$ and $[516, 527, 502, 455]$), produce two sets of 8-state $x^1(2) \ldots x^8(2)$ and $x^{33}(2) \ldots x^{40}(2)$ at stage 2 with the corresponding control actions. Similarly, the other ordered states in $x(1)$, where the arrival-time elements are not the same as each other, provide the corresponding set of states at stage 2 as follows:

$$x^2(1) = [521, 527, 495, 455] \rightarrow x^9(2) \ldots x^{16}(2) \text{ and}$$

$$x^6(1) = [516, 527, 495, 455] \rightarrow x^{41}(2) \ldots x^{48}(2) \;;$$

$$x^3(1) = [521, 521, 502, 455] \rightarrow x^{17}(2) \ldots x^{24}(2) \text{ and}$$

$$x^7(1) = [516, 521, 502, 455] \rightarrow x^{49}(2) \ldots x^{56}(2) \;;$$

$$x^4(1) = [521, 521, 495, 455] \rightarrow x^{25}(2) \ldots x^{32}(2) \text{ and}$$

$$x^8(1) = [516, 521, 495, 455] \rightarrow x^{57}(2) \ldots x^{64}(2) \;;$$

To further elaborate on state evolution at stage 3, an example is shown below:

$$x^1(2) = [647, 654, 622, 575] \rightarrow x^1(3) \ldots x^8(3) \text{ and}$$

$$x^{33}(2) = [647, 654, 622, 575] \rightarrow x^{257}(3) \ldots x^{263}(3);$$

$$x^5(2) = [642, 654, 622, 575] \rightarrow x^{33}(3) \ldots x^{40}(3) \text{ and}$$

$$x^{37}(2) = [642, 654, 622, 575] \rightarrow x^{289}(3) \ldots x^{296}(3);$$

$$x^{17}(2) = [641, 654, 622, 575] \rightarrow x^{129}(3) \ldots x^{136}(3) \text{ and}$$

$$x^{49}(2) = [641, 654, 622, 575] \rightarrow x^{385}(3) \ldots x^{392}(3);$$

$$x^{21}(2) = [636, 654, 622, 575] \rightarrow x^{161}(3) \ldots x^{168}(3) \text{ and}$$

$$x^{53}(2) = [636, 654, 622, 575] \rightarrow x^{417}(3) \ldots x^{424}(3)$$

Even though the optimal solution of train operation can be obtained with the complete state-space diagram, computational demand is inevitably heavy and memory storage requirement become extreme as the number of inter-station runs increases. With Eqns. 6.11 and 6.12, the number of stage transformations in the traffic flow model increases with the number of inter-station runs and the number of states generated escalates with the control actions in later stages.

To minimise the extensive state increment in the state-space traffic flow model in later stages, a certain measure is therefore necessary to reduce the number of states at the intermediate stages for real-time application. Grouping of states in a stage is a viable means.

6.3.4 Operation with state grouping

Fig. 6.3 demonstrates that the number of possible states at the intermediate stages increases significantly with the number of stages. A simplification to reduce the number of states in the model is possible. The notion of state grouping is based on combining some of the states, which result in the same arrival times at successive stations, into a single state at a particular stage prior to the stage optimisation. The optimal solution remains achievable with state grouping.

With reference to the complete state-space traffic flow model, Fig. 6.5 shows that state $x^1(1)$ and $x^5(1)$ from stage 1 produce the same set of states at stage 2 (i.e., $x^1(2)\ldots x^8(2)$ and $x^{33}(2)\ldots x^{40}(2)$) through different control actions. That is, in each pair of state $x^1(2)$ and $x^{33}(2)$, $x^2(2)$ and $x^{34}(2)$, $x^3(2)$ and $x^{35}(2)$, ………, $x^8(2)$ and $x^{40}(2)$, they have the same corresponding arrival-time elements. These 8 pairs of states can be combined together to form new states $\hat{x}^1(2)\ldots\hat{x}^8(2)$ at stage 2 through state grouping, based on the comparison of $M\left(x^1(2)\right)$ and $M\left(x^{33}(2)\right)$, ……, $M\left(x^8(2)\right)$ and $M\left(x^{40}(2)\right)$. With state grouping, for each new state, $x^g(2)$ (i.e., each of $\hat{x}^1(2)\ldots\hat{x}^{32}(2)$), in this example, the corresponding set of $F\left(x^g(2)\right)$, is formed. Similarly, the sets of cost reaching each new state,

$\hat{x}^1(2)\ldots\hat{x}^{64}(2)$, at stage 3 are attained. State grouping is performed for each stage and the number of states in each stage is significantly reduced.

Details of the possible state simplification at stages 2 and 3 are listed in Tables 6.2a and 6.2b respectively. The reduction in the number of states in stage 3 is even more significant. With state grouping, the stage-to-stage transformation with simplified state-space diagram is given in Fig. 6.6. To further illustrate the advantage of state grouping, Table 6.3 shows the total number of possible states before and after grouping with different number of control variable steps and inter-station runs.

Table 6.2a – State reduction at stage 2 with state grouping

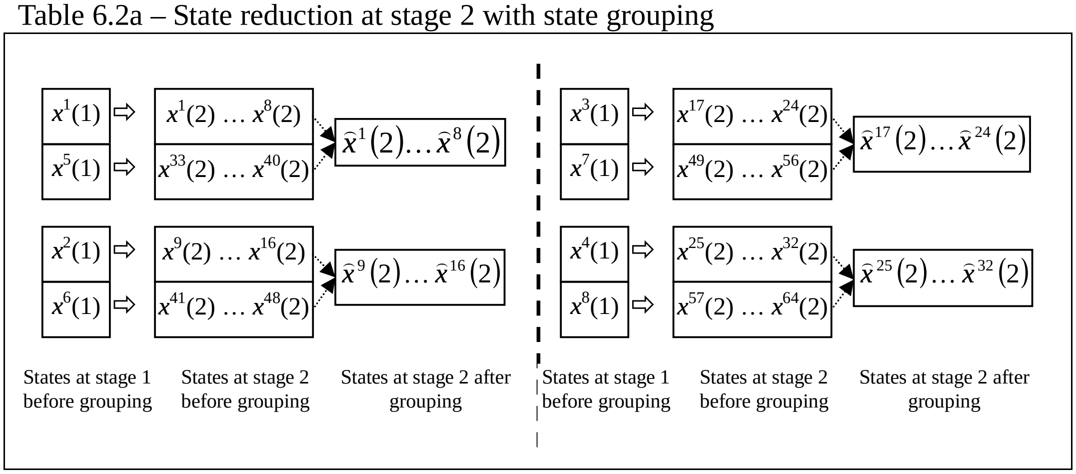


Table 6.2b – State reduction at stage 3 with state grouping

States at stage 2 before grouping	States at stage 3 before grouping	States at stage 3 after grouping	States at stage 2 before grouping	States at stage 3 before grouping	States at stage 3 after grouping
$x^1(2)$	$x^1(3) \dots x^8(3)$		$x^9(2)$	$x^{65}(3) \dots x^{72}(3)$	
$x^5(2)$	$x^{33}(3) \dots x^{40}(3)$		$x^{13}(2)$	$x^{97}(3) \dots x^{104}(3)$	
$x^{17}(2)$	$x^{129}(3) \dots x^{136}(3)$		$x^{25}(2)$	$x^{193}(3) \dots x^{200}(3)$	
$x^{21}(2)$	$x^{161}(3) \dots x^{168}(3)$	$\hat{x}^1(3) \dots \hat{x}^8(3)$	$x^{29}(2)$	$x^{225}(3) \dots x^{232}(3)$	$\hat{x}^{33}(3) \dots \hat{x}^{40}(3)$
$x^{33}(2)$	$x^{257}(3) \dots x^{264}(3)$		$x^{41}(2)$	$x^{321}(3) \dots x^{328}(3)$	
$x^{37}(2)$	$x^{289}(3) \dots x^{296}(3)$		$x^{45}(2)$	$x^{353}(3) \dots x^{360}(3)$	
$x^{49}(2)$	$x^{385}(3) \dots x^{392}(3)$		$x^{57}(2)$	$x^{449}(3) \dots x^{456}(3)$	
$x^{53}(2)$	$x^{417}(3) \dots x^{424}(3)$		$x^{61}(2)$	$x^{481}(3) \dots x^{488}(3)$	
$x^2(2)$	$x^9(3) \dots x^{16}(3)$		$x^{10}(2)$	$x^{73}(3) \dots x^{80}(3)$	
$x^6(2)$	$x^{41}(3) \dots x^{48}(3)$		$x^{14}(2)$	$x^{105}(3) \dots x^{112}(3)$	
$x^{18}(2)$	$x^{137}(3) \dots x^{144}(3)$		$x^{26}(2)$	$x^{201}(3) \dots x^{208}(3)$	
$x^{22}(2)$	$x^{169}(3) \dots x^{176}(3)$	$\hat{x}^9(3) \dots \hat{x}^{16}(3)$	$x^{30}(2)$	$x^{233}(3) \dots x^{240}(3)$	$\hat{x}^{41}(3) \dots \hat{x}^{48}(3)$
$x^{34}(2)$	$x^{265}(3) \dots x^{272}(3)$		$x^{42}(2)$	$x^{329}(3) \dots x^{336}(3)$	
$x^{38}(2)$	$x^{297}(3) \dots x^{304}(3)$		$x^{46}(2)$	$x^{361}(3) \dots x^{368}(3)$	
$x^{50}(2)$	$x^{393}(3) \dots x^{400}(3)$		$x^{58}(2)$	$x^{457}(3) \dots x^{464}(3)$	
$x^{54}(2)$	$x^{425}(3) \dots x^{432}(3)$		$x^{62}(2)$	$x^{489}(3) \dots x^{496}(3)$	
$x^3(2)$	$x^{17}(3) \dots x^{24}(3)$		$x^{11}(2)$	$x^{81}(3) \dots x^{88}(3)$	
$x^7(2)$	$x^{49}(3) \dots x^{56}(3)$		$x^{15}(2)$	$x^{113}(3) \dots x^{120}(3)$	
$x^{19}(2)$	$x^{145}(3) \dots x^{152}(3)$		$x^{27}(2)$	$x^{209}(3) \dots x^{216}(3)$	
$x^{23}(2)$	$x^{177}(3) \dots x^{184}(3)$	$\hat{x}^{17}(3) \dots \hat{x}^{24}(3)$	$x^{31}(2)$	$x^{241}(3) \dots x^{248}(3)$	$\hat{x}^{49}(3) \dots \hat{x}^{56}(3)$
$x^{35}(2)$	$x^{273}(3) \dots x^{280}(3)$		$x^{43}(2)$	$x^{337}(3) \dots x^{344}(3)$	
$x^{39}(2)$	$x^{305}(3) \dots x^{312}(3)$		$x^{47}(2)$	$x^{369}(3) \dots x^{376}(3)$	
$x^{51}(2)$	$x^{401}(3) \dots x^{408}(3)$		$x^{59}(2)$	$x^{465}(3) \dots x^{472}(3)$	
$x^{55}(2)$	$x^{433}(3) \dots x^{440}(3)$		$x^{63}(2)$	$x^{497}(3) \dots x^{504}(3)$	
$x^4(2)$	$x^{25}(3) \dots x^{32}(3)$		$x^{12}(2)$	$x^{89}(3) \dots x^{96}(3)$	
$x^8(2)$	$x^{57}(3) \dots x^{64}(3)$		$x^{16}(2)$	$x^{121}(3) \dots x^{128}(3)$	
$x^{20}(2)$	$x^{153}(3) \dots x^{160}(3)$		$x^{28}(2)$	$x^{217}(3) \dots x^{224}(3)$	
$x^{24}(2)$	$x^{185}(3) \dots x^{192}(3)$	$\hat{x}^{25}(3) \dots \hat{x}^{32}(3)$	$x^{32}(2)$	$x^{249}(3) \dots x^{256}(3)$	$\hat{x}^{57}(3) \dots \hat{x}^{64}(3)$
$x^{36}(2)$	$x^{281}(3) \dots x^{288}(3)$		$x^{44}(2)$	$x^{345}(3) \dots x^{352}(3)$	
$x^{40}(2)$	$x^{313}(3) \dots x^{320}(3)$		$x^{48}(2)$	$x^{377}(3) \dots x^{384}(3)$	
$x^{52}(2)$	$x^{409}(3) \dots x^{416}(3)$		$x^{60}(2)$	$x^{473}(3) \dots x^{480}(3)$	
$x^{56}(2)$	$x^{441}(3) \dots x^{448}(3)$		$x^{64}(2)$	$x^{505}(3) \dots x^{512}(3)$	

Table 6.3 – Number of possible states at the final stage before and after grouping

Number of inter-station runs	Number of states at the final stage before grouping		
	Number of steps in control variable*		
	2	3	4
1	2	3	4
2	16	81	256
3	512	19683	262144
4	65536	43046721	4294967296
5	33554432	$8.472886094 \times 10^{11}$	$1.125899907 \times 10^{15}$
Number of inter-station runs	Number of states at the final stage after grouping		
	Number of steps in control variable*		
	2	3	4
1	2	3	4
2	8	27	64
3	64	729	4096
4	1024	59049	1048576
5	32768	14348907	1073741824

* Run-time adjustment is the control variable in successive inter-station runs in this application.

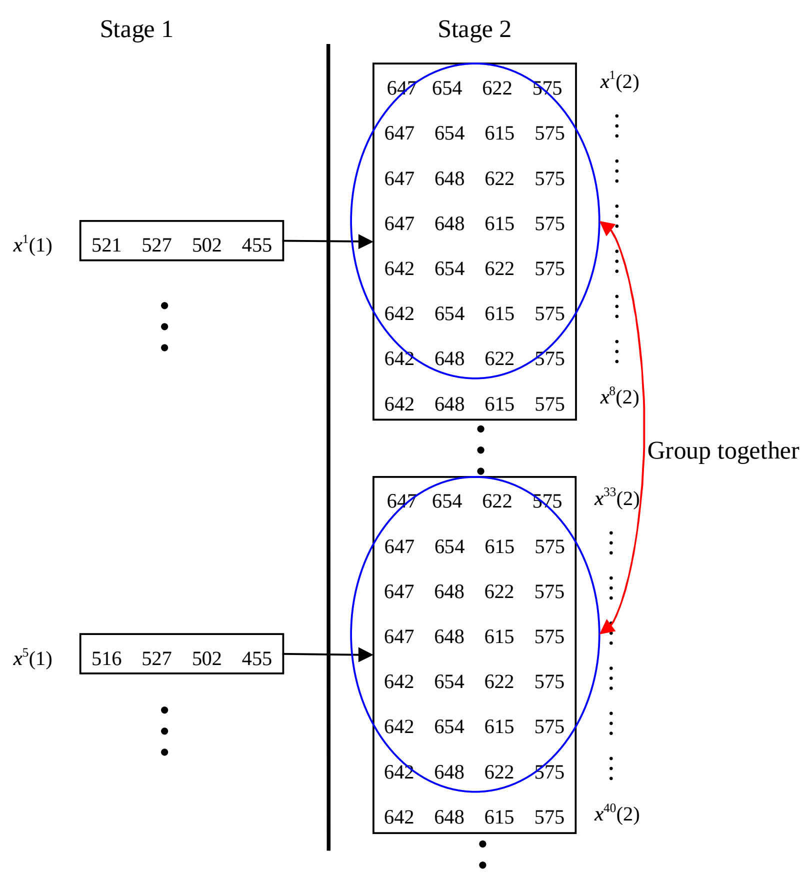
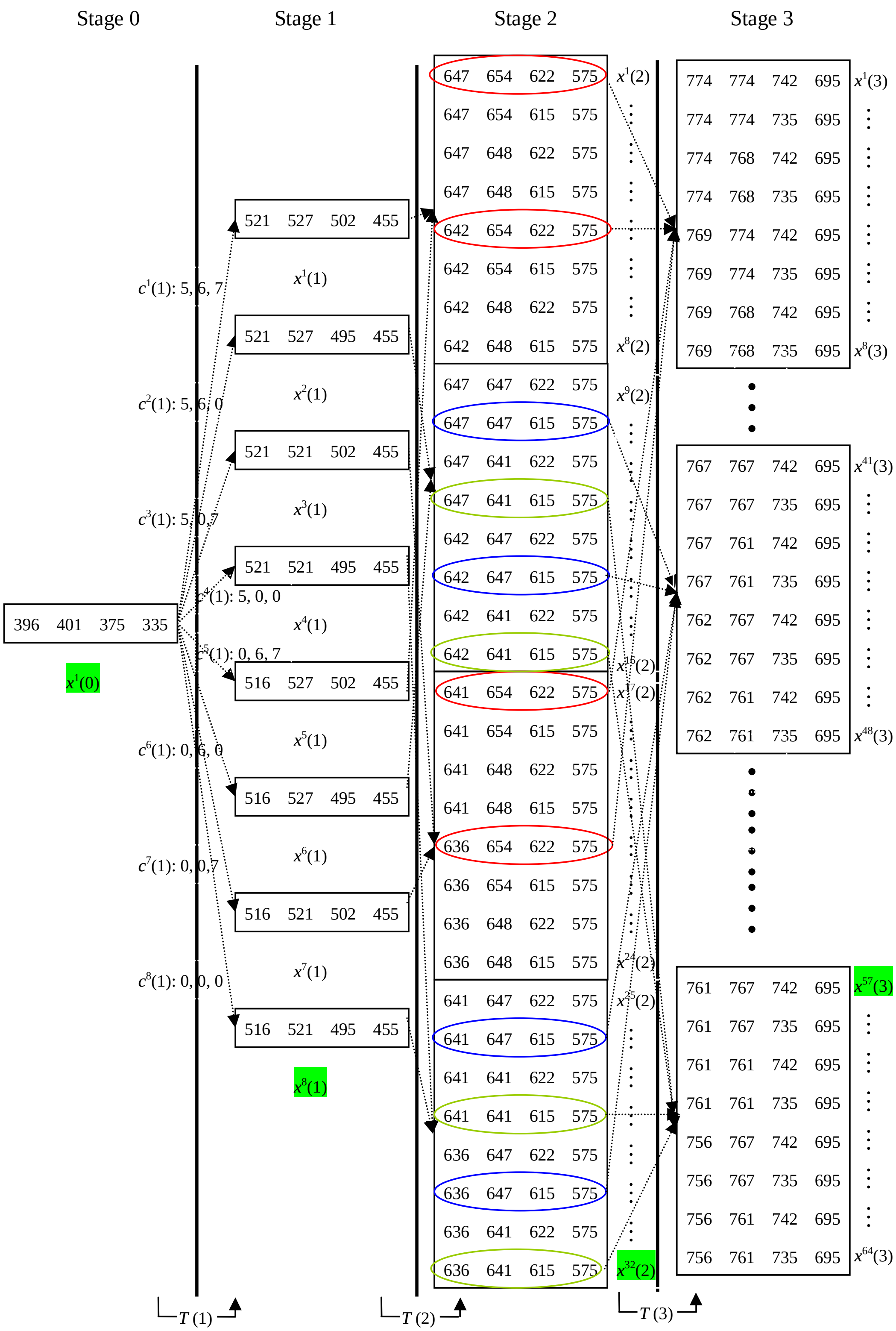


Fig. 6.5 State reduction through grouping



$c^m(k)$: x, y, z are the corresponding run-time extensions with respect to the nominal schedule in successive inter-station runs from stage k-1 to k

AT_0	AT_1	AT_2	AT_3

indicates the corresponding arrival times of trains at successive stations

Fig. 6.6 Stage-to-stage transformation with simplified state-space diagram

6.3.5 Optimal path

With reference to Fig. 6.6, there are two run-time extension levels (i.e., control steps) in each of the three inter-station runs, i.e., either 0 or (/) 5 sec in inter-station 1; 0 / 6 sec in inter-station run 2; and 0 / 7 sec in the inter-station run 3. The run-time is extended by 5% with respect to the nominal train service. To demonstrate the approach to obtain the optimal solution with the given operational requirements in DP, $\{F(x^g(1)),\ for\ 1 \le g \le 8\}$, $\{F(x^g(2)),\ for\ 1 \le g \le 32\}$ and $\{F(x^g(3)),\ for\ 1 \le g \le 64\}$ denote the three sets of costs of reaching all the possible states $\{x^1(1), \ldots, x^8(1)\}$, $\{x^1(2), \ldots, x^{32}(2)\}$ and $\{x^1(3), \ldots, x^{64}(3)\}$ in the successive stages respectively. At the beginning, the minimum cost $M(x(1)) = \{\min F(x^g(1)),\ for\ 1 \le g \le 8\}$ of reaching states $\{x^1(1), \ldots, x^8(1)\}$ in stage 1 are computed first. $\{x^1(1), \ldots, x^8(1)\}$ can then be evolved with the corresponding run-time extensions of (5,6,7), (5,6,0), (5,0,7), (5,0,0), (0,6,7), (0,0,7) and (0,0,0) sec (i.e., $c(2)$) in successive inter-station runs, from the initial stage. Determination of the minimum cost, $M(x(j))$, of reaching the given states $x^g(j)$ is illustrated in Fig. 6.7. Since there is only one initial state in stage 0, no optimisation is required. The control action, $c^*(1)$, which minimizes $\{F(x^g(1)),\ for\ 1 \le g \le 8\}$ is recorded.

With the minimum cost $M(x(1))$ for each element in $x(1)$, the set $M(x(2))$ can be deduced as shown in Table 6.4a. The states in stage 1 reaching each element in $x(2)$ are available, and the corresponding costs are denoted by $F(x^g(2))$. Table 6.4b also reveals the minimum cost, $M(x(3))$, to reach all the possible states in $x(3)$ in stage 3. When the control action $c^*(3)$ is deduced, the overall optimal path with the corresponding control decisions and states in successive stages is obtained.

$$\textit{Optimal path}: \quad x^*(3) \xrightarrow{c^*(3)} x^*(2) \xrightarrow{c^*(2)} x^*(1) \xrightarrow{c^*(1)} x^*(0) \tag{6.15}$$

With the given traffic conditions and operational requirements in Table 6.1, the optimal path is $x^1(0) \rightarrow x^8(1) \rightarrow x^{32}(2) \rightarrow x^{57}(3)$, which is highlighted in the simplified state-space diagram in Fig. 6.6. The corresponding control actions of run-time extension made in each of the three inter-station runs are (0,0,0) sec in the stage transformations '0 to 1' and '1 to 2', while they are (5,6,7) sec in stage transformation '2 to 3'.

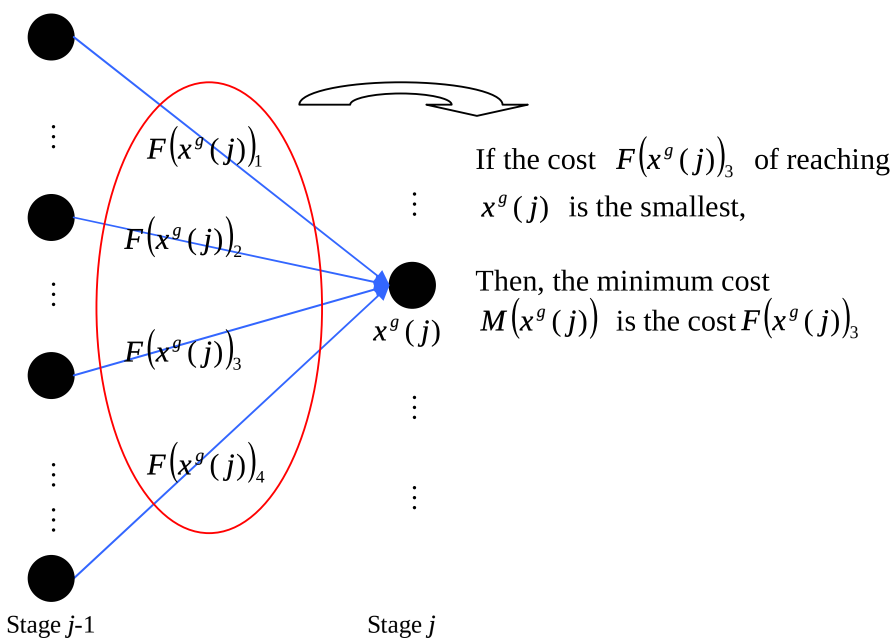


Fig. 6.7 Minimum cost $M(x^g(j))$ of reaching state $x^g(j)$ in stage j

Table 6.4a – Optimisation at stage 2 with state grouping

States in stage 2	Possible states in $x(1)$ to reach each element in $x(2)$	Minimum cost $M(x(2))$ in stage 2	Control action
$x^1(2)$	$x^1(1)$ $x^5(1)$	$\underset{c(2)}{Min}\left\{\begin{matrix} F(x^1(2))_1 + M(x^1(1)) \\ F(x^1(2))_2 + M(x^5(1)) \end{matrix}\right\}$	(5,6,7)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x^{10}(2)$	$x^2(1)$ $x^6(1)$	$\underset{c(2)}{Min}\left\{\begin{matrix} F(x^{10}(2))_1 + M(x^2(1)) \\ F(x^{10}(2))_2 + M(x^6(1)) \end{matrix}\right\}$	(5,6,0)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x^{19}(2)$	$x^3(1)$ $x^7(1)$	$\underset{c(2)}{Min}\left\{\begin{matrix} F(x^{19}(2))_1 + M(x^3(1)) \\ F(x^{19}(2))_2 + M(x^7(1)) \end{matrix}\right\}$	(5,0,7)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x^{28}(2)$	$x^4(1)$ $x^8(1)$	$\underset{c(2)}{Min}\left\{\begin{matrix} F(x^{28}(2))_1 + M(x^4(1)) \\ F(x^{28}(2))_2 + M(x^8(1)) \end{matrix}\right\}$	(5,0,0)

Table 6.4b – Optimisation at stage 3 with state grouping

States in stage 3	Possible states in $x(2)$ to reach each element in $x(3)$	Minimum cost $M(x(3))$ in stage 3	Control action
$x^1(3)$	$x^1(2)$ $x^5(2)$ $x^{17}(2)$ $x^{21}(2)$	$\underset{c(3)}{Min}\left\{\begin{matrix} F(x^1(3))_1 + M(x^1(2)) \\ F(x^1(3))_2 + M(x^5(2)) \\ F(x^1(3))_3 + M(x^{17}(2)) \\ F(x^1(3))_4 + M(x^{21}(2)) \end{matrix}\right\}$	(5,6,7)
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x^{42}(3)$	$x^{10}(2)$ $x^{14}(2)$ $x^{26}(2)$ $x^{30}(2)$	$\underset{c(3)}{Min}\left\{\begin{matrix} F(x^{42}(3))_1 + M(x^{10}(2)) \\ F(x^{42}(3))_2 + M(x^{14}(2)) \\ F(x^{42}(3))_3 + M(x^{26}(2)) \\ F(x^{42}(3))_4 + M(x^{30}(2)) \end{matrix}\right\}$	(5,6,0)
$\vdots$ $\vdots$	$\vdots$ $\vdots$	$\vdots$ $\vdots$	$\vdots$ $\vdots$
$x^{64}(3)$	$x^{12}(2)$ $x^{16}(2)$ $x^{28}(2)$ $x^{32}(2)$	$\underset{c(3)}{Min}\left\{\begin{matrix} F(x^{64}(3))_1 + M(x^{12}(2)) \\ F(x^{64}(3))_2 + M(x^{16}(2)) \\ F(x^{64}(3))_3 + M(x^{28}(2)) \\ F(x^{64}(3))_4 + M(x^{32}(2)) \end{matrix}\right\}$	(0,0,0)

Note: The total number of states is 1, 8, 64, and 512 respectively at stages 0 to 3, when state grouping is not adopted. However, with state grouping, the total number of states in $x(2)$ is reduced from 64 to 32. These 32 states are in eight 4-state sets, and each set produces 8 states at stage 3 with the corresponding control actions. Hence, with the eight 4-state sets, there are finally 64 states in $x(3)$, and the ratio of reduction on the number of states at stage 3 is $\frac{512}{64} = 8$.

6.3.6 Cost functions

To evaluate the possible optimal solution with the given operation demand, an objective function is needed to determine how well the chosen control actions lead to the desired service quality and/or energy consumption of train operation. Punctuality and regularity are the prime concerns for passengers and operators in maintaining a regular and highly stable train service. Three objective functions on service quality, and one on energy consumption, as well as a combined one, are defined to determine the stage-to-stage cost in the state-space traffic flow model:

6.3.6.1 Service quality : Passenger expectation (I)

In general, service quality is more easily affected in a short inter-station run, with the same amount of run-time and dwell-time regulation, than in a long inter-station run. This is because the time regulation, either with dwell-time or run-time control, takes a larger proportion of the total operation time (i.e., nominal dwell-time and run-time) in short inter-station runs. The stage-to-stage cost function of the service quality, as shown in Eqn. 6.16, is therefore proposed. It first determines a set of costs for train operation, $\left(\frac{\left| T_D{}^i - T_A{}^i \right|}{T_D{}^i}, \; for\; 1 \le i \le k \right)$, for successive inter-station runs with the corresponding control actions. It indicates how well the actual train service adheres to the expected one. With the given set of costs, the overall average cost, $Cost_{ser}$, in a region is deduced. The function returns a lower value for a better train service.

$$Cost_{Ser} = \frac{1}{k}\left(\sum_{i=1}^{k} \frac{\left| T_D{}^i - T_A{}^i \right|}{T_D{}^i}\right) \tag{6.16}$$

$$T_D{}^i = (T_{SRT}{}^i + T_{SDW}{}^i) \times (1+\varphi)$$

$$T_A{}^i = \begin{cases} (T_{SRT}{}^i + T_{SDW}{}^i + T_{CDW}{}^i); & Dwell-time\ control \\ (T_{SRT}{}^i + T_{SDW}{}^i + T_{CRT}{}^i); & Run-time\ control \\ (T_{SRT}{}^i + T_{SDW}{}^i + T_{CDW}{}^i + T_{CRT}{}^i) & Mixed \end{cases}$$

Subject to: $T_{CDW}{}^i \le T_{SDW}{}^i$;

$$T_{CRT}{}^i \le T_{SRT}{}^i \times \tau$$

where,

k is the total number of inter-station runs in a region.

$T_D{}^i$ is the desired operation time (dwell-time plus run-time) of the train in the i^{th} inter-station run (i.e., train runs from station i to i+1).

$T_A{}^i$ is the actual run-time and dwell-time of the train in the i^{th} inter-station run.

$T_{SDW}{}^i$ is the nominal dwell-time of the train at station i.

$T_{SRT}{}^i$ is the nominal run-time of the train in the i^{th} inter-station run.

φ is the percentage of the operation time deviation with respect to the nominal schedule in an inter-station run. φ can be either positive or negative. For example, φ is a positive value when the operation time of a train is prolonged in an inter-station run.

$T_{CDW}{}^{i}$ is the dwell-time adjustment of the train at station *i* with respect to the nominal dwell-time.

$T_{CRT}{}^{i}$ is the run-time regulation of the train in the i^{th} inter-station with respect to the nominal run-time.

τ is the coefficient of run-time regulation between stations with respect to nominal run-time. It can be set by operators to determine the range of run-time regulation in the simulation.

6.3.6.2 Service quality: Passenger expectation (II)

A longer inter-station run-time is more acceptable to passengers than a longer dwell-time at a station, if ever the overall travelling time has to be lengthened. An energy saving can also be achieved with a run-time extension in an inter-station run. With Eqn. 6.17, the cost is increased when regulation of train service with dwell-time control is applied. The function is very similar to that given in Eqn. 6.16, and the weight W_i, for the additional dwell-time imposed on the quality of train service in an inter-station run is introduced. W_i is one when only run-time regulation is adopted and it becomes larger than one when additional dwell-time at stations is introduced and the corresponding cost is thereby increased in a particular run.

$$Cost_{Ser} = \frac{1}{k}\left(\sum_{i=1}^{k} W_i \frac{\left|T_D{}^{i} - T_A{}^{i}\right|}{T_D{}^{i}}\right) \qquad (6.17)$$

$$W_i = \frac{T_{SDW}{}^{i} + T_{SRT}{}^{i} + T_{CDW}{}^{i}}{T_{SDW}{}^{i} + T_{SRT}{}^{i}};$$

$$T_D{}^{i} = (T_{SRT}{}^{i} + T_{SDW}{}^{i}) \times (1+\varphi)$$

$$T_A{}^i = \begin{cases} (T_{SRT}{}^i + T_{SDW}{}^i + T_{CDW}{}^i); & Dwell-time\ control \\ (T_{SRT}{}^i + T_{SDW}{}^i + T_{CRT}{}^i); & Run-time\ control \\ (T_{SRT}{}^i + T_{SDW}{}^i + T_{CDW}{}^i + T_{CRT}{}^i) & Mixed \end{cases}$$

Subject to: $T_{CDW}{}^i \le T_{SDW}{}^i$;

$$T_{CRT}{}^i \le T_{SRT}{}^i \times \tau$$

where,

W_i is the corresponding weight of the additional dwell-time imposed on the train at station *i*.

Other definitions remain the same as stated in Eqn. 6.16.

6.3.6.3 Service quality: Operators

Service regularity is the major concern to passengers and operators when accessing quality of train service. Suppose all trains are carrying the same traction characteristics and the separations of trains are scheduled to be the same, headway is the indicator of the service regularity. The following cost function penalises deviation from the nominal headway, and a lower cost implies the chosen solution is closer to the desired service.

$$Cost_{Ser} = \frac{\sum_{i=0}^{k} \left| \frac{(H_d \times (1+\varepsilon)) - (T_{k-i}{}^{i+2} - T_{k-i}{}^{i+1})}{H_d \times (1+\varepsilon)} \right|}{k} \tag{6.18}$$

where,

k is the total number of inter-station runs in a region.

H_d is the nominal headway between the trains.

ε is the maximum allowable percentage of headway deviation from the nominal schedule. ε can be either positive or negative. Headway of trains is lengthened when ε is positive.

T_{k-i}^{i+2} and T_{k-i}^{i+1} are the arrival times of trains '*i*+2' and '*i*+1' at the station '*k-i*' respectively. The headway of trains is therefore the arrival-time difference between the two trains at station '*k-i*'.

6.3.6.4 Energy consumption

Energy supplied by the power supply system is the other major concern for the operators. A higher energy consumption implies a higher operation cost. The following cost function of energy consumption is given to encourage energy reduction in the inter-station run.

$$Cost_{Energy} = Sgn\left(\frac{1}{k}\frac{\sum_{r=1}^{k} E_A{}^r - (1+\vartheta)\sum_{r=1}^{k} E_S{}^r}{(1+\vartheta)\sum_{r=1}^{k} E_S{}^r}\right) \tag{6.19}$$

where,

k is the total number of inter-station runs in a region.

$E_S{}^r$ is the energy consumption of a train with the nominal run-time in an inter-station run '*r*'.

$E_A{}^r$ is the actual energy consumption of a train in an inter-station run '*r*'.

ϑ is the percentage of energy consumption deviation with respect to that in the nominal run-time. ϑ can be set as either positive or negative by the operators. Energy reduction is attained with the corresponding run-time extension of train when ϑ is negative.

The energy cost function is only applicable for run-time control.

6.3.6.5 Overall cost function

To reflect the relative importance of service quality and energy consumption on the overall cost function, the following expression is adopted.

$$Cost_{Overall} = W_{Ser} \cdot Cost_{Ser} + W_{Energy} \cdot Cost_{Energy} \qquad (6.20)$$

Subject to: $W_{Ser} + W_{Energy} = 1$

$$0 \le W_{Ser} \le 1$$

$$0 \le W_{Energy} \le 1$$

W_{Ser} *and* W_{Energy} are the corresponding weightings assigned to service quality and energy consumption respectively. The possible range of the overall cost function is between 0 and 1.

6.4 Operation with disturbance

6.4.1 Arrival time adjustments

The approach to obtain the optimal states for trains with the specified operational requirements through DP is given in Section 6.3.5. Trains run along the line to provide service for passengers according to the arrival times at stations. However, service disruption to trains is almost unavoidable in reality. For instance, trains may stop at station with an additional dwell-time for unexpected reasons, like time extension due to interruption of door closure. The optimal set of states obtained through DP, or called ‘DP-I’ in later discussions, in conditions without disruption, is therefore no longer valid for train operation. New states are needed to replace the states before disruption in order to allow for another formulation of optimal control action by DP to achieve service recovery. The process to obtain the new states for recovery of service through another round of DP is defined as ‘DP-II’ in the following discussions. To explore the new states for the recovery of train service from disruption, some operation rules are introduced to ensure valid states and stage transformation. They are explained in Section 6.4.3.

Fig. 6.8 depicts the set of states, (i.e., arrival times of trains at stations) on an optimal path to meet the specified operational requirements, through DP-I under a given traffic condition before disruption. Trains run along the line sequentially with the corresponding indices at successive stations shown in Fig. 6.9.

State (i.e., Arrival times of trains at stations (sec))

Stage	‘1’	‘2’	‘3’	‘4’	‘5’
0	455	495	521	516	497
1	575	615	641	636	617
2	695	735	761	756	737
3	815	855	881	876	857

Time delay is introduced to the train ‘*i*’ at station ‘3’ in stage 2

Disrupted state

Great care is needed to derive the arrival time of trains in the initial state for recovery

Fig. 6.8 Introduction of delay to one train in stage 2

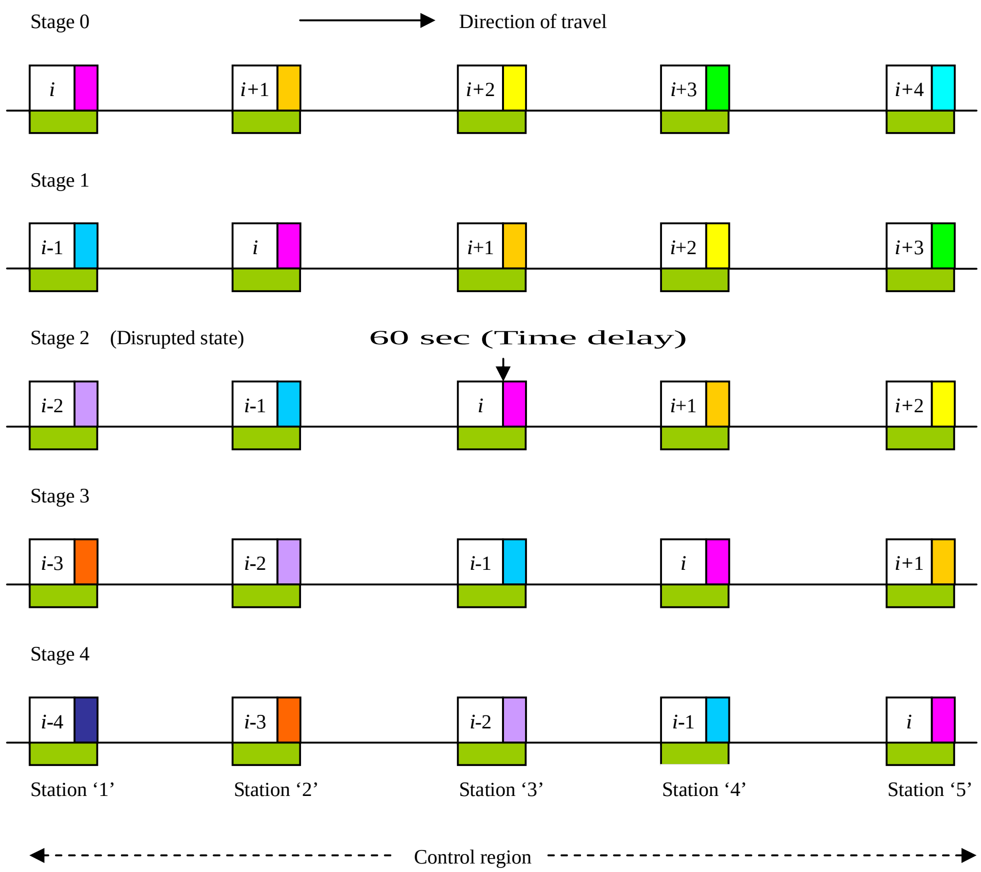


Fig. 6.9 Event based traffic flow conditions

In stage 2, if the traffic condition is disturbed, the train '*i*' waits at station '3' for an additional time duration of 60 sec. The given state [695, 735, 761, 756, 737] in stage 2 is described as the 'disrupted (disturbed) state'. The two states, [815, 855, 881, 876, 857] and [935, 975, 1001, 996, 977] for trains at successive stations before disruption, should have inevitably distorted in the later stages 3 and 4, because of the delay of train '*i*' at station '3' in the disrupted state.

In the event-based traffic flow model, an event represents the arrival times of trains at successive stations. Each train has its own corresponding arrival time at stations, and they are different from each other. The passage of time is irregular in the model and the updates of the train movement are not carried out synchronously. Therefore, great care in updating the event from one to another is needed, because the states before disruption in stage 3 and 4 through DP-I may be invalid from the viewpoint of signalling or other constraints, when an additional time delay for train '*i*' is introduced in the disrupted state. A new state evolved from the disrupted state is defined as the 'initial state for recovery'. This new state can be achieved through some specific rules regarding the adjustments to the arrival times of trains, as discussed in Section 6.4.3, and it is the initial state for DP-II to attain optimal control action to recover the service.

Because of the irregular time passage of train operation (i.e. the arrival time of train at stations are not identical) in the traffic flow model, the arrival time of the disturbed train '*i*' (i.e., 761 sec) at station '3' in the disrupted state [695, 735, 761, 756, 737] is taken as the reference to check the sequence of train operation. With reference to Fig. 6.10, when the arrival time of the train '*i*-1' at station '4' is smaller than that of the disturbed train '*i*' at station '3' in the disrupted state, the arrival time of the train '*i*-1' at station '5' (i.e., one element in the initial state for recovery) is the same as attained in stage 3 through DP-I. The reason is that the inter-station run of train '*i*-1' between stations '4' and '5' was determined by DP-I prior to the disruption. On the other hand, if the arrival time of train '*i*-1' at station '4' is later than the time of disruption, its subsequent inter-station run should be determined by DP-II.

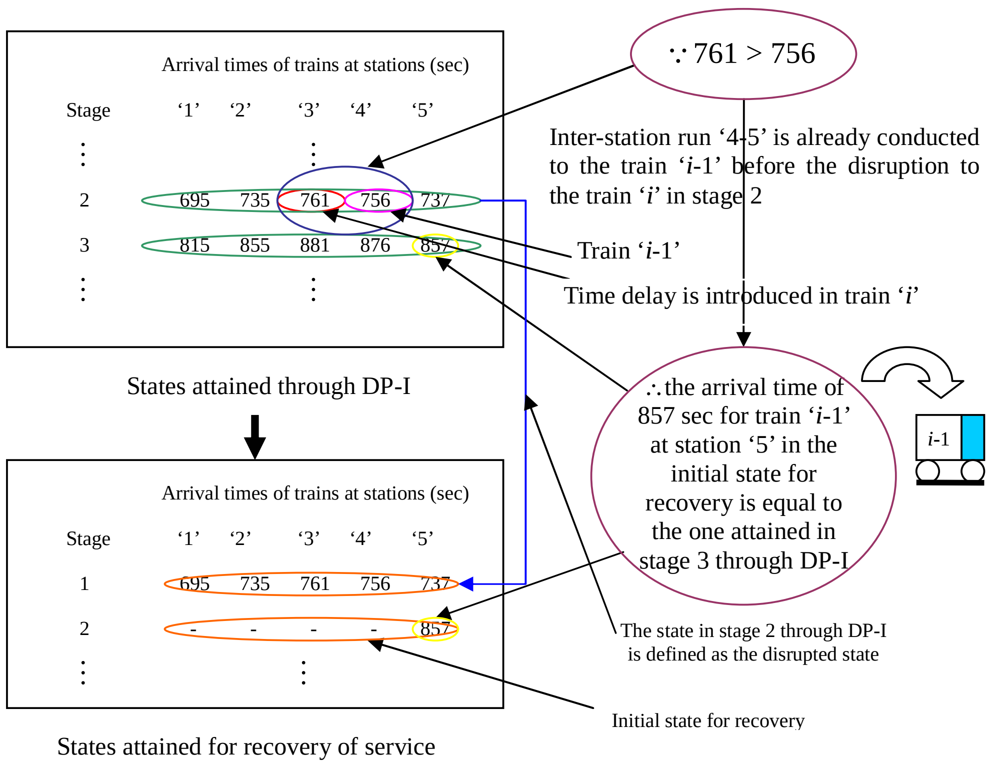


Fig. 6.10 Time reference to check the sequence of train operation

Adjustments to the arrival times of trains at stations in the initial state for recovery can be classified into two categories: modifications to the arrival times for the trains (i.e., '*i*+1' and '*i*+2') following the 'disturbed' train '*i*', and the trains (i.e., '*i*-1' and '*i*-2') ahead the 'disturbed' train '*i*'. The adjustments made to the arrival times for the trains ahead at stations in the initial state for recovery are the same as mentioned in the last paragraph. However, the 'disturbed' train '*i*' usually carries the impact of disruption to the trains following it, where schedules are more distorted than those of the trains ahead because of signalling constraints. It is therefore necessary to adjust the arrival times of following trains at stations, and details are given in Section 6.4.3. A special term for the adjustment of the schedules of following trains, 'inductive time delay', is introduced. Further, the approach adopted to achieve a valid state with arrival time for trains in the initial state for recovery will be explained through three examples with different disturbed traffic conditions in terms of additional time delays to train(s) at station(s) in the subsequent sections.

6.4.2 Inductive time delay

Under normal traffic conditions, trains run along the line sequentially with the given headway time separation to serve passengers optimally. To meet the passenger flow which varies throughout a day, changes in headway are inevitably needed to maintain the train service. However, there is a limit to the reduction of headway because of signalling constraints and the maximum headway reduction (*MHR*) is defined as:

$$MHR = Schedule\ headway - Minimum\ headway\,(HD_{\mathrm{min}}) \qquad (6.21)$$

To determine the safe operation for the trains at stations behind the disturbed train after the disruption, a term 'inductive time delay (*ITD*)' is introduced. The time delay induced to the trains behind at stations imposed by the actual time disruption to the preceding train is defined as the inductive time delay. With the given *MHR* on trains, the *ITD* for train behind, 'i+n', at stations 'x-n' (i.e., $ITD_{x\text{-}n}^{i+n}$), imposed by the time of disruption (i.e., $ATD_x^{i}\,|_k$) to the train 'i' at station 'x' in the disrupted state in stage k, can be deduced by the following equation.

$$ITD_{x-n}^{i+n}\,|_k = ATD_x^{i}\,|_k - (MHR \times n) \qquad n>0 \qquad (6.22)$$

Eqn. 6.22 shows that $ITD_{x-n}^{i+n}\,|_k = ATD_x^{i}\,|_k$ when n = 0. The inductive time delay of train 'i+n' at station 'x-n' behind the disturbed train 'i' at station 'n' in the disrupted state in stage k can be calculated with this equation for n>0. When $ITD_{x-n}^{i+n}\,|_k \geq 0$, the arrival time for the train 'i+n+1' behind at station 'x-n' in the initial state for recovery in stage k+1 (i.e., $AT_{x\text{-}n}^{i+n+1}|_{k+1}$) is devised because of the signalling constraints. Given $ITD_{x-n}^{i+n}|_k \geq 0$, $AT_{x-n}^{i+n+1}|_{k+1}$ can be determined by the following equation:

$$AT_{x-n}^{i+n+1}|_{k+1} = AT_{x-n}^{i+n}|_k + ITD_{x-n}^{i+n}|_k + HD_{\min} \tag{6.23}$$

Nevertheless, there is no inductive time delay imposed on the trains behind when $ITD_{x-n}^{i+n}|_k < 0$, and no disruption is thus caused to the trains behind. In addition, with $ITD_{x-n}^{i+n}|_k < 0$, the approach used to obtain the arrival time for the train '*i*+*n*+1' at station '*x-n*' in the initial state for recovery is similar to that for the train '*i-n*' ahead of the disturbed train '*i*', as described in Section 6.4.1.

6.4.3 Rules of the adjustments on the arrival times for trains

To determine the valid initial state for DP-II to attain optimal control action for recovery from disruption, an example is hereby given:

An unexpected delay is introduced to the train '*i*' at station '*x*' and another delay is given to train '*i*+*m*' at station '*x-m*' in the disrupted state in stage *k*. To explore the appropriate arrival time for each train in the initial state for recovery of service from the disrupted state, the following rules are applied.

Rule 1: Adjustment to the arrival time of train, '*i-p*', at station '*x*+*p*+1', in the initial state for recovery (i.e., stage *k*+1). Train '*i-p*' at station '*x*+*p*' is in front of the disturbed train '*i*' at station '*x*' in the disrupted state in stage *k*.

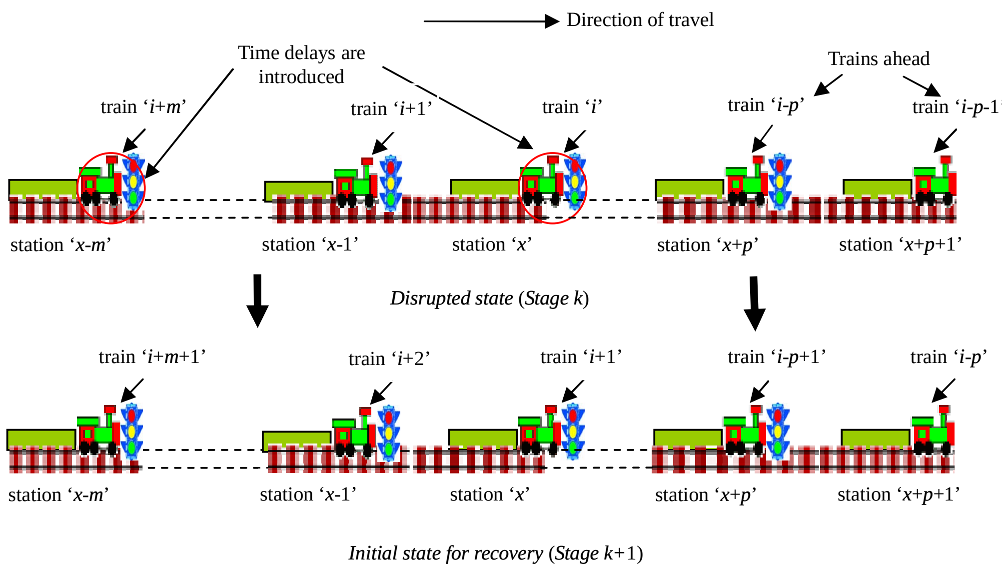


Fig. 6.11 Adjustment to the arrival time of train '*i-p*' in the initial state for recovery

If the arrival time of train '*i-p*' at station '*x*+*p*' is smaller than that of the disturbed train '*i*' at station '*x*' in the disrupted state in stage *k* (i.e., $AT_{x+p}^{\ \ i-p}|_k < AT_x^{\ i}|_k$), $AT_{x+p+1}^{\ \ i-p}|_{k+1}$ is equal to the given arrival time for train '*i-p*' at station '*x*+*p*+1' as derived by DP-I. It is because the train '*i-p*' has already started its inter-station train control from station '*x*+*p*' before the disruption, train '*i-p*' should carry on with the on-going control action and arrive at station '*x*+*p*+1'. However, if $AT_{x+p}^{\ \ i-p}|_k \geq AT_x^{\ i}|_k$, $AT_{x+p+1}^{\ \ i-p}|_{k+1}$ (i.e., $AT_{x+p+1}^{\ \ i-p}|_{k+1}$) is to be devised by DP-II. The above conditions are summarised as follows:

a. When $AT_{x+p}^{\ \ i-p}|_k < AT_x^{\ i}|_k$, $AT_{x+p}^{\ \ i-p}|_k \xrightarrow{DP-I} AT_{x+p+1}^{\ \ i-p}|_{k+1}$

b. When $AT_{x+p}^{\ \ i-p}|_k \geq AT_x^{\ i}|_k$, $AT_{x+p}^{\ \ i-p}|_k \xrightarrow{DP-II} AT_{x+p+1}^{\ \ i-p}|_{k+1}$

Rule 2: Adjustment to the arrival time of disturbed train '*i*' at station '*x*+1' in the initial state for recovery in stage *k*+1.

With the time delay to train 'i' at station 'x' in the disrupted state, the arrival time of disturbed train 'i' at station 'x+1' in the initial state for recovery in stage k+1 (i.e., $AT_{x+1}^{\ i}\ |_{k+1}$) is to be devised by DP-II.

$$\text{i.e.,}\quad AT_x^{\ i}\ |_k \xrightarrow{DP-II} AT_{x+1}^{\ i}\ |_{k+1}$$

Rule 3: Adjustment to the arrival time of train '$i+r$' at station 'x-r+1' in the initial state for recovery in stage k+1. Train '$i+r$' is between the two disturbed trains 'i' and '$i+m$' in the disrupted state in stage k.

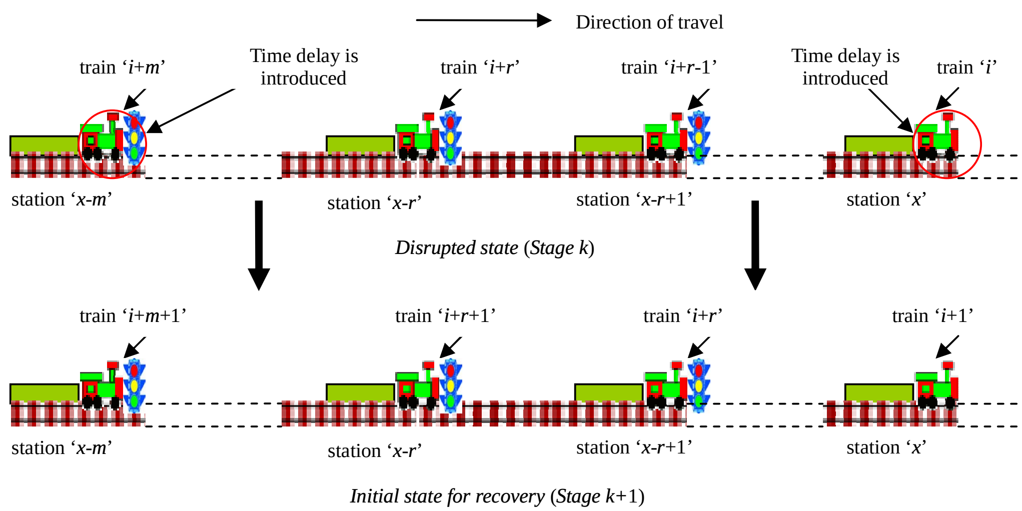


Fig. 6.12 Adjustment to the arrival time of train '$i+r$' between the two disturbed trains 'i' and '$i+m$' in the initial state for recovery

For train '$i+r$' at station 'x-r' which is between the two disturbed trains 'i' and '$i+m$' in the disrupted state, the disruption to the train 'i' may lead to inductive time delay on the train '$i+r$-1' at station 'x-r+1' (i.e., $ITD_{x-r+1}^{\ i+r-1}|_k$) in stage k. $ITD_{x-r+1}^{\ i+r-1}|_k$ is given by Eqn. 6.22. The arrival time of train '$i+r$' at station 'x-r+1' in the initial state for recovery in stage k+1 can be determined with the following two conditions:

a. When $ITD_{x-r+1}^{\ i+r-1}|_k \geq 0$, there are signalling constraints for the train 'i+r' in its subsequent run between stations 'i+r' and 'i+r+1'. The arrival time of train 'i+r' at station 'x-r+1' in stage k+1 is given by Eqn. 6.23:

$$AT_{x-r+1}^{\ i+r}\ |_{k+1} = AT_{x-r+1}^{\ i+r-1}\ |_k + ITD_{x-r+1}^{\ i+r-1}\ |_k + HD_{\min}$$

b. When $ITD_{x\text{-}r+1}^{\ i+r\text{-}1}|_k < 0$, it implies no signalling constraints are imposed on the train 'i+r' in the disrupted state in stage k, and

I. if the arrival time of train 'i+r' at station 'x-r' is larger than or equal to that of the train 'i' at station 'x' in the disrupted state, the arrival time of train 'i+r' at station 'x-r+1' in the initial state for recovery is to be devised by DP-II.

i.e., $AT_{x-r}^{\ i+r}\ |_k \geq AT_x^{\ i}\ |_k \xrightarrow{DP-II} AT_{x-r+1}^{\ i+r}\ |_{k+1}$

II. if the arrival time of train 'i+r' at station 'x-r' is smaller than that of the train 'i' at station 'x' in the disrupted state, the arrival time of train 'i+r' at station 'x-r+1' in the initial state for recovery is the same as derived by DP-I.

i.e., $AT_{x-r}^{\ i+r}\ |_k \geq AT_x^{\ i}\ |_k \xrightarrow{DP-I} AT_{x-r+1}^{\ i+r}\ |_{k+1}$

Rule 4: Adjustment to the arrival time of disturbed train 'i+m' at station 'x-m+1' in the initial state for recovery in stage k+1.

a. When the inductive time delay in train 'i+m' at station 'x-m' (i.e., $ITD_{x\text{-}m}^{\ i+m}|_k$) imposed by the actual time delay to the preceding train

'i' at station 'x' (i.e., $ATD_x^{\,i}|_k$) is larger than the actual time delay in train 'i+m' at station 'x-m' (i.e., $ATD_{x-m}^{\,i+m}|_k$) in the disrupted state in stage k, the arrival time of train 'i+m' at station 'x-m+1' in the initial state for recovery in stage k+1 (i.e., $AT_{x-m+1}^{\,i+m}|_{k+1}$) is calculated by Eqn. 6.23. The condition is similar to ***Rule*** 3a.

b. When $ITD_{x\text{-}m}^{\,i+m}|_k$ is smaller than the $ATD_{x-m}^{\,i+m}|_k$ in the disrupted state, the disruption to train 'i' at station 'x' imposes no effect on the trains 'i+m', 'i+m+1', …'i+m+n' at stations 'x-m', 'x-m-1', … 'x-m-n' respectively. With the introduction of $ATD_{x-m}^{\,i+m}|_k$ in the disrupted state, $AT_{x-m+1}^{\,i+m}|_{k+1}$ in the initial state for recovery is to be devised by DP-II.

$$\text{i.e., } \left(ITD_{x-m}^{\,i+m}|_k\right) < \left(ATD_{x-m}^{\,i+m}|_k\right) \xrightarrow{DP-II} AT_{x-m+1}^{\,i+m}|_{k+1}$$

Rule 5: Adjustment to the arrival time of train 'i+t' at station 'x-t+1' in the initial state for recovery in stage k+1. Train 'i+t' at station 'x-t' follows the disturbed train 'i+m' at station 'x-m' in the disrupted state in stage k.

Since there are two actual time delays introduced to trains 'i' and 'i+m' at stations 'x' and 'x-m' in the disrupted state respectively, the two corresponding inductive time delays on the train 'i+t-1' at station 'x-t+1' in stage k are computed first. These two inductive time delays are then compared with each other and the larger one is used to determine the arrival time of train 'i+t' at station 'x-t+1' in the initial state for recovery in stage k+1 with Eqn. 6.23.

Rules 3 and *4* are not applicable to the adjustment of the arrival times of trains behind in the initial state for recovery, when there is only one delayed train. To further explain the service recovery from disruption with the stated rules, three examples of different disturbed traffic conditions are given in the next sections.

6.4.4 One delay to one train

The following example illustrates the approach used to obtain the arrival time of trains at successive stations in the initial state for recovery in stage k+1, when an actual time delay of 60 sec is introduced to train 'i' at station '3' in the disrupted state in stage k=2 , as shown in Fig. 6.10. Headway and dwell-time at stations are set at 120 and 25 sec respectively. A maximum of 10 % reduction on headway (i.e., 12 sec) is allowed. The arrival time of trains at successive stations 1 through 5 are [695, 735, 761, 756 and 737] sec in the disrupted state and [815, 855, 881, 876 and 857] sec in the initial state for recovery respectively. When the train 'i' stops at station '3' with a time delay of 60 sec for any reason in the disrupted state, the corresponding inductive time delay for trains behind, 'i+1' and 'i+2', at stations '2' and '1' can be calculated as follows:

Actual time delay of train 'i' at station '3', $ATD_3^i|_2 = 60$ sec

With Eqn. 6.22, $ITD_{(3-1)}^{i+1}|_k = ITD_2^{i+1}|_2 = 60 - 12\times 1 = 48$ sec

$ITD_{(3-2)}^{i+2}|_k = ITD_1^{i+2}|_2 = 60 - 12\times 2 = 36$ sec

With the given time delay of 60 sec, the train 'i' departs from station '3' at 846 sec (i.e., 761+25+60) in the disrupted state, and the arrival time for the train at station '4' in the initial state for recovery can be obtained with DP-II (i.e., *Rule 2*). However, the arrival time for the train '*i*+1' behind at station '3' has been determined as 881 sec in stage 3 through DP-I. The headway becomes 60 sec (i.e., 881-846+25), which is much less than the minimum allowable value of 108 sec (i.e., 120-12). The arrival time, 881 sec, of the train '*i*+1' at station '3' obtained from DP-I is therefore not valid, and it is adjusted to 929 sec (i.e., 761+60+108) with Eqn. 6.23 in the initial state for recovery (i.e., *Rule 5*). Similarly, the arrival time of the following trains, '*i*+2' and '*i*+3', at stations are revised in the initial state for recovery, according to the signalling constraints (i.e., minimum headway), and they become:

Arrival time of train '*i*+2' at station '2', $AT_2^{i+2}|_3$

$= AT_2^{i+1}|_2 + ITD_2^{i+1}|_2 + HD_{\min} = 735 + 48 + 108 = 891$ sec

Arrival time of train '*i*+3' at station '1', $AT_1^{i+3}|_3$

$= AT_1^{i+2}|_2 + ITD_1^{i+2}|_2 + HD_{\min} = 695 + 36 + 108 = 839$ sec

Further, the arrival time of the train ahead, '*i*-1', (i.e., 857 sec) at station '5' in stage 3, obtained through DP-I is used as that in the initial state for recovery according to *Rule 1a*, because the arrival time of train '*i*-1' at station '4' is smaller (i.e., 756 < 761 sec) than that of the disturbed train '*i*' at station '3' in the disrupted state.

An initial state for recovery [839, 891, 929, ♣, 857] is thus formed. The symbol '♣' denotes the arrival time of train '*i*' at station '4', which is to be devised by DP-II.

6.4.5 Extent of time delay

The following example is to demonstrate the effect of different magnitudes of time delay, introduced to a train at a station, on the arrival times of trains at stations in the initial state for recovery. The traffic conditions remain identical, as given in Section 6.4.4 but the time delay to train '*i*' at station '3', $ATD_3^i|_2$, is set to 20 sec in the disrupted state in stage 2. The two corresponding inductive time delays for the trains behind, '*i*+1' and '*i*+2', at stations '2' and '1', become:

Actual time delay of train at station '3', $ATD_3^i|_2$, = 20 sec

$ITD_{(3\text{-}1)}^{i+1}|_k = ITD_2^{i+1}|_2 = 20 - 12\times(3\text{-}2) = 8$ sec

$ITD_{(3\text{-}2)}^{i+2}|_k = ITD_1^{i+2}|_2 = 20 - 12\times(3\text{-}1) = -4$ sec

With $ITD_1^{i+2}|_2 < 0$, the preceding train '*i*' imposes no time delay to the train '*i*+2' at station '1' in the disrupted state. The arrival time of the train '*i*+3' at station '1' in the initial state for recovery remains the same (i.e., 815 sec) as that attained by DP-I in stage 3, $AT_1^{i+3}|_3^{DP\text{-}I}$. Similarly, the arrival time of trains '*i*+2' and '*i*+1' at stations '2' and '3' in the initial state for recovery are defined by Eqn. 6.23 (i.e., *Rule 5*):

Arrival time of train '*i*+1' at station '3', $AT_3^{i+1}|_3$

$= AT_3^{i}|_2 + ATD_3^{i}|_2 + HD_{\min} = 761 + 20 + 108 = 889$ sec

Arrival time of train '*i*+2' at station '2', $AT_2^{i+2}|_3$

$= AT_2^{i+1}|_2 + ITD_2^{i+1}|_2 + HD_{\min} = 735 + 8 + 108 = 851$ sec

In addition, the arrival time for trains '*i*' and '*i*-1' at stations '4' and '5' in the initial state for recovery can be determined through *Rules 2* and *1* respectively. The initial state for recovery therefore becomes [815, 851, 889, ♣, 857].

6.4.6 Disruption to multiple trains

This example evaluates the impact of disruption to several trains to obtain the arrival times of trains at stations in the initial state for recovery. The traffic conditions remain the same as shown in Fig. 6.13, but two time delays of 50 sec and 60 sec are introduced to the trains ‘*i*+1’ and ‘*i*-1’ at stations ‘2’ and ‘4’ in the disrupted state in stage 2 respectively.

Arrival times of trains at stations (sec)

Stage	‘1’	‘2’	‘3’	‘4’	‘5’
0	455	495	521	516	497
1	575	615	641	636	617
2	695	735	761	756	737
3	815	855	881	876	857
4	935	975	1001	996	977

Two time delays of 50 and 60 sec are introduced to the trains ‘*i*+1’ and ‘*i*-1’ at station ‘2’ and ‘4’ respectively in stage 2

Disrupted state

The state attained through DP-I in stage 3 is used as the reference to determine the arrival times for trains at stations in the initial state for recovery

Fig. 6.13 Introduction of two corresponding time delays to trains in stage 2

With the time delay of 60 sec on train ‘*i*-1’ at station ‘4’, the inductive time delay to the trains behind (i.e., ‘*i*’, ‘*i*+1’ and ‘*i*+2’) at stations ‘3’, ‘2’ and ‘1’ are calculated as,

Actual time delay of train ‘*i*-1’ at station ‘4’, $ATD_4^{i\text{-}1}|_2$, = 60 sec

$$ITD_{(4\text{-}1)}^{i}|_k = ITD_3^{i}|_2 = 60 - 12\times 1 = 48 \text{ sec}$$

$$ITD_{(4\text{-}2)}^{i+1}|_k = ITD_2^{i+1}|_2 = 60 - 12\times 2 = 36 \text{ sec}$$

$$ITD_{(4\text{-}3)}^{i+2}|_k = ITD_1^{i+2}|_2 = 60 - 12\times 3 = 24 \text{ sec}$$

The arrival time of trains at stations in the initial state for recovery are deduced by *Rule 3a*:

$$AT_4^{i}|_3 = AT_4^{i-1}|_2 + ATD_4^{i-1}|_2 + HD_{\min} = 756 + 60 + 108 = 924 \text{ sec}$$

$$AT_3^{i+1}|_3 = AT_3^{i}|_2 + ITD_3^{i}|_2 + HD_{\min} = 761 + 48 + 108 = 917 \text{ sec}$$

$$AT_2^{i+2}|_3 = AT_2^{i+1}|_2 + ITD_2^{i+1}|_2 + HD_{\min} = 735 + 36 + 108 = 879 \text{ sec}$$

$$AT_1^{i+3}|_3 = AT_1^{i+2}|_2 + ITD_1^{i+2}|_2 + HD_{\min} = 695 + 24 + 108 = 827 \text{ sec}$$

The arrival time of train '*i*-1' at station '5' in the initial state for recovery, $AT_5^{i\text{-}1}|_3$, has to be achieved through DP-II with *Rule* 2. However, the arrival times of trains '*i*+1', '*i*+2' and '*i*+3' at stations '3', '2' and '1' need further adjustments in the initial state for recovery because of the time delay of 50 sec to the train '*i*+1' at station '2' in the disrupted state. The inductive time delay at station '1' is then updated with 50 sec time delay in train '*i*+1', which is denoted by '*'.

Actual time delay of train '*i*+1' at station '2', $ATD_2^{i+1}|_2$, = 50 sec

$ITD_{(2\text{-}1)}^{i+2}|_k^{\,*} = ITD_1^{i+2}|_2^{\,*} = 50 - 12\times1 = 38$ sec

Since the $ITD_2^{i+1}|_2$ (i.e., 36 sec) for train '*i*+1' induced by the time delay to the train '*i*-1' at station '4' is less than the actual time delay to train '*i*+1' at station '2' (i.e., 50 sec) in the disrupted state, the arrival time of train '*i*+1' at station '3' in the initial state for recovery is to be obtained through DP-II with *Rule 4b*. The arrival times of trains '*i*+3' and '*i*+2' at stations '1' and '2' in the initial state for recovery are revised according to *Rule 5*;

$$AT_2^{i+2}|_3 = AT_2^{i+1}|_2 + ATD_2^{i+1}|_2 + HD_{\min} = 735 + 50 + 108 = 893 \text{ sec}$$

$$AT_1^{i+3}|_3 = AT_1^{i+2}|_2 + ITD_1^{i+2}|_2^{*} + HD_{\min} = 695 + 38 + 108 = 841 \text{ sec}$$

The initial state for recovery thus becomes [841, 893, ♣, 924, ♣].

Nevertheless, the arrival times of trains '*i*+1', '*i*+2' and '*i*+3' at stations '3', '2' and '1' in the initial state for recovery need no further revision, if $ITD_2^{i+1}|_2$ for train '*i*+1' induced by the disruption to the train '*i*-1' ahead is larger than the actual time delay to the train '*i*+1' at station '2' (i.e., *Rule 4a*). The initial state [827, 879, 917, 924, ♣] is thus formed.

In general, adjustments to the arrival time for trains at stations in the initial state for recovery are in accordance with: i) the magnitude of time delay(s) to train(s), and ii) its/their relative position(s). With reference to Fig. 6.13, to further explain the adjustments to the arrival times for trains at stations with different magnitude of time delay to trains, two time delays to trains '*i*-1' and '*i*+1' at stations '4' and '2', 10 and 30 sec, are given in the disrupted state respectively. With the time delay of 10 sec to the train '*i*-1' at station '4', the corresponding inductive time delays to the trains behind (i.e., '*i*', '*i*+1' and '*i*+2') at station '3', '2' and '1' are all negative (i.e., -2, -14 and -26). However, the inductive time delay for the train '*i*+2' at station '1' is revised to 18 sec in the disrupted state, when the actual time delay of 30 sec to train '*i*+1' at station '2' is introduced. The arrival times for trains '*i*+2' and '*i*+3' at stations '2' and '1' in the initial state for recovery then become:

$$AT_2^{i+2}\,|_3 = AT_2^{i+1}\,|_2 + ATD_2^{i+1}\,|_2 + HD_{\min} = 735 + 30 + 108 = 873 \text{ sec}$$

$$AT_1^{i+3}\,|_3 = AT_1^{i+2}\,|_2 + ITD_1^{i+2}\,|_2 + HD_{\min} = 695 + 18 + 108 = 821 \text{ sec}$$

The arrival times for the disturbed trains '*i*+1' and '*i*-1' at stations '3' and '5' in the initial state for recovery are both deduced through DP-II (i.e., *Rule 2* and *Rule 4b*). Further, since $ITD_3^i|_2$ for train '*i*' at station '3' is negative and the arrival time of train '*i*-1' at station '4' is smaller than that of train '*i*' at station '3' (i.e., 756<761) in the disrupted state, the arrival time of train '*i*' at station '4' in the initial state for recovery is the same as attained by DP-I (i.e., Rule *3b*-II). As a result, the state becomes [821, 873, ♣, ♣, ♣].

6.5 Coordination in multiple regions

With the *HTRC*, a railway line is divided into a number of regions to carry out the multi-train control. The section of track between two feeder substations is defined as a control region, which normally covers a number of passenger stations in DC metro systems. Usually, a line is divided into more than two regions and each of them regulates the train operation under given traffic conditions and operational requirements independently. From the viewpoint of application, a control window/cycle for the regional controller to update the train operation decisions to meet the operational requirements is the time taken for the train movement in successive inter-station runs within a region, as shown in Fig. 6.14. In other words, in DP, the regional control cycle is the number of stage transformations (i.e., events). A larger number of inter-station runs implies a larger number of stage transformations. However, a longer computation time and larger memory demand is needed to obtain the optimal train operation with a higher number of stage transformations in DP.

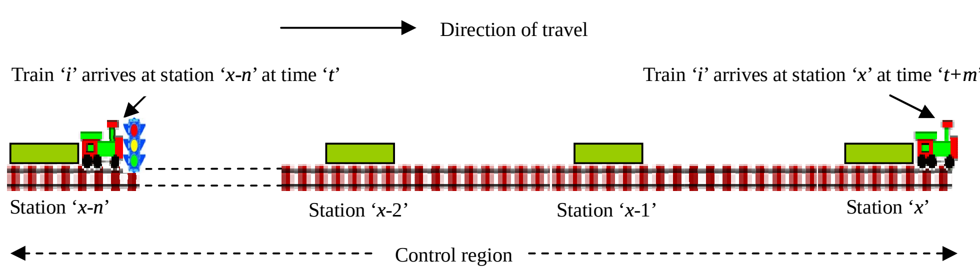


Fig. 6.14 Time interval to update the train instructions

In case of recovery of train service from disruption or adjusting the service with given operational requirements, an irregular headway may occur in a region and it is likely to be carried through to the following region via the 'overlapping' station, as shown in Fig. 6.15. Because of the sequential nature of train operation, train coordination in successive regions is thereby important.

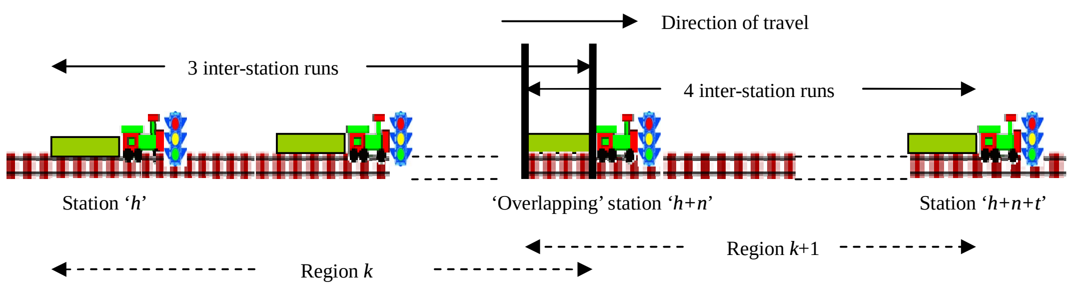


Fig. 6.15 Trains coordination in multiple regions

Further, it is worth noting that the controller's flexibility in adjusting the train service depends upon not only the number of steps in the control variable (i.e., a larger number of control steps implies more choices for the adjustment of train operation), but also the magnitude and location of the time disruption to train(s) at station(s). To obtain effective train control for recovery, the decision on 1) the number of steps in the control variable, e.g., two steps 5, 10 sec, and 2) the control direction according to the nominal schedule (i.e., positive or negative adjustment of the control steps), like ±5 and ± 10 sec, with respect to the disruption(s) are essential. In general, a larger (time) disruption(s) to train(s) usually requires a smaller resolution on the control steps within a given control space (e.g., 10% adjustment of run-time with respect to the nominal schedule).

The number of stations is usually different from one region to another. With the two possibly different numbers of inter-station runs in two successive regions, the numbers of stage transformations in DP in those regions are therefore not necessarily identical. For instance, with 3 inter-station runs in region k and 4 in region k+1 as depicted in Fig. 6.16, there are three optimal arrival-time elements at the 'overlapping' station (i.e., AT_3^6, AT_3^7 and AT_3^8), under given operational constraints at the n^{th} control cycle in region k. However, the number of arrival-time elements required to successfully carry out the train control in region k+1 is 4 (i.e., AT_3^6, AT_3^7, AT_3^8 and AT_3^9). Hence, the arrival-time instructions for trains at the 'overlapping' station cannot be simply used by the controller in region k+1, because AT_3^9 is not available. To achieve a good and easier coordination of train operation across two regions with different inter-station runs, a common number of stage transformations in those regions is highly desirable. $Stage_{\min}|_{k,k+1}$ is the least common multiple (i.e., *L.C.M.*) of the two possible numbers of stage transformations required in control cycles in regions k and k+1 (i.e., the number of stage transformations required in DP is the number of inter-station runs in a region), and it is defined as:

$$Stage_{\min}|_{k,k+1} = L.C.M.(Stage_k\,,\, Stage_{k+1}) \tag{6.24}$$

where $Stage_k$ *and* $Stage_{k+1}$ are the numbers of stages of control cycles required in regions 'k' and 'k+1' respectively.

A larger number of stage transformations is not essential for train control when $Stage_k$ *and* $Stage_{k+1}$ are the same. It is because when the number of inter-station runs is identical in two regions, the arrival-time set at the 'overlapping' station given by the controller in region k can be carried forward to that in region k+1 directly.

Three elements are given at station ‘3’ in one control cycle in region k, but four elements are needed for carrying on the train control in region k+1

Arrival time at the overlapping station is not available to region k+1 in stage 4

Stage	Arrival time of trains at stations (sec)							
	Region k				Region k+1			
	0	1	2	**3***	4	5	6	7
Initial state 0	AT_0^{8}	AT_1^{7}	AT_2^{6}	AT_3^{5}	AT_4^{4}	AT_5^{3}	AT_6^{2}	AT_7^{1}
1	AT_0^{9}	AT_1^{8}	AT_2^{7}	AT_3^{6}	AT_4^{5}	AT_5^{4}	AT_6^{3}	AT_7^{2}
2	AT_0^{10}	AT_1^{9}	AT_2^{8}	AT_3^{7}	AT_4^{6}	AT_5^{5}	AT_6^{4}	AT_7^{3}
3	AT_0^{11}	AT_1^{10}	AT_2^{9}	AT_3^{8}	AT_4^{7}	AT_5^{6}	AT_6^{5}	AT_7^{4}
4	**?**	**?**	**?**	AT_3^{9}**	AT_4^{8}	AT_5^{7}	AT_6^{6}	AT_7^{5}

$AT_x^{\,i}$ is the arrival time of train ‘i’ at station ‘x’

* Station ‘3’ is the overlapping station in the two successive regions.

** Arrival time of train 9 at station ‘3’, AT_3^{9}, is not attained at the n^{th} control cycle in region k.

Fig. 6.16 Interruption of one arrival time exchange at the overlapping station with different numbers of inter-station runs in two successive regions

Fig. 6.17 is an example to illustrate an interchange of the arrival-time sets at the ‘overlapping’ station for train control between regions for recovery. One region consists of 5 stations (i.e., 0-4) whilst the other includes 4 stations (i.e., 4-7). Station ‘4’ is the ‘overlapping’ station, and trains pass station ‘4’ while approaching region 2 from region 1. Since there are four and three stage transformations in each control cycle of decision-making in region 1 and 2 respectively, $Stage_{\min}$ becomes,

$$\begin{aligned} Stage_{\min}\,|_{1,2} &= L.C.M.(Stage_1,\, Stage_2) \\ &= L.C.M.(4,\,3) \\ &= 12 \end{aligned}$$

Stage	Arrival time of trains at stations (sec)							
	Region 1 (12 stages with 3 cycles)				Overlap station	Region 2 (12 stages with 4 cycles)		
	0	1	2	3	4	5	6	7
0	AT_0^{8}	AT_1^{7}	AT_2^{6}	AT_3^{5}	AT_4^{4}	AT_5^{3}	AT_6^{2}	AT_7^{1}
1	AT_0^{9}	AT_1^{8}	AT_2^{7}	AT_3^{6}	AT_4^{5}	AT_5^{4}	AT_6^{3}	AT_7^{2}
2	AT_0^{10}	AT_1^{9}	AT_2^{8}	AT_3^{7}	AT_4^{6}	AT_5^{5}	AT_6^{4}	AT_7^{3}
3	AT_0^{11}	AT_1^{10}	AT_2^{9}	AT_3^{8}	AT_4^{7}	AT_5^{6}	AT_6^{5}	AT_7^{4}
4	AT_0^{12}	AT_1^{11}	AT_2^{10}	AT_3^{9}	AT_4^{8}	AT_5^{7}	AT_6^{6}	AT_7^{5}
5	AT_0^{13}	AT_1^{12}	AT_2^{11}	AT_3^{10}	AT_4^{9}	AT_5^{8}	AT_6^{7}	AT_7^{6}
6	AT_0^{14}	AT_1^{13}	AT_2^{12}	AT_3^{11}	AT_4^{10}	AT_5^{9}	AT_6^{8}	AT_7^{7}
7	AT_0^{15}	AT_1^{14}	AT_2^{13}	AT_3^{12}	AT_4^{11}	AT_5^{10}	AT_6^{9}	AT_7^{8}
8	AT_0^{16}	AT_1^{15}	AT_2^{14}	AT_3^{13}	AT_4^{12}	AT_5^{11}	AT_6^{10}	AT_7^{9}
9	AT_0^{17}	AT_1^{16}	AT_2^{15}	AT_3^{14}	AT_4^{13}	AT_5^{12}	AT_6^{11}	AT_7^{10}
10	AT_0^{18}	AT_1^{17}	AT_2^{16}	AT_3^{15}	AT_4^{14}	AT_5^{13}	AT_6^{12}	AT_7^{11}
11	AT_0^{19}	AT_1^{18}	AT_2^{17}	AT_3^{16}	AT_4^{15}	AT_5^{14}	AT_6^{13}	AT_7^{12}
12	AT_0^{20}	AT_1^{19}	AT_2^{18}	AT_3^{17}	AT_4^{16}	AT_5^{15}	AT_6^{14}	AT_7^{13}

$AT_x^{\,i}$ is the arrival time of train 'i' at station 'x'

Fig. 6.17 Transformation of arrival time for trains through the two successive regions

To ensure the continuity of train operation across regions by following the arrival-time sets for trains at the 'overlapping' station, the number of control cycles in region 1 is defined as:

$$Region\ 1 = \frac{12}{4} = 3\ cycles$$

Similarly, the number of control cycles in region 2 is calculated by:

$$Region\ 2 = \frac{12}{3} = 4\ cycles$$

With 3 cycles in region 1 and 4 cycles in region 2, the interchange of the arrival-time set at the 'overlapping' station across regions for train control is established.

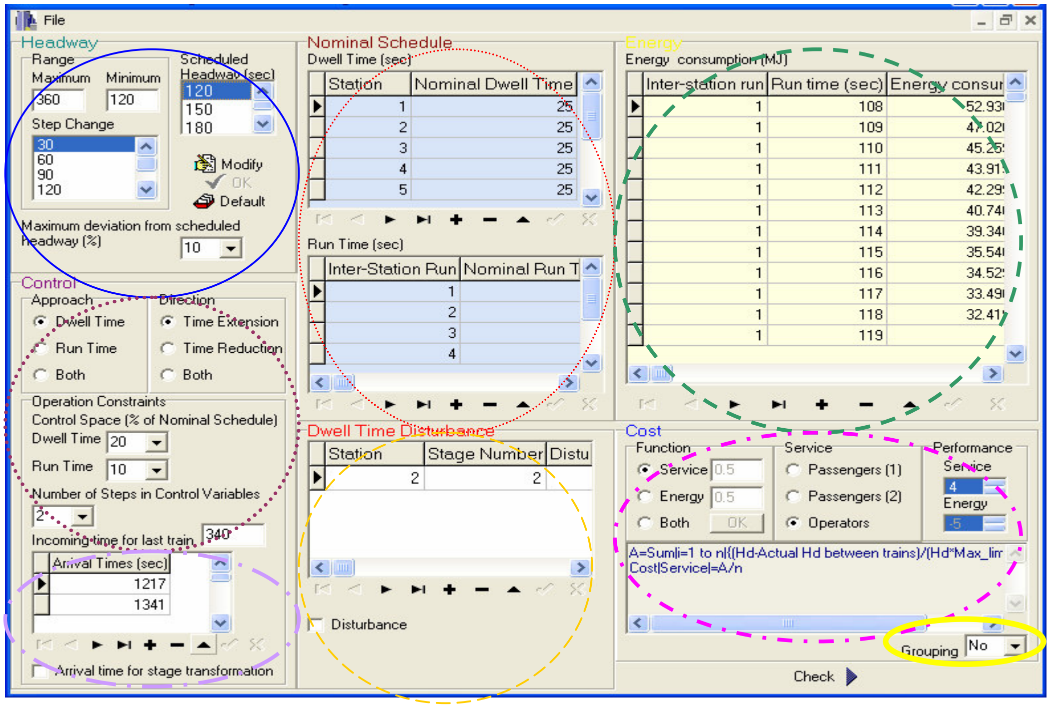
6.6 Software implementation

6.6.1 Interfaces

The *RTC* has been implemented in Borland C++ Builder [85,86], which not only provides an effective user interface, but also ensures fast calculation in the application of the event-based traffic flow model with C++ language. Fig. 6.18 shows the input interface of the *RTC*, a number of operational constraints and system requirements should be defined through the input interface prior to the simulation:

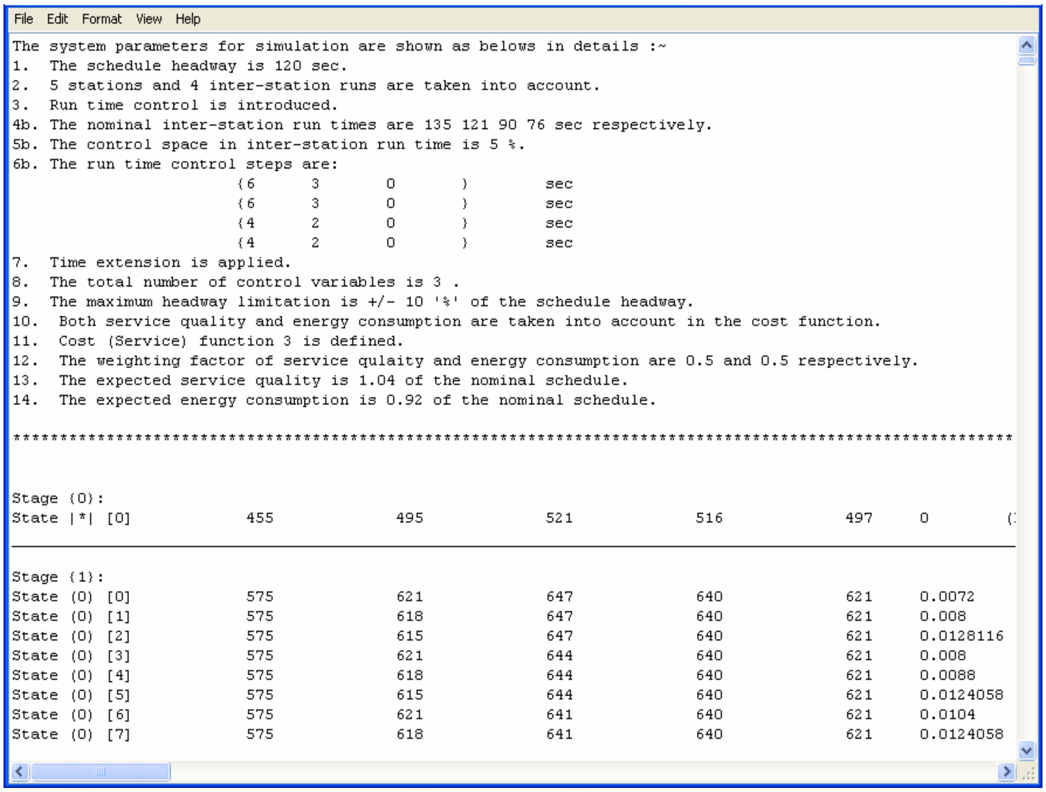
1. Headway pattern;
2. Nominal train schedule;
3. A 'lookup table' of different run-times and their corresponding energy consumption of trains in successive inter-station runs. Energy consumption of trains with different run-times in a run is obtained with a single train simulator;
4. Control method (i.e., run-time and/or dwell-time control) to meet the operational requirements and the availability of the solution space;
5. Number of steps in control variables;
6. Cost functions; and
7. Operational requirements in terms of service quality and energy consumption. The weightings on the service quality and energy consumption of train operation are defined according to the users' expectation.

Given the operational constraints and requirements, the state-space traffic flow model is established. Dwell-time and/or run-time for trains and hence the arrival time of trains at the successive stations in a particular region are calculated with DP, and the result are displayed on the output interface as illustrated in Fig. 6.19.



Line	Name	Functionality
	Headway	Operational headway of trains
	Control	Approach (i.e., dwell-time and/or run-time control) to be introduced in the event-based traffic flow model. The control space in dwell-time and run-time is also to be defined
	Nominal schedule	Nominal dwell-time of train at stations and run-times of trains in successive inter-station runs
	Energy	Energy consumption with the corresponding run-times of train. The possible range of run-time for train with respect to the nominal schedule is stated
	Cost	Determination of how well the chosen control actions lead to the desired operational requirements
	Dwell-time disturbance	Introduction of time disruption(s) to train(s) at station(s)
	Coordination in regions	Arrival times for trains at the multi-region station for the coordination of train operation in two successive regions for the recovery of service after the disruption
	Grouping	Operation with state grouping

Fig. 6.18 Input interface for regional train controller



```
File  Edit  Format  View  Help
The system parameters for simulation are shown as belows in details :~
1.  The schedule headway is 120 sec.
2.  5 stations and 4 inter-station runs are taken into account.
3.  Run time control is introduced.
4b. The nominal inter-station run times are 135 121 90 76 sec respectively.
5b. The control space in inter-station run time is 5 %.
6b. The run time control steps are:
                         {6      3       0       }         sec
                         {6      3       0       }         sec
                         {4      2       0       }         sec
                         {4      2       0       }         sec
7.  Time extension is applied.
8.  The total number of control variables is 3 .
9.  The maximum headway limitation is +/- 10 '%' of the schedule headway.
10.  Both service quality and energy consumption are taken into account in the cost function.
11.  Cost (Service) function 3 is defined.
12.  The weighting factor of service qulaity and energy consumption are 0.5 and 0.5 respectively.
13.  The expected service quality is 1.04 of the nominal schedule.
14.  The expected energy consumption is 0.92 of the nominal schedule.

*****************************************************************************************************************


Stage (0):
State |*| [0]           455             495             521             516             497     0         (:

Stage (1):
State (0) [0]           575             621             647             640             621     0.0072
State (0) [1]           575             618             647             640             621     0.008
State (0) [2]           575             615             647             640             621     0.0128116
State (0) [3]           575             621             644             640             621     0.008
State (0) [4]           575             618             644             640             621     0.0088
State (0) [5]           575             615             644             640             621     0.0124058
State (0) [6]           575             621             641             640             621     0.0104
State (0) [7]           575             618             641             640             621     0.0124058
```

Fig. 6.19 Output interface for regional train controller

6.6.2 Functional structure

In the event-based traffic flow model, a number of input parameters are required to conduct the traffic flow calculation. The functions of the corresponding input parameters in the model are given in Fig. 6.18. In general, input parameters are classified into two categories and they can be defined by the operators: i) operational constraints for the traffic flow calculation and ii) operational requirements. The operational constraints for trains and their corresponding functions in the traffic flow calculation are stated as follows:

1. Headway of trains – Ensure safe train operation.
2. Nominal dwell-time for trains at stations and run-times of trains in successive

inter-station runs – They are the two parameters in the train movement calculation for the stage transformation in the state-space traffic flow model.

3. Energy consumption with the corresponding run-times of trains in successive inter-station runs – Energy consumption with different run-times of train are necessary to determine how well the chosen control actions (i.e., dwell-time or run-time control) lead to the desired energy consumption of train operation.
4. Control method – Dwell-time, run-time and their combination, are used to adjust the train operation in the successive inter-station runs in the model.
5. Time disruption(s) to train(s) – Train(s) with the corresponding magnitude of time delay(s) at station(s) can be defined in the model. The states with the given arrival time of trains at stations for the recovery of service from disruption are provided in the model.

In addition, the objective functions, for service quality and energy consumption, as defined in Section 6.3.6, are provided. The operational requirement of trains can be easily defined by the operators.

Given the operational constraints and requirements, the controller first determines a state with the arrival times for trains at stations in the initial stage, according to the headway and nominal train schedule in successive inter-station runs. When the initial state is fixed, the complete state diagram is established with the given arrival times for trains at stations and the possible control actions and nominal schedule in the subsequent stages. The optimal solution with the corresponding cost and control actions in successive stage transformations is finally attained through DP. Details of the structure of the regional train controller are depicted in Fig. 6.20.

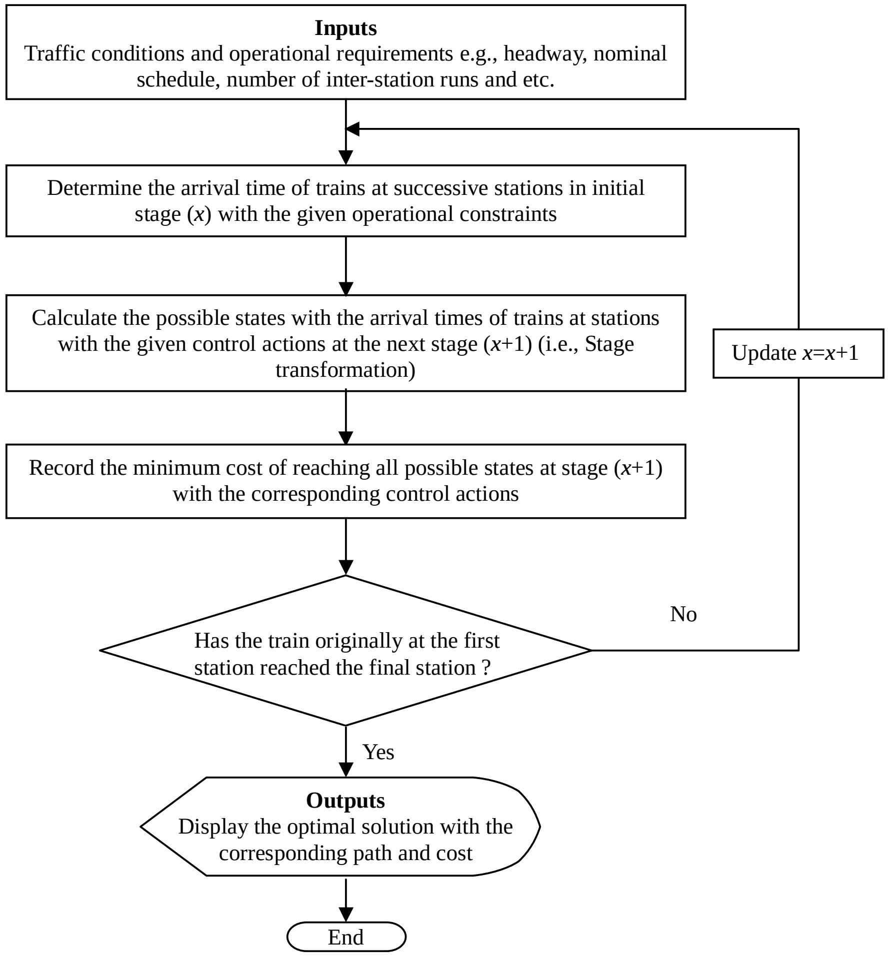


Fig. 6.20 Structure of the *RTC*

6.6.3 Data storage

The data storage requirement is the major concern in the traffic flow calculation. The number of states with given arrival times of trains at stations escalates significantly with the number of stations in a region and the number of possible control actions in the event-based model. A larger number of states generated implies a higher data storage requirement. To reduce the number of states in the model and hence the data storage requirement, state grouping has been adopted to combine the states, which have the same arrival times of trains at stations. Table

6.3 shows the advantage to be obtained by grouping with different numbers of control actions and inter-station runs. For example, when 3 inter-station runs and 4 steps of control variable are involved, the number of states evolved at the last stage is reduced by a ratio of 64 with grouping and the data storage memory of the overall traffic flow model is decreased from 26MB to 1.27MB. The advantage of state simplification will be further explained in Section 7.2.1.

6.7 Practical considerations

With *HTRC*, an advisory system of dwell-time and run-time control, *RTC*, can be integrated in the central control room to provide the necessary instructions for trains within a region. At the system level of control, the *CTC* determines the necessary service headway for *RTC*s in accordance with the timetable. With the given headway, a *RTC* then provides an appropriate set of dwell-times and run-times for trains and forwards the recommended control actions to the on-board *TBC*. In addition, the *RTC* devises the updated control actions on trains for service recovery when service disruption occurs. All control actions devised by means of the *RTC* are subject to supervision by the ATP to ensure safe train operation.

The *RTC* requires two sets of data to devise the necessary instructions for trains in the decision-making process. Some of them are operational-dependent, such as the number of steps in control variables; while the others are infrastructure-related, such as the number of passenger stations within a region. As there are usually a few regions along a line, exchange of arrival-times sets for trains at the 'overlapping' stations across regions is needed to link up train control through a number of *RTC*s.

Since the number of possible states (i.e., arrival-time sets) escalates with the number of stations and control actions in the traffic flow model, a large data storage and fast microprocessor platform are necessary to carry out the multi-train operation in this level of control. To enhance the availability and reliability of regional control, a duplicate and hot standby hardware and software are required to take over the train control immediately in case of any failure.

Chapter 7
Performance Analysis of Regional Train Controller

This chapter looks into studies on the feasibility and versatility of the regional train controller in carrying out dwell-time and run-time control within a region in *HTRC*, under various traffic conditions and operational requirements. With the given headway constraints imposed by the central train controller, *RTC* enables multi-train coordination to maintain the train service within its own operating region. To test the functions of the controller both for normal traffic conditions and for the recovery of service from disruption, numerous case studies are carried out and described in this chapter.

7.1 Simulation setup

With the aid of the event-based traffic flow model, a number of studies were carried out to explore the controller's performance, in terms of memory requirement, computational time and optimality solution. The subjects of the studies include:

1. Advantage of state grouping for memory storage and simulation time.
2. Weighting factors on service and energy of the cost function to train operation.
3. Comparison of dwell-time, run-time and their combinations on train control.
4. Impact of track layouts on train control.
5. Impact of time delay on train control.
6. Recovery of train service from time disruption at several stations.
7. Recovery of train service from time disruption in coordination between regions.
8. Different disruption levels at two stations.

The nominal headway on trains is set at 120 sec and nominal schedule of dwell-times for trains at stations is 25sec. Other operational constraints and requirements are shown against the experiments. The simulations were run on IBM-compatible PC with Pentium III CPU in all tests.

7.2 Case studies

7.2.1 State simplification

This study reveals the possible state simplification at stages in the state-space traffic flow model through grouping. The traffic conditions and operational requirements are given in Table 7.1.

Table 7.1 – Traffic conditions and operational requirements

Traffic conditions			
Number of inter-station runs	3		
Allowable headway deviation	± 15%		
Nominal inter-station run-times	Run 0-1	Run 1-2	Run 2-3
	135 sec	121 sec	90 sec
Control and operational requirements			
Control method	Mixed (i.e., Dwell-time and run-time)		
Control direction	Time extension		
Control space*	Run-time: +10%	Dwell-time: +20%	
Number of steps in control variable	Dwell-time		
	3, 2, 1, 0 sec		
	Run 0-1	Run 1-2	Run 2-3
	12, 8, 4, 0 sec	12, 8, 4, 0 sec	9, 6, 3, 0 sec
Operational requirements	Service: 10% extension		
Cost function	Service regularity		

*Nominal schedule is used as a reference to regulate the service.

Results:

Table 7.2 – Number of states at stages

Case I: Operation with complete state-space diagram (i.e., without grouping)				
	Stage 0	Stage 1	Stage 2	Stage 3
Number of states	1	64	4096	262144
Case II: Operation with dynamic programming (i.e., state grouping)				
	Stage 0	Stage 1	Stage 2	Stage 3
Number of states	1	64	1024	4096

Table 7.3 – Optimal cost and computational demand

	Case I	Case II
Cost	0.1269	0.1269
Computation time (sec)	35.6	0.73
Physical memory storage (MB)	26	1.27

Table 7.4 – Dwell-time and run-time extensions with the corresponding optimal paths

Stage transformation	Dwell-time (sec)*			
	Station ‘0’	Station ‘1’	Station ‘2’	Station ‘3’
0→1	2	2	3	3
1→2	3	2	3	3
2→3	3	3	3	3
	Run-time (sec)*			
	Station ‘0-1’	Station ‘1-2’	Station ‘2-3’	
0→1	8	8	9	
1→2	12	8	9	
2→3	12	12	9	

*Dwell-time and run-time solutions are the same in both cases with the corresponding optimal paths

Discussion:

In Table 7.2, it can be shown that the number of possible states at the intermediate stages (i.e., stage 2 and 3) increases substantially in Case I. Likewise, in Case II, where grouping is introduced, the number of states also increases, however, the numbers are much smaller when compared with Case I, especially at stage 3.

In the state-space model, the number of stations, ST, deduces the number of possible stages; while the number of steps in the control variable, CV, indicates the number of possible extended dwell-times at stations and run-times in inter-station runs. ST and CV then determine the number of possible solutions to be expanded for a state, according to $(CV)^{ST}$. For example, when ST is 3, CV is 2, and the number of states in stage 1 is 9, then for each single state, there could be 2^3 possible solutions in the proceeding stage. In other words, in stage 2, there are altogether 9x8 states with complete state-space diagram and the multiplication goes on thereafter (i.e., the number of states in the next stage is equal to (number of states in present stage) x $(CV)^{ST}$). With dynamic programming, states containing the same trains' arrival times at stations are grouped together at a particular stage. A number of states in the proceeding stages can be reduced and a significant reduction on the scale of states expansions can then be accomplished in later stages with grouping.

Simulation results also show that the controller delivers the same cost at the final state in Case II, when compared with Case I. Given the same cost, it has been verified that the controller provides the same solution with identical traffic conditions and operational requirements with and without state grouping. The computation time and memory requirement are significantly reduced as a result of the state grouping.

7.2.2 Service and energy

This study highlights the flexibility and performance of the controller with different weightings on the service quality and energy demand in the cost function, in case studies I to V.

Table 7.5 – Traffic conditions and operational requirements

Traffic conditions					
Number of inter-station runs	4				
Allowable headway deviation	± 10%				
Nominal inter-station run-times	Run 0-1	Run 1-2	Run 2-3	Run 3-4	
	135sec	121 sec	90 sec	76 sec	
Control and operational requirements					
Control method	Run-time				
Control direction	Time extension and reduction				
Control space*	±5%				
Number of steps in control variable	Run 0-1	Run 1-2	Run 2-3	Run 3-4	
	7, 0, -7 sec	6, 0, -6 sec	4, 0, -4 sec	4, 0, -4 sec	
Operational requirements	Service**: 4% extension		Energy: 8% reduction		
Weighting factor	Case I	Case II	Case III	Case IV	Case V
	W_S=1 W_E=0	W_S=0.75 W_E=0.25	W_S=0.5 W_E=0.5	W_S=0.25 W_E=0.75	W_S=0 W_E=1
Grouping	Yes				

*Nominal schedule is used as a reference to regulate the service.
** Regularity is the prime concern in the service cost function.

Table 7.6 – Nominal and extended schedule in successive inter-station runs and their corresponding energy consumption

Inter-station runs	Nominal schedule (sec)	Energy consumption (MJ)	Extended schedule (sec)	Energy consumption (MJ)
0-1	135	20.3	135 + 7 = 142	17.2
1-2	121	38	121 + 6 = 127	36.8
2-3	90	34.38	90 + 4 = 94	31.6
3-4	76	20	76 + 4 = 80	17.05

Results:

Table 7.7 – Optimal costs

	Case I	Case II	Case III	Case IV	Case V
Cost*	0.079	0.062	0.044	0.025	0

* The cost is determined by Eqns (6.18), (6.19) and (6.20)

Table 7.8 – Run-time extensions with the corresponding optimal costs

Stage Transformation	Run-time (sec)							
	Case I				Case II			
	Run 0-1	Run 1-2	Run 2-3	Run 3-4	Run 0-1	Run 1-2	Run 2-3	Run 3-4
0→1	0	6	4	4	0	6	4	4
1→2	7	6	0	4	7	6	0	4
2→3	7	6	4	4	7	0	4	4
3→4	7	6	0	4	7	6	4	4
Stage Transformation	Run-time (sec)							
	Case III				Case IV			
	Run 0-1	Run 1-2	Run 2-3	Run 3-4	Run 0-1	Run 1-2	Run 2-3	Run 3-4
0→1	0	6	4	4	7	6	4	4
1→2	7	6	0	4	7	0	4	4
2→3	7	0	4	4	7	0	4	4
3→4	7	6	4	4	7	0	4	4
Stage Transformation	Run-time (sec)							
	Case V							
	Run 0-1	Run 1-2	Run 2-3	Run 3-4				
0→1	7	0	4	4				
1→2	7	0	4	4				
2→3	7	0	4	4				
3→4	7	0	4	4				

Discussion:

With the given traffic conditions and operational requirements, the controller provides feasible solutions in all cases and delivers the optimal path with zero cost in Case V, as shown in Table 7.7, where energy demand plays a dominant role in the cost function. It can also be shown that the optimal cost is gradually reduced when the weighting factors of energy demand (i.e.,W_E) in the cost function increases.

Further, Table 7.8 shows that the run-time extension (i.e., 6 sec) in an inter-station run '1-2' is taken in each stage transformation when W_S is set to 1 in Case I. However, the nominal schedule (i.e., 121 sec) is preferable when energy demand becomes dominant in the cost function in Case V, since the energy reduction with the extended schedule (i.e.,127 sec) is less significant when compared with other inter-station runs (i.e., inter-station runs '0-1', '2-3' and '3-4') as shown in Table 7.6. Details of the run-time extensions in successive inter-station runs with the corresponding optimal costs are given in Table 7.8.

7.2.3 Dwell-time and run-time

This study examines the impact on train performance of different control criteria under given traffic conditions and operational requirements. To investigate the applicability of different train control criteria with dynamic programming, different train operation requirements are introduced – Case A: service extension, Case B: energy reduction and Case C: service extension with energy saving.

Table 7.9 – Traffic conditions and operational requirements

<table>
<tr><td colspan="10">Traffic conditions</td></tr>
<tr><td>Number of inter-station runs</td><td colspan="9">4</td></tr>
<tr><td>Allowable headway deviation</td><td colspan="9">± 15%</td></tr>
<tr><td rowspan="2">Nominal inter-station run-times</td><td colspan="2">Run 0-1</td><td colspan="2">Run 1-2</td><td colspan="2">Run 2-3</td><td colspan="3">Run 3-4</td></tr>
<tr><td colspan="2">135sec</td><td colspan="2">121 sec</td><td colspan="2">90 sec</td><td colspan="3">76 sec</td></tr>
<tr><td colspan="10">Control and operational requirements</td></tr>
<tr><td rowspan="3">Control method</td><td colspan="3">Case I</td><td colspan="3">Case II</td><td colspan="3">Case III</td></tr>
<tr><td>A</td><td>B</td><td>C</td><td>A</td><td>B</td><td>C</td><td>A</td><td>B</td><td>C</td></tr>
<tr><td colspan="3">Dwell-time</td><td colspan="3">Run-time</td><td colspan="3">Mixed</td></tr>
<tr><td>Control direction</td><td colspan="9">Time extension</td></tr>
<tr><td>Control space*</td><td colspan="4">Run-time: +10%</td><td colspan="5">Dwell-time: +10%</td></tr>
<tr><td rowspan="4">Number of steps in control variable</td><td colspan="9">Dwell-time</td></tr>
<tr><td colspan="9">2, 1, 0 sec</td></tr>
<tr><td colspan="2">Run 0-1</td><td colspan="2">Run 1-2</td><td colspan="2">Run 2-3</td><td colspan="3">Run 3-4</td></tr>
<tr><td colspan="2">14, 7, 0 sec</td><td colspan="2">12, 6, 0 sec</td><td colspan="2">8, 4, 0 sec</td><td colspan="3">8, 4, 0 sec</td></tr>
<tr><td>Operational requirements**</td><td colspan="4">Service: 4% extension</td><td colspan="5">Energy: 8% reduction</td></tr>
<tr><td rowspan="3">Cost function***</td><td colspan="3">Case I</td><td colspan="3">Case II</td><td colspan="3">Case III</td></tr>
<tr><td>A</td><td>B</td><td>C</td><td>A</td><td>B</td><td>C</td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>S</td><td>E</td><td>S&E</td><td>S</td><td>E</td><td>S&E</td><td>S</td><td>E</td><td>S&E</td></tr>
<tr><td>Weighting factor</td><td colspan="9">$W_S = 0.5, \quad W_E = 0.5$ (For all case C only)</td></tr>
<tr><td>Grouping</td><td colspan="9">Yes</td></tr>
</table>

*Nominal schedule is used as a reference to regulate the service.

**Regularity is the prime concern when service quality (i.e., $W_S \neq 0$) is taken into account in the cost function.

***S: service dominant (i.e., $W_S = 1$); E: energy dominant (i.e., $W_E = 1$); S&E: Both service and energy are taken into account with their corresponding weights.

Results:

Table 7.10 – Optimal costs

	Case I		
	A	B*	C
Cost	0.128	-	0.107
	Case II		
	A	B	C
Cost	0.075	0	0.042
	Case III		
	A	B	C
Cost	0.086	0	0.046

*Adjustments of dwell-time of trains at stations is not applicable to achieve energy saving

Table 7.11 – Dwell-time extensions with the corresponding optimal paths

Stage Transfor -mation	Case I – Dwell-time control														
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
	Station 0			Station 1			Station 2			Station 3			Station 4		
0→1	2	-	2	2	-	2	1	-	2	0	-	2	2	-	2
1→2	2	-	2	2	-	2	2	-	1	0	-	2	2	-	2
2→3	2	-	2	2	-	2	2	-	2	2	-	0	2	-	2
3→4	2	-	2	2	-	2	2	-	2	2	-	2	2	-	2

Table 7.12 – Run-time extensions with the corresponding optimal paths

Stage Transfor -mation	Case II – Run-time control											
	A	B	C	A	B	C	A	B	C	A	B	C
	Run 0-1			Run 1-2			Run 2-3			Run 3-4		
0→1	7	14	7	6	6	6	4	4	4	4	0	4
1→2	7	7	7	0	12	0	4	8	4	4	0	4
2→3	7	7	7	6	12	6	8	8	8	4	0	0
3→4	14	0	14	12	6	12	8	4	8	4	8	0

Table 7.13 – Dwell-time and run-time extensions with the corresponding optimal paths

Stage Transfor -mation	Case III - Mixed control														
	Dwell-time (sec)														
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
	Station 0			Station 1			Station 2			Station 3			Station 4		
0→1	1	2	1	1	1	1	1	1	1	1	0	1	2	2	2
1→2	1	1	1	0	2	0	1	2	1	1	0	1	2	2	2
2→3	1	1	1	1	2	1	2	2	2	1	0	0	2	2	2
3→4	2	0	2	2	1	2	2	1	2	1	2	0	2	2	2

	Run-time (sec)											
	A	B	C	A	B	C	A	B	C	A	B	C
	Station 0-1			Station 1-2			Station 2-3			Station 3-4		
0→1	7	14	7	6	6	6	4	4	4	4	0	4
1→2	7	7	7	0	12	0	4	8	4	4	0	4
2→3	7	7	7	6	12	6	8	8	8	4	0	0
3→4	14	0	14	12	6	12	8	4	8	4	8	0

Discussion:

Table 7.10 shows that run-time control in successive inter-station runs is the best approach to maintain the train service with different operational requirements, since the controller delivers the optimal path with the lowest cost, as shown in Case II of tests A, B and C, when compared with those in Case I and Case III respectively. Further, run-time adjustment in successive inter-station runs provides high flexibility in train control and also achieves energy saving, when compared with dwell-time control.

Simulation results reveal that regulation of train service can also be achieved with dwell-time control in test A of Case I, but it is not effective to maintain the train service when energy saving is the prime concern, as in test B, because the run-times in successive inter-stations runs are kept constant. Further, the optimal cost in test C of Case I is lower than that in test A of Case I, since, in test C, service quality and energy demand are both taken into account in the cost function and the change of dwell-times at stations has no impact on the energy demand in the cost function.

In addition, a higher flexibility in train regulation can be attained with mixed control (i.e., both dwell-time and run-time control are allowed at the same time). Nevertheless, it is worth noting that the optimal costs obtained in tests A and C of Case III are slightly higher than those in tests A and C of Case II even though mixed control is applied in Case III. It is because dwell-time extensions at stations inevitably affect the train service quality but has no impact on energy saving. Moreover, the optimisation problem of train scheduling gets more complicated and the solution space becomes larger with the introduction of mixed control. More computation time and higher memory storage requirement are expected. Furthermore, dwell-time control is the second best option to maintain the train schedule from the view point of energy saving, for it can only be introduced when the run-times in successive inter-station runs can no longer be extended.

7.2.4 Track layout

This study verifies the controller's capability in obtaining the optimal adjustment on run-times for trains in the 3 successive inter-station runs, under given traffic conditions and operational constraints. Four sets of track topologies with three successive inter-station runs are investigated, in Cases I to IV respectively. Two operational requirements for trains, Test A - 5% extension of service regularity, and Test B - 15% reduction on the energy consumption, are adopted for the 4 different runs.

Table 7.14 – Traffic conditions and operational requirements

<table>
<tr><th colspan="10">Traffic conditions</th></tr>
<tr><td colspan="2">Number of inter-station runs</td><td colspan="8">3</td></tr>
<tr><td colspan="2">Allowable headway deviation</td><td colspan="8">± 10%</td></tr>
<tr><td colspan="2">Nominal inter-station run-times (sec)</td><td colspan="2">Case I</td><td colspan="2">Case II</td><td colspan="2">Case III</td><td colspan="2">Case IV</td></tr>
<tr><td rowspan="3"></td><td>Run 0-1</td><td colspan="2">135</td><td colspan="2">90</td><td colspan="2">76</td><td colspan="2">135</td></tr>
<tr><td>Run 1-2</td><td colspan="2">90</td><td colspan="2">135</td><td colspan="2">90</td><td colspan="2">76</td></tr>
<tr><td>Run 2-3</td><td colspan="2">76</td><td colspan="2">76</td><td colspan="2">135</td><td colspan="2">90</td></tr>
<tr><th colspan="10">Control and operational requirements</th></tr>
<tr><td>Control method</td><td colspan="9">Run-time</td></tr>
<tr><td>Control direction</td><td colspan="9">Time extension</td></tr>
<tr><td rowspan="4">Number of steps in control variable</td><td></td><td colspan="2">Case I</td><td colspan="2">Case II</td><td colspan="2">Case III</td><td colspan="2">Case IV</td></tr>
<tr><td>Run 0-1</td><td colspan="2">12, 8, 4, 0 sec</td><td colspan="2">9, 6, 3, 0 sec</td><td colspan="2">6, 4, 2, 0 sec</td><td colspan="2">12, 8, 4, 0 sec</td></tr>
<tr><td>Run 1-2</td><td colspan="2">9, 6, 3, 0 sec</td><td colspan="2">12, 8, 4, 0 sec</td><td colspan="2">9, 6, 3, 0 sec</td><td colspan="2">6, 4, 2, 0 sec</td></tr>
<tr><td>Run 2-3</td><td colspan="2">6, 4, 2, 0 sec</td><td colspan="2">6, 4, 2, 0 sec</td><td colspan="2">12, 8, 4, 0 sec</td><td colspan="2">9, 6, 3, 0 sec</td></tr>
<tr><td rowspan="3">Operational requirements</td><td>Case I</td><td>Case II</td><td>Case III</td><td colspan="2">Case IV</td><td>Case I</td><td>Case II</td><td>Case III</td><td>Case IV</td></tr>
<tr><td colspan="5">A</td><td colspan="4">B</td></tr>
<tr><td colspan="5">Service regularity: 5% extension</td><td colspan="4">Energy: 15% reduction</td></tr>
</table>

*Nominal schedule is used as a reference to regulate the service

Results:

Table 7.15 – Run-time extensions

<table>
<tr><th>Test A</th><th colspan="9">Run-time (sec)</th></tr>
<tr><td rowspan="2">Stage Transfor-mation</td><td colspan="3">Case I</td><td colspan="3">Case II</td><td colspan="3">Case III</td></tr>
<tr><td>Run 0-1</td><td>Run 1-2</td><td>Run 2-3</td><td>Run 0-1</td><td>Run 1-2</td><td>Run 2-3</td><td>Run 0-1</td><td>Run 1-2</td><td>Run 2-3</td></tr>
<tr><td>0→1</td><td>4</td><td>6</td><td>6</td><td>6</td><td>8</td><td>6</td><td>6</td><td>6</td><td>8</td></tr>
<tr><td>1→2</td><td>8</td><td>9</td><td>6</td><td>9</td><td>8</td><td>4</td><td>6</td><td>6</td><td>8</td></tr>
<tr><td>2→3</td><td>12</td><td>9</td><td>6</td><td>9</td><td>12</td><td>4</td><td>6</td><td>9</td><td>8</td></tr>
<tr><td rowspan="2"></td><td colspan="3">Case IV</td><td colspan="6" rowspan="5"></td></tr>
<tr><td>Run 0-1</td><td>Run 1-2</td><td>Run 2-3</td></tr>
<tr><td>0→1</td><td>8</td><td>6</td><td>6</td></tr>
<tr><td>1→2</td><td>12</td><td>4</td><td>6</td></tr>
<tr><td>2→3</td><td>12</td><td>6</td><td>6</td></tr>
</table>

To be continued

Test B	Run-time (sec)								
Stage Transfor-mation	Case I			Case II			Case III		
	Run 0-1	Run 1-2	Run 2-3	Run 0-1	Run 1-2	Run 2-3	Run 0-1	Run 1-2	Run 2-3
0→1	12	6	0	9	8	0	6	9	0
1→2	8	9	0	6	12	0	6	9	0
2→3	0	9	6	0	12	6	0	6	12
	Case IV								
	Run 0-1	Run 1-2	Run 2-3						
0→1	12	6	0						
1→2	12	6	0						
2→3	0	6	9						

Discussion:

In this study, a number of tests with the identical number of inter-station runs but different run-times in successive stations have been examined under the specified operational requirements. As shown in Table 7.15, the train service in test A is extended by 5%. It is worth noting that the maximum run-time extensions (i.e., 6sec) is usually preferred in a short inter-station run. Nevertheless, the run-time is likely to be extended (i.e., 12sec) in a long inter-station run when energy saving is the main concern, like the cases in tests B (i.e., 15% reduction). In a long inter-station run, a train usually travels at a higher speed until it is necessary to brake for the next station. Greater energy consumption is thus required. A significant energy reduction is possible when the run-time between stations can be extended as much as possible.

However, a train spends less time on motoring at high speed in a short inter-station run and a significant energy reduction may not be obtained with a longer run-time. It can therefore be concluded that run time is more likely to be extended in a long inter-station distance from the viewpoint of energy saving, whereas the time extension is introduced to short inter-station runs when service quality is the prime concern.

7.2.5 Impact of time delay on train control

This study investigates the controller's capability in determining the states for recovery of train service in DP, with different levels of disruptions to train service at a station, as shown in Table 7.18. The traffic conditions and operational requirements are given in Table 7.16. The controller regulates the train service within its own region which covers the five stations, and each train is fed into the line at the origin station '0' with 120 sec headway. The arrival time for trains at the stations with 120 sec headway is given in Table 7.17.

Table 7.16 – Traffic conditions and operational requirements

Traffic conditions				
Number of inter-station runs	4			
Allowable headway deviation	± 10% (i.e., ± 12 sec)			
Nominal inter-station run-times	Run 0-1	Run 1-2	Run 2-3	Run 3-4
	135sec	121 sec	90 sec	76 sec
Control and operational requirements				
Control method	Run-time			
Control direction	Time extension and reduction			
Control space	±5%			
Number of steps in control variable	Run 0-1	Run 1-2	Run 2-3	Run 3-4
	7, 0, -7 sec	6, 0, -6 sec	4, 0, -4 sec	4, 0, -4 sec
Cost function	Service regularity			

Table 7.17 – Arrival time for trains at successive stations with the headway of 120sec

Stage	Arrival time for trains at stations (sec)				
	0	1	2	3	4
0	455	495	521	516	497
1	575	615	641	636	617
2*	695	735	761	756	737
3	815	855	881	876	857
4	935	975	1001	996	977

* The disturbed traffic condition is given in stage 2

Table 7.18 – Disturbed traffic conditions

Case	Station	Time delay (sec)
A	3	60
B	3	30
C	3	10

Results:

Table 7.19 – Regulation of run-time for trains in successive inter-station runs for the recovery of train service from disturbance

Stage Transformation	Run 0-1	Run 1-2	Run 2-3	Run 3-4
	Case A			
0→1	0	0	0	- 4
1→2	7	6	4	4
2→3	7	6	4	0
3→4	7	6	4	4
	Case B			
0→1	0	0	0	- 4
1→2	0	6	4	- 4
2→3	0	6	4	4
3→4	0	6	4	4
	Case C			
0→1	0	0	0	- 4
1→2	0	0	0	4
2→3	0	0	0	4
3→4	0	0	0	4

Table 7.20 – Arrival time for trains at successive stations for the recovery of service from disturbance

Stage	Arrival time for trains at stations (sec)				
	0	1	2	3	4
	Case A				
0*	695	735	761	756	737
1**	827	879	917	924	913
2	947	994	1031	1036	1029
3	1067	1114	1146	1150	1137
4	1187	1234	1266	1265	1255
	Case B				
0*	695	735	761	756	737
1**	815	855	887	894	883
2	935	975	1007	1006	991
3	1055	1095	1127	1126	1111
4	1175	1215	1247	1246	1231

To be continued

	Case C				
0*	695	735	761	756	737
1**	815	855	881	876	863
2	935	975	1001	996	981
3	1055	1095	1121	1116	1101
4	1175	1215	1241	1236	1221

State with '*' and '**' are the disrupted state and the initial state for recovery respectively.

Discussion:

With reference to the traffic conditions and operational requirements shown in Table 7.16, headway of trains at stations is set at 120 sec and a maximum of 10% reduction on headway (i.e., 12 sec) is allowed. The two sets of corresponding arrival times of trains at successive stations (i.e., 0-4) are [695, 735, 761, 756 and 737] sec in stage 2, and [815, 855, 881, 876 and 857] sec in stage 3 through DP-I, as depicted in Table 7.17. When a train stops at station '3' with the additional time delay of 60, 30, and 10 sec in Cases A, B and C in the disrupted state respectively, the inductive time delays to the following trains at stations increase with the actual time delay to the train at station '3'. The corresponding inductive time delays for the trains behind at stations are given in Table 7.21.

Table 7.21 – Inductive time delay for trains at the stations behind

	Inductive time delay (sec)		
	Station '0'	Station '1'	Station '2'
Case A	24	36	48
Case B	- 6	6	18
Case C	- 26	- 14	- 2

With the 60 sec time delay to the train at station '3' in Case A, the inductive time delays of trains at the stations 0, 1, 2 are 24, 36, and 48 sec respectively with the time separation governed by the headway deviation (i.e., 12 sec). The arrival time for trains at stations is then adjusted to [827, 879, 917, 924 and 913] sec in the initial state for recovery in stage 1, according to the rules of the adjustments to the arrival times for trains stated in Section 6.4.3. The states after the initial state for recovery can be simply obtained through DP-II.

In addition, the corresponding inductive time delays for trains at stations behind are smaller than 0 in Cases B and C. No significant distortion of the traffic to the following trains is achieved in both cases. Nevertheless, the traffic for the trains at and behind station '3' is still slightly disturbed with the corresponding time disruption. Table 7.20 shows the optimal states obtained to recover the service with DP-II in the three cases.

7.2.6 Time disruptions at several stations

This study shows the controller's capability in delivering the appropriate arrival time for trains at successive stations with run-time control, for recovery of train service with the disturbed trains at two different stations given in Table 7.22. The traffic condition and operational requirements are the same as those in Section 7.2.5, and the arrival times for trains at stations shown in Table 7.17 remain unchanged.

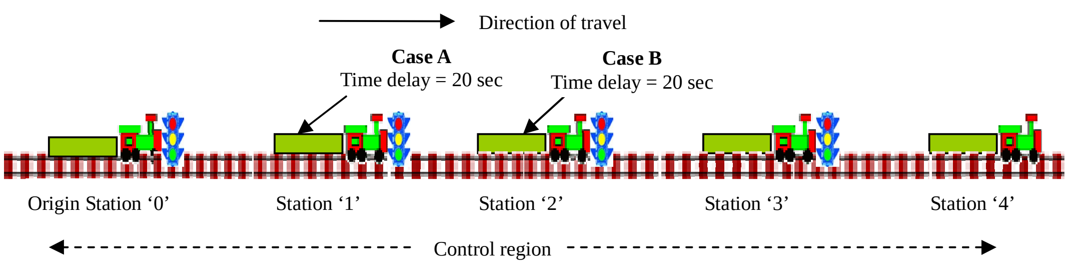


Fig. 7.1 Traffic conditions

Table 7.22 – Disturbed traffic conditions

Case	Station	Time delay (sec)
A	1	20
B	2	20

Results:

Table 7.23 – Regulation of run-time for trains in successive inter-station runs for the recovery of train service from disturbance

Stage Transformation	Run 0-1	Run 1-2	Run 2-3	Run 3-4
	Case A			
0→1	0	- 6	4	4
1→2	7	6	- 4	0
2→3	7	6	- 4	- 4
3→4	7	6	0	- 4
	Case B			
0→1	0	0	- 4	0
1→2	0	6	4	- 4
2→3	0	6	4	0
3→4	0	6	4	4

Table 7.24 – Arrival time for trains at successive stations for the recovery of service from disturbance

Stage	Arrival time for trains at stations (sec)				
	0	1	2	3	4
	Case A				
0*	695	735	761	756	737
1**	811	863	895	880	861
2	931	978	1015	1006	981
3	1051	1098	1130	1126	1103
4	1171	1218	1250	1245	1223
	Case B				
0*	695	735	761	756	737
1**	815	855	889	892	857
2	935	975	1007	1008	989
3	1055	1095	1127	1126	1109
4	1175	1215	1247	1246	1231

State with ‘*’ and ‘**’ are the disrupted state and the initial state for recovery respectively.

Table 7.25 – Changes in headway at stations in successive stages

Stage	Headway (sec)				
	Station ‘0’	Station ‘1’	Station ‘2’	Station ‘3’	Station ‘4’
	Case A				
1	116	128	134	124	124
2	120	115	120	126	120
3	120	120	115	120	122
4	120	120	120	119	120
	Case B				
1	120	120	128	136	120
2	120	120	118	116	132
3	120	120	120	118	120
4	120	120	120	120	122

Discussion:

With 20sec time delay to the train at station ‘1’ in Case A, and ‘2’ in Case B as shown in Table 7.22, the controller is capable of providing the appropriate arrival times for trains at stations in both cases, for the recovery of train service from disruption. Table 7.25 also depicts that the service regularity of trains is gradually improved through a series of successive stage transformations stipulated by DP-II, and the 120 sec of time separation of trains at stations is almost resumed in stage 4 in both cases.

Fig. 7.2 in Case A and Fig. 7.3 in Case B illustrate the changes in headway with stages at stations ‘0’ to ‘4’ respectively. With the time delay of train at station ‘*n*’ in the disrupted state, the change in headway at station ‘*n*+1’ ahead is the largest in both cases (i.e., station ‘2’ in Case A and station ‘3’ in Case B), when compared with that at other stations. The arrival time of the disturbed train at station ‘*n*+1’ ahead in the initial state for recovery is directly imposed by the time delay of that train and its run-times in the subsequent inter-station runs. The headway of trains at station ‘*n*+1’ is then gradually improved with the appropriate control actions on run-times through a number of stage transformations. In other words, a significantly irregular headway is observed at station ‘*n*+1’ ahead when the time disruption is introduced to a train at station ‘*n*’. Further, a larger time disruption to train at station ‘*n*’ implies a larger difference in headway at station ‘*n*+1’ in stages.

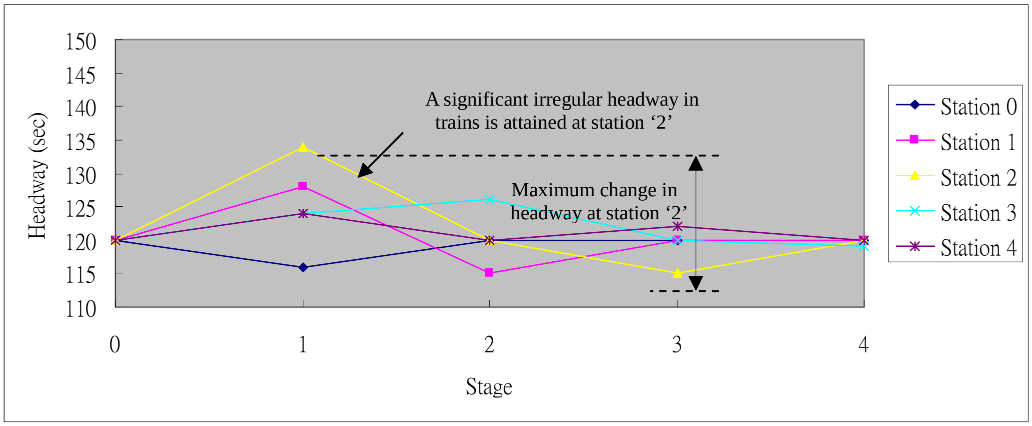


Fig. 7.2 Changes in headway with stages – Case A

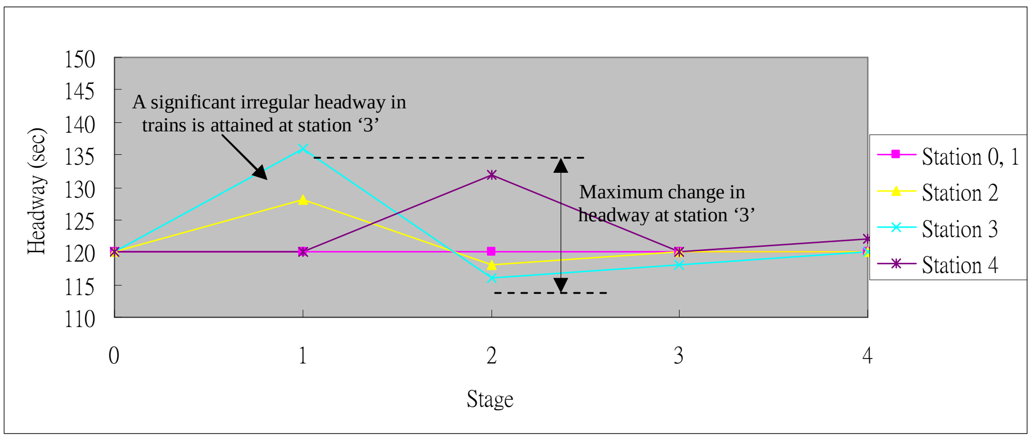


Fig. 7.3 Changes in headway with stages – Case B

Table 7.25 shows that a slight change in headway in stages is achieved at the station ‘4’ in Case A, when compared with that in Case B. It is because there is one more subsequent inter-station run for trains to recover the service in Case A. It thereby demonstrates that a larger control space for recovery is available when the disturbed train is far away from the ‘overlapping’ station in a region. Moreover, the maximum number of stage transformations increases directly with the total number of inter-station runs in a region.

In addition, irregular train separation caused in region *k* is introduced to the following region *k*+1 through the ‘overlapping’ station if the disruption is large and close to the end of region *k*. Regulation of train operation will then be further carried out by the next regional controller. Train coordination to recover the service in two successive regions will be discussed in the next study.

7.2.7 Recovery of train service in multiple regions

This study demonstrates the approach to recover the train service from disruption with arrival-time coordination of trains at the overlapping station in the two successive regions – stations ‘0’ to ‘4’ in region 1; stations ‘4’ to ‘8’ in region 2; and station ‘4’ being the ‘overlapping’ station between the two regions. The traffic condition and operational requirements in region 1 and 2 are the same, as shown in Table 7.16. The arrival times for trains in the two regions are given in Table 7.26, and the disturbed traffic conditions are depicted in Table 7.27.

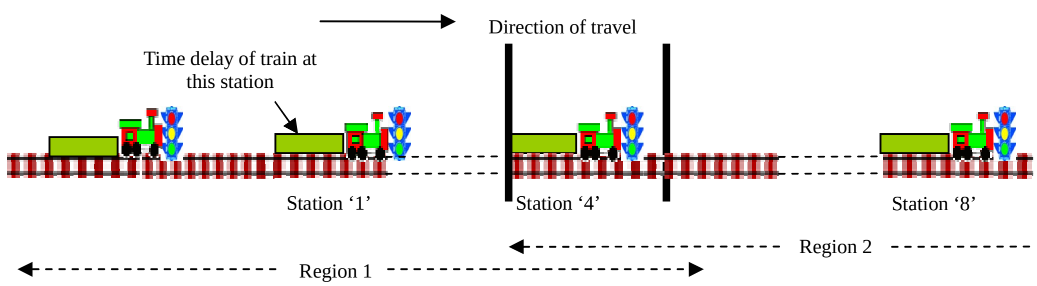


Fig. 7.4 Traffic conditions

Table 7.26 – Arrival time for trains at successive stations with a headway of 120sec in regions 1 and 2

Stage	Arrival time for trains at stations (sec)				
	Region 1				
	0	1	2	3	4
0	935	975	1001	996	977
1	1055	1095	1121	1116	1097
2	1175	**1215**	1241	1236	1217
3	1295	1335	1361	1356	1337
4	1415	1455	1481	1476	1457
	Region 2				
	4	5	6	7	8
0	977	1017	1043	1038	1019
1	1097	1137	1163	1158	1139
2	1217	1257	1283	1278	1259
3	1337	1377	1403	1398	1379
4	1457	1497	1523	1518	1499

Table 7.27 – Disturbed traffic conditions

Case	Station	Region	Time delay (sec)
A	1	1	20
B	1	1	60

Results:

Table 7.28 – Optimal costs in regions 1 and 2

	Region 1	Region 2
	Case A	
Cost	0.088	0.025
	Case B	
Cost	0.408	0.152

Table 7.29 – Regulation of run-time for trains in successive inter-station runs for the recovery of train service from disturbance in regions 1 and 2

Stage Transformation	Case A			
	Region 1			
	Run 0-1	Run 1-2	Run 2-3	Run 3-4
0→1	0	- 6	4	4
1→2	7	6	- 4	0
2→3	7	6	- 4	- 4
3→4	7	6	0	- 4
	Region 2			
	Run 4-5	Run 5-6	Run 6-7	Run 7-8
0→1	0	0	0	0
1→2	- 7	0	0	0
2→3	- 7	6	0	0
3→4	- 7	6	- 4	0
Stage Transformation	Case B			
	Region 1			
	Run 0-1	Run 1-2	Run 2-3	Run 3-4
0→1	0	- 6	4	4
1→2	7	6	- 4	4
2→3	7	6	- 4	- 4
3→4	7	6	0	- 4
	Region 2			
	Run 4-5	Run 5-6	Run 6-7	Run 7-8
0→1	0	0	0	0
1→2	0	0	0	0
2→3	0	0	0	0
3→4	- 7	- 6	- 4	0

Table 7.30 – Arrival time for trains at successive stations for the recovery of service from disturbance

Stage	Arrival time for trains at stations (sec)				
	Case A				
	Region 1				
	0	1	2	3	4
0*	1175	1215	1241	1236	1217
1**	1291	1343	1375	1360	1341
2	1411	1458	1495	1486	1461
3	1531	1578	1610	1606	1583
4	1651	1698	1730	1725	1703
	Region 2				
	4	5	6	7	8
0*	1217	1257	1283	1278	1259
1**	1341	1377	1403	1398	1379
2	1461	1494	1523	1518	1499
3	1583	1614	1646	1638	1619
4	1703	1736	1766	1757	1739

To be continued

Stage	Case B				
	Region 1				
	0	1	2	3	4
0*	1175	1215	1241	1236	1217
1**	1331	1383	1415	1360	1341
2	1451	1498	1535	1526	1465
3	1571	1618	1650	1646	1623
4	1691	1738	1770	1765	1743
	Region 2				
	4	5	6	7	8
0*	1217	1257	1283	1278	1259
1**	1341	1377	1403	1398	1379
2	1465	1501	1523	1518	1499
3	1623	1625	1647	1638	1619
4	1743	1776	1765	1758	1739

State with ‘*’ and ‘**’ are the disrupted state and the initial state for recovery respectively.

Table 7.31 – Changes in headway at stations in successive stages

Stage	Headway (sec)				
	Case A				
	Region 1				
	Station ‘0’	Station ‘1’	Station ‘2’	Station ‘3’	Station ‘4’
1	116	128	134	124	124
2	120	115	120	126	120
3	120	120	115	120	122
4	120	120	120	119	120
	Region 2				
	Station ‘4’	Station ‘5’	Station ‘6’	Station ‘7’	Station ‘8’
1	124	120	120	120	120
2	120	117	120	120	120
3	122	120	123	120	120
4	120	122	120	119	120
Stage	Case B				
	Region 1				
	Station ‘0’	Station ‘1’	Station ‘2’	Station ‘3’	Station ‘4’
1	156	168	174	124	124
2	120	115	120	166	124
3	120	120	115	120	158
4	120	120	120	119	120
	Region 2				
	Station ‘4’	Station ‘5’	Station ‘6’	Station ‘7’	Station ‘8’
1	124	120	120	120	120
2	124	124	120	120	120
3	158	124	124	120	120
4	120	151	118	120	120

Discussion:

With a time delay of 20 sec for a train in Case A and 60 sec for train in Case B in region 1, Table 7.28 shows the two corresponding optimal costs are reduced in region 2 when compared to those in region 1 in both cases. A more significant reduction on the optimal cost (i.e., 0.408-0.152 = 0.256) is attained in Case B with a larger time disruption to train. It has therefore shown that the two controllers are capable of regulating the train operation gradually from disruption with appropriate coordination between two successive regions. To coordinate train operation in the two successive regions, station ‘4’ is located in both regions to transfer the train operation from one region to another in this application. The arrival time for trains at station ‘4’ in the successive stages in region 1 are [1217, 1341, 1461, 1583, 1703] sec in Case A, and [1217, 1341, 1465, 1623, 1743] sec in Case B, and these two sets of arrival times are then introduced into region 2 as its controller’s operational constraints (i.e., the ‘fed-in’ time constraint of each train approaching region 2 through the station ‘4’).

It is also shown in Table 7.31 that the impact of the time delay of train at station ‘1’ is evenly distributed to a number of stations with the given operational constraints, through the optimal path from DP-II in region 1. However, a fixed 120 sec headway of trains cannot be fully achieved yet because the train control does not only depend upon the resolution of the step change in the run-time variables through the inter-station runs, but also the magnitude and location of the time disruption. Train control becomes more sophisticated with a higher number of steps (i.e., higher resolution) in control variables. However, more computation time and higher memory requirement are the drawbacks. To conclude, the controller is successful in regulating the train operation from disruption with appropriate coordination between regions.

7.2.8 Multiple disruptions

The traffic conditions and operational requirements remain the same as those in the case study described in Section 7.2.5, but the two trains departing from stations ‘1’ and ‘3’ in stage 2 are both delayed. To further examine the controller’s versatility under various disturbed traffic conditions, three cases of multiple disruptions with different time delays at stations in a region are given.

Table 7.32 – Introduction of time delay at the two stations

	Time delay (sec)	
	Station ‘1’	Station ‘3’
Case A	30	60
Case B	10	30
Case C	60	30

Results:

Table 7.33 – Regulation of run-time for trains for the recovery of train service from disturbance

Stage Transformation	Run 0-1	Run 1-2	Run 2-3	Run 3-4
	Case A			
$0 \rightarrow 1$	0	0	0	- 4
$1 \rightarrow 2$	7	6	4	4
$2 \rightarrow 3$	7	6	4	0
$3 \rightarrow 4$	7	6	4	4
	Case B			
$0 \rightarrow 1$	0	0	0	- 4
$1 \rightarrow 2$	0	6	4	- 4
$2 \rightarrow 3$	0	6	4	0
$3 \rightarrow 4$	0	6	4	4
	Case C			
$0 \rightarrow 1$	0	- 6	0	- 4
$1 \rightarrow 2$	7	6	- 4	4
$2 \rightarrow 3$	7	6	- 4	- 4
$3 \rightarrow 4$	7	6	0	- 4

Table 7.34 – Arrival time for trains at successive stations for the recovery of service from disturbance

Stage	Arrival time for trains at stations (sec)				
	0	1	2	3	4
	Case A				
0*	695	735	761	756	737
1**	827	879	917	924	913
2	947	994	1031	1036	1029
3	1067	1114	1146	1150	1137
4	1187	1234	1266	1265	1255
	Case B				
0*	695	735	761	756	737
1**	815	855	891	894	883
2	935	975	1007	1010	991
3	1055	1095	1127	1126	1111
4	1175	1215	1247	1246	1231
	Case C				
0*	695	735	761	756	737
1**	851	903	935	894	883
2	971	1018	1055	1046	999
3	1091	1138	1170	1166	1143
4	1211	1258	1290	1285	1263

State with ‘*’ and ‘**’ are the disrupted state and the initial state for recovery respectively.

Table 7.35 – Changes in headway at stations in successive stages

Stage	Headway (sec)				
	Station ‘0’	Station ‘1’	Station ‘2’	Station ‘3’	Station ‘4’
	Case A				
1	132	144	156	168	176
2	120	115	114	112	116
3	120	120	115	114	108
4	120	120	120	115	118
	Case B				
1	120	120	130	138	146
2	120	120	116	116	108
3	120	120	120	116	120
4	120	120	120	120	120
	Case C				
1	156	168	174	138	146
2	120	115	120	152	116
3	120	120	115	120	144
4	120	120	120	119	120

Discussion:

With time delay of train at station ‘3’ shown in Cases A, B and C of Table 7.32, the corresponding inductive time delay to the following trains are listed below:

Table 7.36 – Inductive time delay for the following trains with the disruptions at station '3'

	Inductive time delay (sec)		
	Station '0'	Station '1'	Station '2'
Case A	24	36	48
Case B	-6	6	18
Case C	-6	6	18

Table 7.37 – Inductive time delay for the following trains with the two combined disruptions at stations '1' and '3'

	Inductive time delay (sec)		
	Station '0'	Station '1'	Station '2'
Case A	24	36	48
Case B	-2	10	18
Case C	48	60	18

In Case A, the disruption for the train at station '3' imposes a significant inductive time delay on the traffic in the trains behind. The inductive time delay of the train at station '1' is even larger than the actual disruption of the train at the same station (i.e., 36>30 sec). The traffic behind is therefore dominated by the disruption of the train at station '3'. In Cases B and C, when combining with the two disruptions at stations '1' and '3', the inductive delays for the trains behind are updated as in Table 7.37. The inductive delay at station '0' is revised with the time disruption of train at station '1', instead of that at station '3', and the inductive time delay of 6 sec at station '1' is replaced by the actual disruption of the train at the station and increased to 10 sec in Case B, and 60 sec in Case C.

With the given inductive time delay for the following trains at stations, the arrival times for trains at stations is adjusted to [827, 879, 917, 924 and 913] sec in Case A, [815, 855, 891, 894, 883] sec in Case B and [851, 903, 935, 894, 883] sec in Case C in the initial state for recovery. Table 7.35 also reveals that the headway varies with the two time disruptions at stations ‘1’ and ‘3’. A larger disruption of the train at stations results in a more significant irregular headway to the following region through station ‘4’. To conclude, the controller enables the train movement coordination for recovery of train service when multiple disruptions occur.

7.3 Concluding remarks

A regional train controller to adjust the dwell-time of trains at stations and run-time of trains in successive inter station runs has been presented, for which a dynamic programming approach has been adopted. The complexity and size of the state-space traffic flow model depends on the number of stations and thus inter-stations runs and the number of possible sets of control actions. The number of states in the model is significantly increased with the number of inter-station runs. To reduce the possible number of states, states with the same arrival times are combined through grouping. Simulation results show that the possible number of states at the intermediate stages is significantly reduced through grouping in the model. Shorter simulation time and lower memory requirements are therefore possible.

Regulation of the run-times in successive inter-station runs and of the dwell-times at stations with respect to the nominal schedule are the two viable alternatives to maintain train service with respect to the time-varying service demand. From the viewpoint of energy saving, run-time control is superior to dwell-time control because a longer run-time between stations implies a lower energy consumption. However, energy reduction cannot be achieved by lengthening the waiting time of train at station as the run-time in inter-station run remains the same, unless regenerative braking is taken into account. Further, a higher flexibility of train regulation can be achieved with run-time control when compared to dwell-time control. It is because run-times between stations are usually much longer than the dwell-time and hence a larger control space is available for the operators.

On the recovery of train service from disruptions, the controller decides the appropriate arrival times for trains at stations for various disruption(s). Inductive delay is introduced to determine the valid state from the disruption, with the rules of the adjustments on the arrival times for trains. Coordination and regulation of train operation in multiple regions are adopted to maintain the service when the disruption of the train at a station is large and the disturbed train is located toward the end of a region.

With the application of dynamic programming, the controller is capable of delivering the optimal path in the state diagram for real-time control under various traffic conditions and operational requirements. Given that a train usually takes 10 minutes or more to run through three or more stations in metro systems and the computational demand in the proposed control escalates drastically with the number of stages, DP still ensures an optimal solution for real-time applications. Heuristic methods can be applied to solve this problem and they usually consume less memory. However, they do not guarantee optimal solution. In practice, the regional train controllers can be integrated into the central control room.

Chapter 8
Central Train Controller

In daily railway operation, the frequency of train service is adjusted to cater for the varying passenger demand, in accordance with the predetermined ‘timetable’. During the transition between two levels of service headway, two patterns cover the same line simultaneously and the traffic pattern is disrupted. The peak kVA demand on the supply system may be either higher or lower because of the transition. A penalty on the tariff for train operation is imposed when the peak demand exceeds the specified threshold during the transition. From the operational point of view, a sudden and short period of peak demand may be avoided by splitting a one-step change of headway into a number of smaller changes. An advisory system to provide the appropriate sequence of changes of headway and minimise the overall peak kVA demand on the supply system is desirable.

8.1 Introduction

8.1.1 Methodology

This chapter describes the design and implementation of the central train controller (*CTC*) at the system level of the *HTRC*. The function of the *CTC* is to determine the appropriate sequence of headway-step changes to update one level of headway to another, either from small to large (e.g., 120sec→240sec) or vice versa. The achieved transitions minimise the overall peak kVA demand on the supply system and maintain a regular train service for passengers.

As dwell-time and run-time adjustments on each train are not conducted at the system level of control and headway is the only viable control variable, the complexity of the control may not be as high as that in the *RTC*. Bearing this in mind, the time required to find the appropriate sequence of changes of headway is not substantial. Dynamic programming (DP) and even exhaustive search are the possible approaches to seek for the optimal headway-sequence. Studies on feasibility and performance of headway control by *CTC* are then undertaken for the purpose of real-time operations in the next chapter.

8.1.2 Literature reviews

A number of methods for reducing the peak kVA demand on the traction supply system have been proposed in recent years. Takeuchi and Goodman [87,88] investigated the starting behaviour of a queue of trains for recovery of train service from disturbance under both FBS and MBS. Since each section of track is occupied by no more than one train at a time in FBS, a "natural" starting time delay is given to the following trains for recovery from disturbance. Trains are prohibited to start from a stop till the leading train clears the block section in front. However, trains start at the same time in MBS when the leading train moves. The peak demand incurred on the supply system is therefore lower with FBS than that with MBS during service recovery. Starting time delays (STD) and acceleration rate limits (ARL) are introduced to minimise the peak kVA demand on the supply system in metro operations under MBS. Simulation results show that ARL achieve better peak demand reduction under certain traffic conditions. However, the approaches of STD and ARL are not further extended to regulate train operation in normal traffic conditions.

Sanso and Girard [89] developed a heuristic-based train controller to reduce the peak kVA demand on the supply system under normal traffic condition. The controller provides a set of optimised dwell-time schedules at successive stations. Simulation results reveal that the peak demand on the supply system is reduced by 13~17% with a maximum of 25 seconds of delay in the service. However, the complexity of the control is greatly increased as the number of station grows.

These two studies show the successful reduction of peak demand on the supply system with either normal traffic condition or recovery of service from disturbance. Details of train operational instructions (i.e., STD or ARL of trains) are required, and a high complexity of train control is formed to carry out regulation of train service. To demonstrate the capability of train control for real-time applications, a low complexity of headway control is proposed at the *CTC* level for normal traffic regulation.

8.2 Problem formulation

8.2.1 Operation to timetable

In normal traffic condition, the frequency of train service follows the predetermined 'timetable' to meet the passenger demand variations. The timetable is determined by a number of factors, such as the timing of peak hours (usually in the mornings and after-work hours) and the location of stations (demand is higher if the station is located in a densely populated area), and it is designed to meet the flow of passengers along the line by dispatching an appropriate number of trains during a given time period, as described by the service headway.

An example of the ‘timetable’ describing the appropriate frequency of train service in each period, is illustrated in Table 8.1. For example, a 2-minute (i.e., 120sec) headway is adopted during the morning peak hours of 07:00 to 10:00, followed by a 5-minute (i.e., 300sec) headway till noon and so on. The time and distance profile of train operation is given in Fig. 8.1. Such practice is applied and followed in normal traffic conditions. For simplicity in the following descriptions, a series of changes in the headway of trains throughout a day according to the timetable is defined as the ‘nominal daily headway change schedule’ and the necessary headway allocated in each time period is called the ‘nominal service headway’. The time period to move from one level of nominal headway to another is referred to as the ‘change-over’.

Table 8.1 – Nominal daily headway changes schedule

Time interval	06:00 ~ 07:00	07:00 ~ 10:00	10:00 ~ 12:00	…	14:00 ~ 16:00	…	17:00 ~ 19:00	…	22:00 ~ 24:00
Headway (sec)	240	120	300	…	640	…	120	…	360

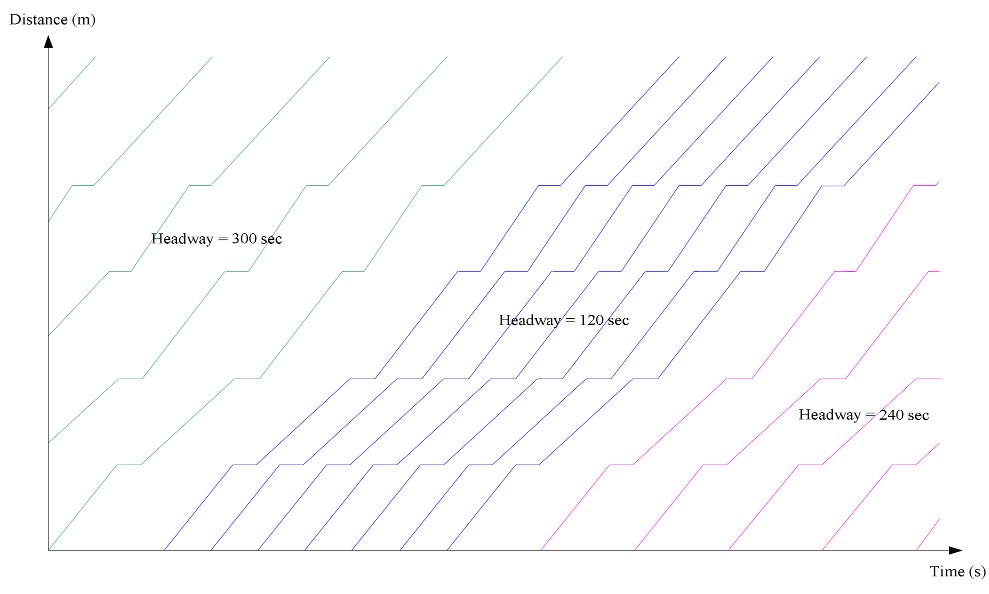


Fig. 8.1 Time and distance profile

When there is a significant delay in the traffic, the pre-determined time schedule for trains over a certain time period is no longer useful for the operators as a long and dense queue of trains is usually formed along the line. To recover the service from such disturbance, a common practice is to make the disrupted train depart the station as soon as possible and to introduce the pre-set minimum service headway to the trains behind (i.e., a more frequent service) to catch up with the schedule and cope with the passenger crowds at stations until the service is totally recovered [90]. Inevitably, a shorter headway train service places a higher power demand on the supply system.

As far as the operational cost is concerned, there are two major parameters of power demand in train operation: peak kVA demand and kWh consumption. Peak kVA demand is the instantaneous maximum power demand of trains, while kWh consumption is the energy usage of trains over a given time period. There is no room for reducing the kVA demand and kWh consumption is placed on the supply system when a constant nominal headway is maintained.

The traffic pattern changes when the nominal headway is to be changed from one level to another. The peak kVA demand on the supply system may be either higher or lower during the 'change-over'. In the case when the peak kVA demand is higher than the specific threshold, because of a number of trains accelerating simultaneously, a higher operational cost is charged by the power utility. However, an analytical model to link peak kVA demand and service headway may not be a technically viable option because there are so many inter-dependent parameters involved. Some of the parameters are system-related while the others are track-topology related. In addition, a special tariff, which encourages a low peak kVA demand, is usually negotiated for the metro operators. Hence, it is attractive for the operators to minimise the peak kVA demand as much as possible, provided that the service is maintained. A supervisory system is thus preferred which can provide the appropriate sequence of changes of headway and to achieve a low overall peak kVA demand on the supply system.

8.2.2 Headway control

To minimise kVA demand on the supply system in 'change-over', it is possible to change one level of nominal headway to another in stages, instead of one-step change. For example, when the nominal headway for trains is to be changed from 120sec to 240sec and the minimum resolution on the headway control variable, σ_{Hd}, is set at 30sec, the maximum number of step changes available to arrive at the subsequent nominal headway in a headway-sequence during the 'change-over' is given by:

$$\Psi = \frac{|Hd_E - Hd_S|}{\sigma_{Hd}} \tag{8.1}$$

$$= \frac{|120 - 240|}{30} = 4$$

where Hd_E and Hd_S are the headways before and after the 'change-over' respectively. From the viewpoint of application, given Ψ, the possible sets of headway-sequence to reach Hd_S can be expressed by:

$$\ell = 2^{\Psi - 1} \tag{8.2}$$

As shown in Fig. 8.2, there are: $2^{4-1} = 8$ possible sets of headway-sequence to reach the nominal headway of 240sec under the given operational constraints. Each sequence produces different peak kVA demand on the supply system. Details of the approach used to represent the 8 sets of headway-sequence in the *HTRC* are given in Section 8.3.

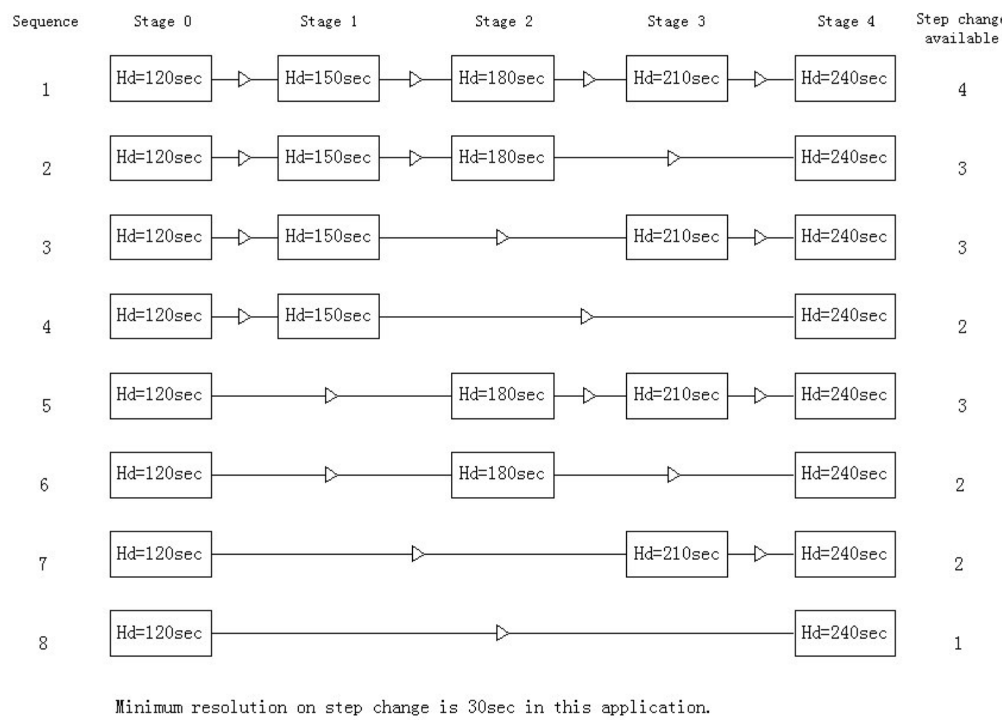


Fig. 8.2 Possible sets of headway-sequence to change the headway from 120sec to 240sec when $\sigma_{Hd} = 30\,\text{sec}$

8.2.3 Solution space

Given the required headways before and after the 'change-over', a smaller step σ_{Hd} results in a greater number of options of headway-sequences to reach Hd_S. In other words, the solution space of the headway control grows as the step size reduces. However, very small headway step changes are not appropriate in reality because of the time concern. The proposed *CTC* accommodates a moderate resolution which is appropriate and feasible from the operational viewpoint.

8.3 Event-based traffic flow model

A railway line is split into a number of regions in the *HTRC* and the headway control can be carried out by changing the headway in stages when trains move from one region to another. An event-based traffic flow model is adopted to represent the headway change with multi-train operation and the optimal headway-sequence is obtained by means of DP for real time applications. In the model, each event allows one step-change of headway for trains in regions and a state represents the corresponding service headway in each region of *HTRC*. In order to demonstrate the headway change with an event-based model, a 3-region railway line, as given in Fig. 8.3, is set up as follows. Regions 3 and 1 are the first and last regions of the line in this example.

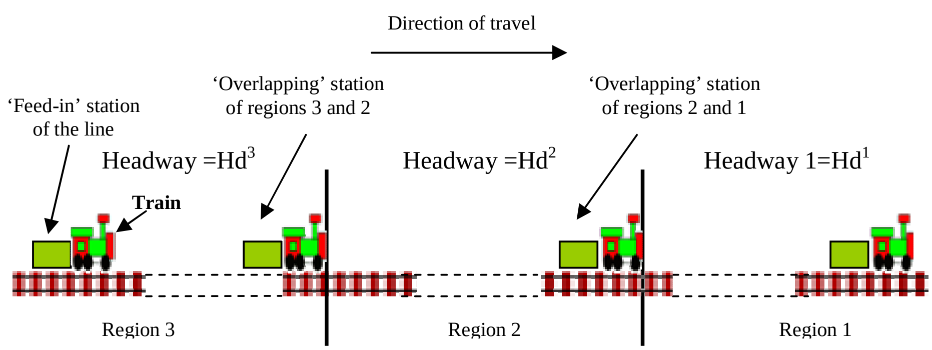


Fig. 8.3 A 3-region railway line of *HTRC*

8.3.1 State definition

In event-based modelling, a state is defined as a set of service headways allocated to regions on a line. In other words, each region has its own operating service headway at a particular stage. From Fig. 8.3, a state can be represented as:

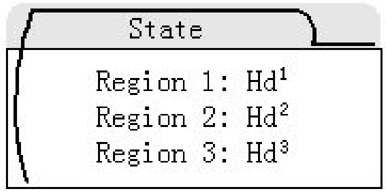


Fig. 8.4 State formulation

For the sake of simplicity, $[Hd^3, Hd^2, Hd^1]_k$ is used to describe a state in stage k on a 3-region railway line in the subsequent descriptions.

8.3.2 State evolution

The headway in a region is forwarded to the subsequent region in each state evolution, and the headway allocated in regions 3 and 2 in stage 'k' are directed to regions 2 and 1 in stage 'k+1' respectively, in a 3-region railway line. The process of changing one level of nominal headway to another through a series of state evolutions is described as the 'regional-headway sequence' in this study. For the 8 possible headway-sequences in the example shown in Fig. 8.2, the state evolutions in each of the 8 regional headway-sequences are shown below.

Sequence 1: 120→150→180→210→240
$$[120,120,120]_1 \rightarrow [150,120,120]_2 \rightarrow [180,150,120]_3 \rightarrow [210,180,150]_4 \rightarrow [240,210,180]_5 \rightarrow [240,240,210]_6 \rightarrow [240,240,240]_7$$

Sequence 2: 120→150→180→240
$$[120,120,120]_1 \rightarrow [150,120,120]_2 \rightarrow [180,150,120]_3 \rightarrow [240,180,150]_4 \rightarrow [240,240,180]_5 \rightarrow [240,240,240]_6$$

Sequence 3: 120→150→210→240
$$[120,120,120]_1 \rightarrow [150,120,120]_2 \rightarrow [210,150,120]_3 \rightarrow [240,210,150]_4 \rightarrow [240,240,210]_5 \rightarrow [240,240,240]_6$$

Sequence 4: 120→150→240
$$[120,120,120]_1 \rightarrow [150,120,120]_2 \rightarrow [240,150,120]_3 \rightarrow [240,240,150]_4 \rightarrow [240,240,240]_5$$

Sequence 5: 120→180→210→240

$$[120,120,120]_1 \rightarrow [180,120,120]_2 \rightarrow [210,180,120]_3 \rightarrow [240,210,180]_4 \rightarrow [240,240,210]_5 \rightarrow [240,240,240]_6$$

Sequence 6: 120→180→240

$$[120,120,120]_1 \rightarrow [180,120,120]_2 \rightarrow [240,180,120]_3 \rightarrow [240,240,180]_4 \rightarrow [240,240,240]_5$$

Sequence 7: 120→210→240

$$[120,120,120]_1 \rightarrow [210,120,120]_2 \rightarrow [240,210,120]_3 \rightarrow [240,240,210]_4 \rightarrow [240,240,240]_5$$

Sequence 8: 120→240

$$[120,120,120]_1 \rightarrow [240,120,120]_2 \rightarrow [240,240,120]_3 \rightarrow [240,240,240]_4$$

However, with different arrival and departure times of trains at stations, the headway distributions of trains in a state may be different from the desired values in evolutions. Details are given in the following.

8.3.3 Implementation

From Fig. 8.5, stations '*x-t*' '*x-n*' and '*x-g*' are the three corresponding starting stations in regions 3, 2 and 1 respectively; while stations '*x-n*', '*x-g*' and '*x*' are the last stations in the three regions: 3, 2 and 1. Stations '*x-n*' and '*x-g*' are the two 'overlapping' stations of regions 3 and 2 and regions 2 and 1 respectively. The headway needs to be changed from Hd^{α} to Hd^{δ} by the sequence of $Hd^{\alpha} \rightarrow Hd^{\beta} \rightarrow Hd^{\delta}$ within a certain time period, and the line is fully occupied initially by the trains separated with a headway of Hd^{α}.

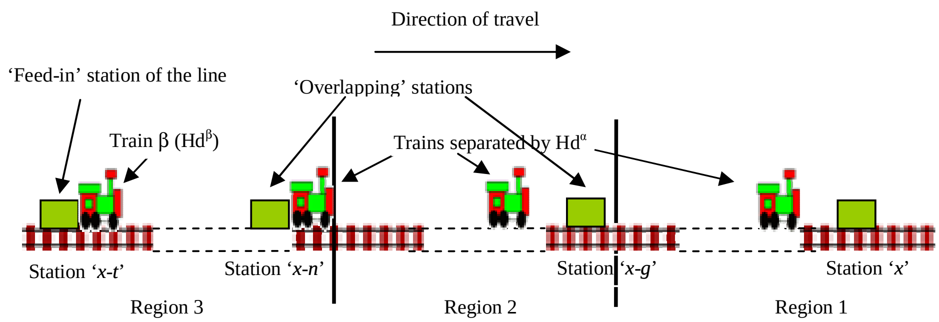


Fig. 8.5 State $[Hd^{\alpha}, Hd^{\alpha}, Hd^{\alpha}]_1$

Trains β is the first train carrying the headway of Hd^{β}, while train δ is the first train carrying the headway of Hd^{δ} behind train β. The arrival times of train β and train δ at stations '*x-t*' '*x-n*' and '*x-g*' are compared, and used to determine new state in evolution. To begin with, the headway of Hd^{β} is brought by train β at the 'feed-in' station '*x-t*' (i.e., the first station of the line) at time $t=i$, and the headway distributions (i.e., states) of trains in regions are described as $[Hd^{\alpha}, Hd^{\alpha}, Hd^{\alpha}]_1$ in this stage.

The headway of Hd^{δ} is then introduced to train δ at station '*x-t*'at $t=i+a$, when the train β reaches region 2 (i.e., station '*x-n*') as depicted in Fig. 8.6. Hd^{β} is fully spread over region 3 at this time instant (i.e., the trains behind train β and in front of train δ is operating with the headway of Hd^{β}). The arrival time of train β at region 2 implies the departure time of the train at station '*x-n*'. The state is changed to $[Hd^{\beta}, Hd^{\alpha}, Hd^{\alpha}]_2$ in this stage.

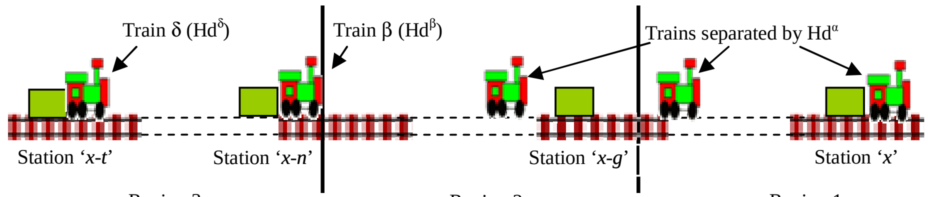


Fig. 8.6 State [Hd^{β}, Hd^{α}, $Hd^{\alpha}]_2$

As described in Section 8.3.2, [Hd^{δ}, Hd^{β}, $Hd^{\alpha}]_3$ is the desired state in the subsequent evolution. However, as shown in Fig. 8.7, it is possible that the train β has not yet arrived at region 1 (i.e., station '*x-g*') when the train δ carrying the headway of Hd^{δ} reaches region 2 (i.e., station '*x-n*'). The headway of Hd^{β} is not fully occupied in region 2, and this traffic condition is identified as [Hd^{δ}, Hd^{α}, $Hd^{\alpha}]_3$ by taking the arrival time at region 2 as the reference to evolve new states. On the other hand, state [Hd^{β}, Hd^{β}, $Hd^{\alpha}]_3$ is also necessary when the arrival time of train β reaching region 1 is earlier than that of train δ in region 2. However, to ensure each headway-step is introduced to trains in region 3 at each evolution, [Hd^{β}, Hd^{β}, $Hd^{\alpha}]_3$ is not taken in this application.

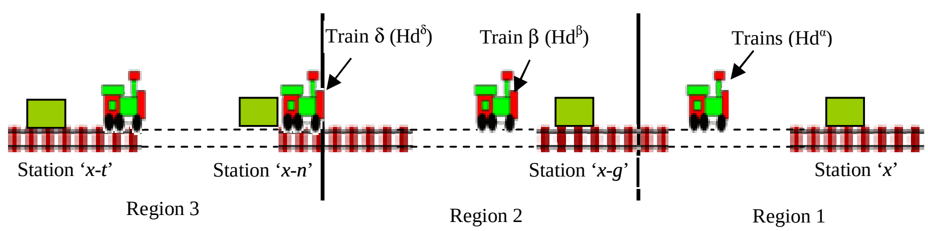


Fig. 8.7 State [Hd^{δ}, Hd^{α}, $Hd^{\alpha}]_3$

Once the train β reaches region 1, the state needs to be updated and described by [Hd^{δ}, Hd^{β}, $Hd^{\alpha}]_4$ as in Fig. 8.8. Arrival-time sequences of trains δ and β in regions 2 and 1 are the key parameters to determine the possible regional headways in evolutions.

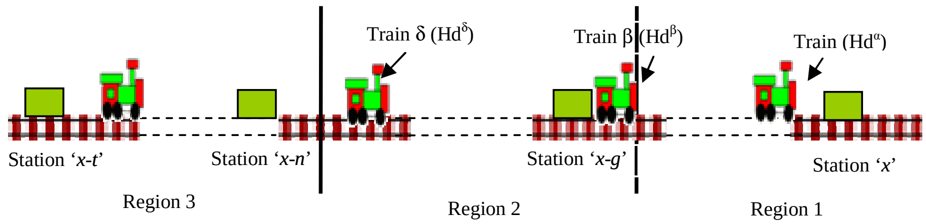


Fig. 8.8 State [Hd^{δ}, Hd^{β}, Hd^{α}]$_4$

Similarly, as shown in Figs 8.9 and 8.10, states [Hd^{δ}, Hd^{δ}, Hd^{α}]$_5$ and [Hd^{δ}, Hd^{δ}, Hd^{β}]$_6$ can be deduced in stages 5 and 6 when train δ reaches region 1. Finally, Fig. 8.11 reveals that the headway of Hd^{δ} (i.e., [Hd^{δ}, Hd^{δ}, Hd^{δ}]$_7$) is fully achieved on the line when the train δ arrives at station '*x*'.

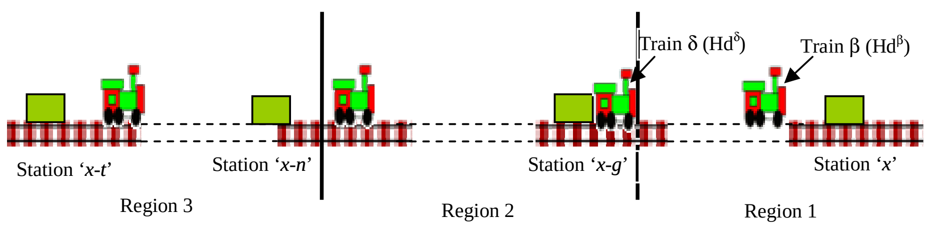


Fig. 8.9 State [Hd^{δ}, Hd^{δ}, Hd^{α}]$_5$

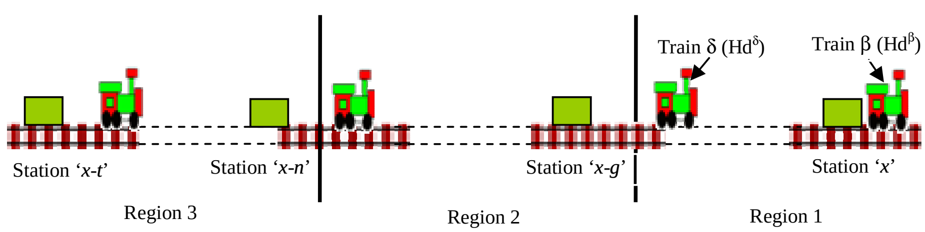


Fig. 8.10 State [Hd^{δ}, Hd^{δ}, Hd^{β}]$_6$

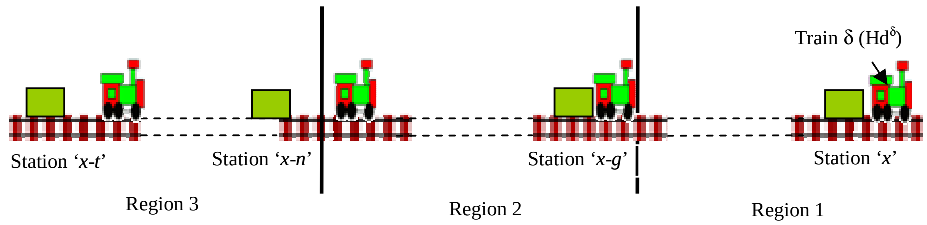


Fig. 8.11 State [Hd^{δ}, Hd^{δ}, Hd^{δ}]$_7$

To demonstrate the approach to determine each new state in a particular regional-headway sequence for updating one level of nominal value to another, two examples are given.

8.3.3.1 Example I

The traffic condition of the first example is illustrated in Fig. 8.12. A railway line is divided into 3 regions – 4 inter-stations runs (i.e., station 4 to 8) in region 2, 3 runs in regions 1 (i.e., stations 8 to 11) and 3 (i.e., stations 1 to 4). Region 3 is to referred as the first region of a line. The train dispatching sequence is in increasing order of train number (i.e., 0→1→2→3→…). The nominal dwell-times at stations are set at 30sec. The corresponding travelling time for each region of the 3-region railway line in *HTRC* is listed in Table 8.2. Minimum resolution on the headway step change σ_{Hd} is set at 30sec and the headway of 120sec is to be changed to 240sec.

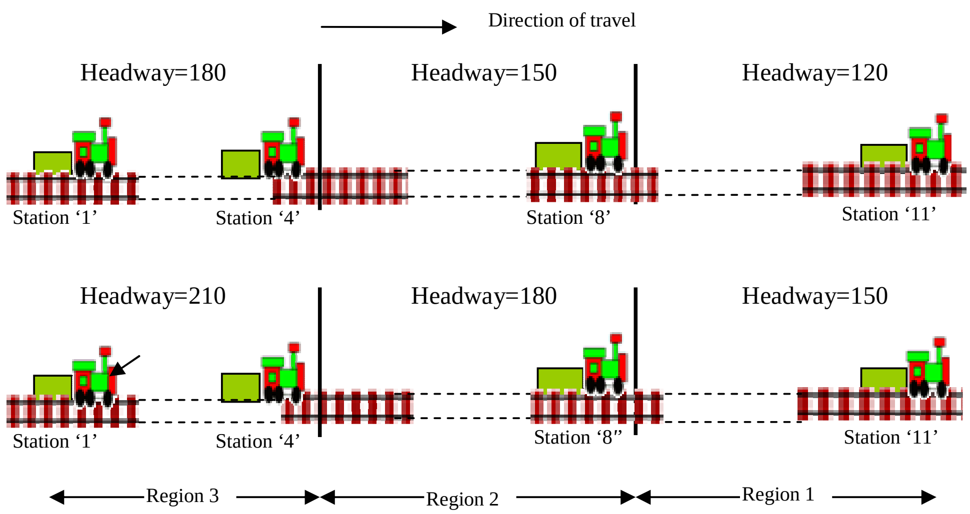


Fig. 8.12 Traffic conditions in a 3-region railway line

Table 8.2 – Travelling time of train in each region

Region	Stations	Number of inter-station runs in regions	*Travelling time (sec)
3	1→4	3	352
2	4→8	4	441
1	8→11	3	381

*The stated travelling time comprises a series of successive dwell-times and run-times of a train in a region.

The process to update the headway in stages with the regional-headway sequence 3 stated in Section 8.3.2 are described as follows:

Step 1:

Suppose train '0' departs at station 1 when time t=0, and the headway of 120sec is changed to 150sec on train '15'. The departure times of the train '15' at stations 1, 4, 8 and 11 are calculated as follows:

$$DT_1^{15} = 120 \times 15 + (150 - 120) = 1830\,\text{sec}$$

$$DT_4^{15} = 1830 + 352 = 2182\,\text{sec}$$

$$DT_8^{15} = 2182 + 441 = 2623\,\text{sec}$$

$$DT_{11}^{15} = 2623 + 381 = 3004\,\text{sec}$$

where $DT_x^{\ y}$ represents the departure time of y^{th} train at x^{th} station. At t=1830sec, the line is fully occupied by trains with a headway of 120sec and the headway distributions of trains are described as [120,120,120]sec in the initial stage (i.e., stage 1).

To change the headway from 150sec to 210sec at station 1, trains '17' and '18' are taken into account, where

$$DT_1^{17} = 1830 + 150 \times 2 = 2130 \sec$$

$$DT_1^{18} = 1830 + 150 \times 3 = 2280 \sec$$

From Fig. 8.13, DT_4^{15} is larger than DT_1^{17} but smaller than DT_1^{18} when the headway of 150sec is introduced to trains '17' and '18'. The headway of 210sec should not be given to train '17' at station 1 because train '15' carrying the headway of 150sec does not reach region 2. However, train '15' has already departed station 4 at t=2182, and the headway for train '18' is therefore changed to 210sec. A new DT_1^{18} is defined as:

$$DT_1^{18} = 1830 + 150 \times 2 + 210 = 2340 \sec$$

At t=2340sec, the headway distributions become [150,120,120]sec in stage 2.

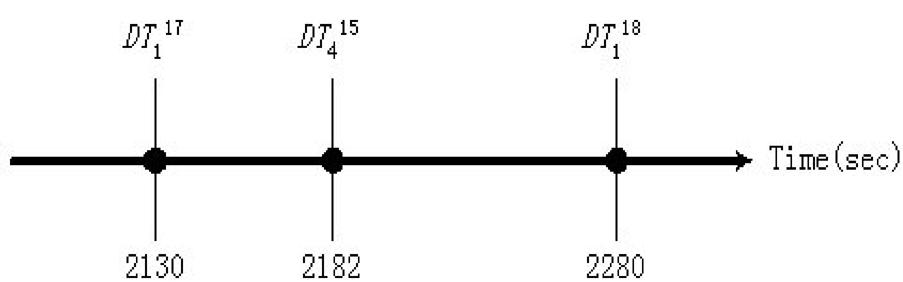


Fig. 8.13 Departure time sequence of trains 15, 17 and 18 at stations

Step 2:

Given the departure time of train '18' at station 1 and the travelling time in the successive inter-region runs (i.e., Table 8.2), the departure time of train '18' at stations 4, 8 and 11 are obtained.

$$DT_4^{18} = 2340 + 352 = 2692\,\text{sec}$$

$$DT_8^{18} = 2692 + 441 = 3133\,\text{sec}$$

$$DT_{11}^{18} = 3133 + 381 = 3514\,\text{sec}$$

To decide if train '19' or '20' should carry the change of headway from 210sec to 240sec at station 1, the preliminary values of DT_1^{19} and DT_1^{20} are calculated first.

$$DT_1^{19} = 2340 + 210 = 2550\,\text{sec}$$

$$DT_1^{20} = 2340 + 210 \times 2 = 2760\,\text{sec}$$

Fig. 8.14 shows that DT_4^{18} (i.e., 2692) is larger than DT_8^{15} (i.e.,2623). It thus implies that train '18' reaches region 2 at a later time than the train '15' arriving at region 1. In other words, the headway of 210sec is not completely achieved in region 3 at *t*=2623sec when train '15' arrives at station 8. The headway distributions become [150,150,120]sec but not the desired [210,150,120]sec. To ensure the headway-step change on each evolution in region 3, the departure time of train '18' at station 4 is taken as the reference in this example. The headway distributions become [210,150,120]sec when train '18' arrives at region 2 at t=2692sec in stage 3.

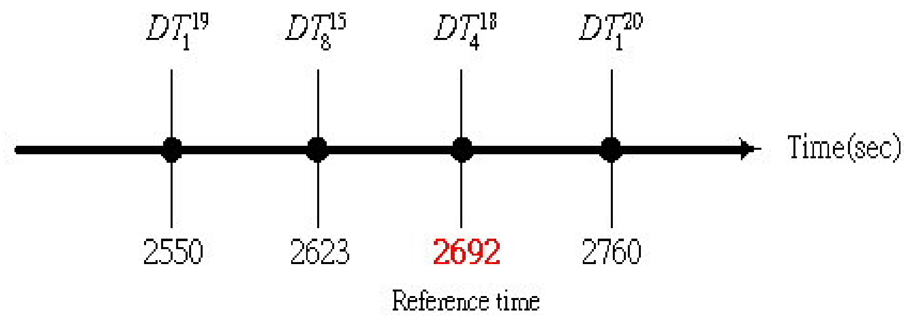


Fig. 8.14 Departure time sequence of trains 15, 18, 19 and 20 at stations

Similar to the approach used in step 1 for the decision of the headway sequence of train, the headway of 240sec is introduced to train '20' and its departure time at station 1 can be defined as:

$$DT_1^{20} = 2340 + 210 + 240 = 2790\,\text{sec}$$

Step 3:

Given the departure time of train '20' from station 1 at t=2790sec, train '20' departure times for stations 4, 8 and 11 are obtained:

$$DT_4^{20} = 2790 + 352 = 3142\,\text{sec}$$

$$DT_8^{20} = 3142 + 441 = 3583\,\text{sec}$$

$$DT_{11}^{20} = 3583 + 381 = 3964\,\text{sec}$$

Since $DT_4^{20} > DT_8^{18} > DT_{11}^{15}$ as depicted in Fig. 8.15, trains '15' and '18' have already departed from stations 11 and 8 respectively when train '20' arrives at station 4 at t=3142sec. The headway distributions become [240,210,150]sec in stage 4.

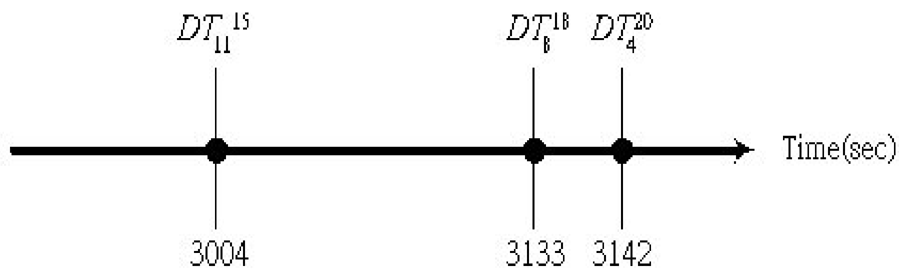


Fig. 8.15 Departure time sequence of trains 15, 18, and 20 at stations

Step 4:

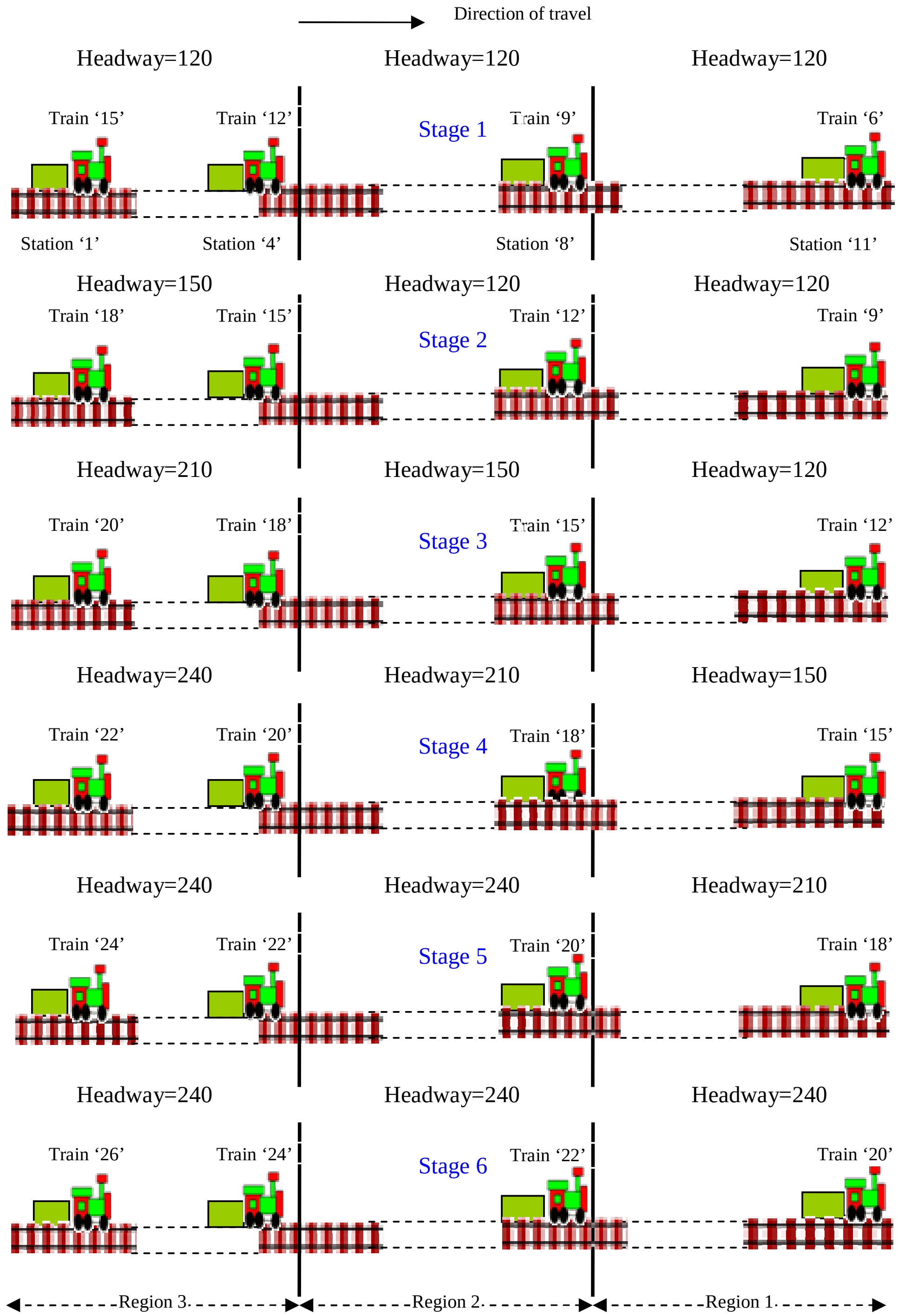
When train '20' reaches station 8 at *t*=3583sec, the first train carrying the headway of 210sec (i.e., train '18') has already arrived at station 11. The headway distribution is thereby described as [240,240,210]sec in stage 5.

Step 5:

The nominal headway of 240sec for trains is fully updated from 120sec at *t*=3964sec when train '20' has already arrived at station 11. The headway distributions hence finally become [240,240,240]sec in stage 6. The length of time period of changing the headway of $Hd_E = 120\,\text{sec}$ to $Hd_S = 240\,\text{sec}$, $TL_{Hd^E}^{Hd^s}$, is:

$$TL_{Hd^E}^{Hd^s} = 3964 - 1830 = 2134\,\text{sec}$$

The corresponding traffic conditions described under the given headway sequence of 120→150→210→240 in this 3-region railway line is illustrated in Fig. 8.16a and 8.16b.



* Trains depart from stations at different times in each stage.

Fig. 8.16a Regional-headway sequence in regions of *HTRC*

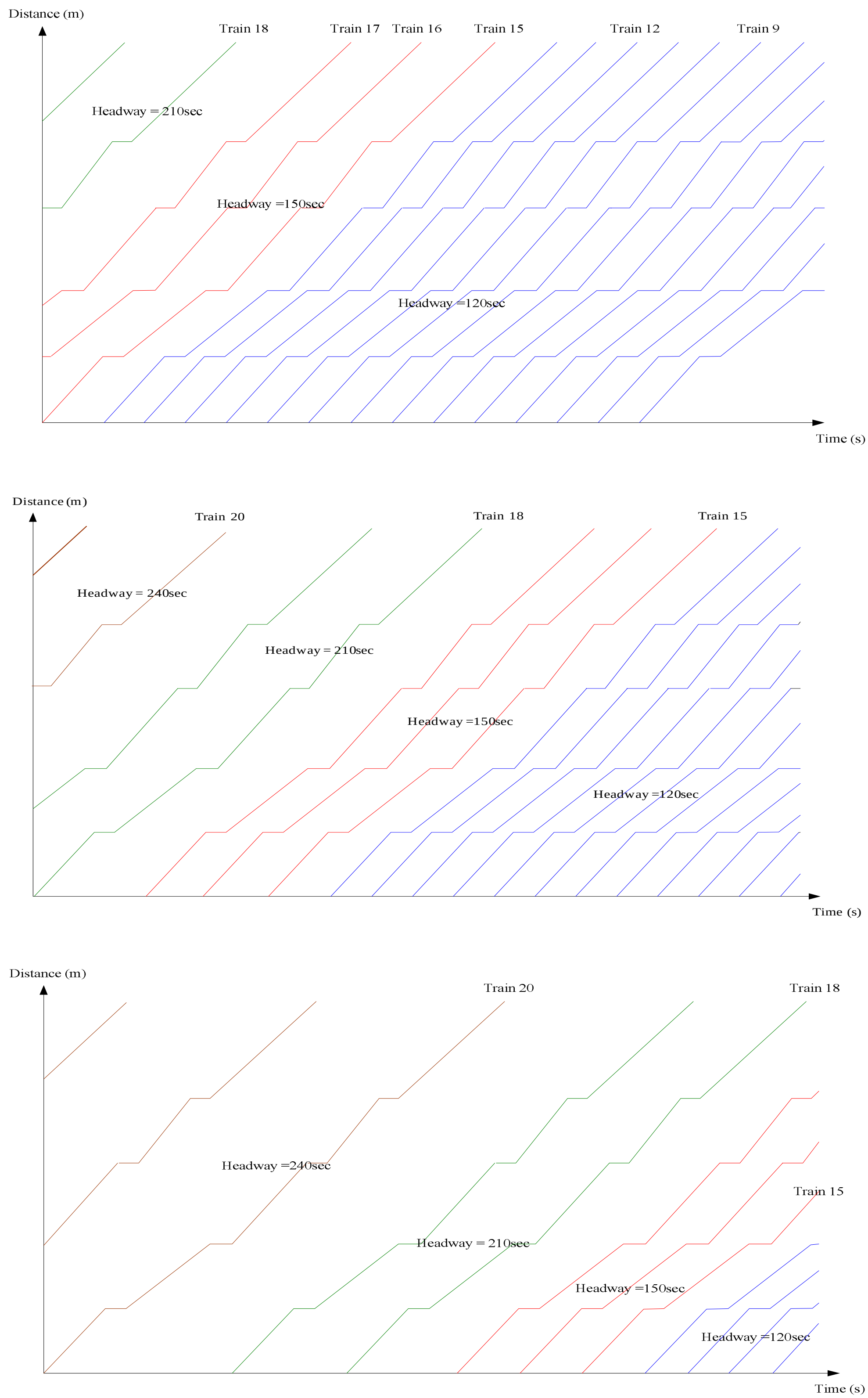


Fig.16b Time and distance profile of the headway sequence of

120sec→150sec→240sec

8.3.3.2 Example II

In some traffic conditions, the two corresponding headways in regions 3 and 2 may not be sequentially forwarded to the subsequent regions 2 and 1 in each stage transition. The headway distributions in a state of the sequence are not the same as the desired distributions. For example, with the sequence of 120sec→150sec→180sec→240sec, the following time schedule for trains is achieved across the 3-region line. The traffic conditions given in Example I remain.

Step 1:

The time schedule for trains updating the headway of [120,120,120] in stage 1 to [150,120,120]sec in stage 2 is similar to example I.

Step 2:

With the given sequence of 150sec→180sec in region 3 and the travelling times in successive regions, the departure time of train '18' at station 1, 4, 8 and 11 are devised as,

$$DT_1^{18} = 1830 + 150 \times 2 + 180 = 2310\,\text{sec}$$

$$DT_4^{18} = 2662\,\text{sec}\,;\;\; DT_8^{18} = 3103\text{sec} \;\;\text{and}\;\; DT_{11}^{18} = 3484\,\text{sec}$$

To change the headway from 180sec to 240sec at station 1, trains '19' and '20' are taken into account,

$$DT_1^{19} = 2310 + 180 = 2490\,\text{sec}$$

$$DT_1^{20} = 2310 + 180 \times 2 = 2670\,\text{sec}$$

Since $DT_1^{19} < DT_4^{18} < DT_1^{20}$, a new DT_1^{20} is calculated:

$$DT_1^{20} = 2310 + 180 + 240 = 2730\,\text{sec}$$

At t=2730sec, the headway distributions become [180,150,120]sec in stage 3.

Step 3:

From the departure time of train '20' at station 1 as t=2730sec, the departure time of train '20' at stations 4, 8 and 20 are determined.

$$DT_4^{20} = 2730 + 352 = 3082\,\text{sec}$$

$$DT_8^{20} = 3082 + 441 = 3523\,\text{sec}$$

$$DT_{11}^{20} = 3523 + 381 = 3904\,\text{sec}$$

From Fig. 8.17, $DT_8^{18} > DT_4^{20} > DT_{11}^{15}$ and this departure time sequence of trains at stations represents that the headway of 150sec has already been spread over region 1 and the headway of 180sec has not yet fully covered region 2 when train '20' arrives at station 4. The headway distributions are therefore described as [240,150,150]sec instead of the desired state [240,180,150]sec in stage 4.

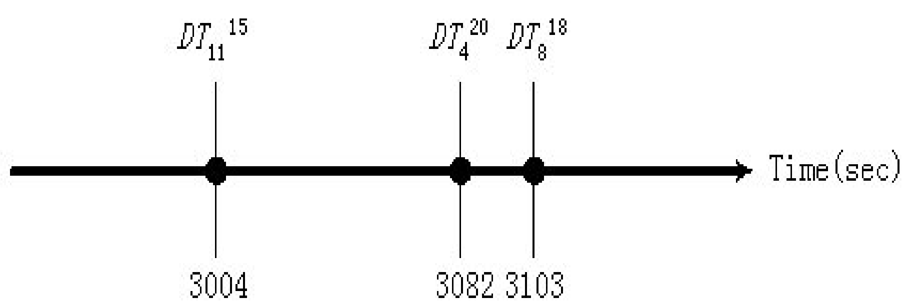


Fig. 8.17 Departure time sequence of trains 15, 18, and 20 at stations with the sequence of 120sec→150sec→180sec→240sec

Step 4:

Given $DT_{11}^{18} = 3484\,\text{sec}$ and $DT_8^{20} = 3523\,\text{sec}$, the headway distributions are changed to [240,240,180]sec and [240,240,240]sec at t=3523sec and 3904sec respectively. The overall regional-headway sequence is therefore calculated as:

$$[120,120,120] \to [150,120,120] \to [180,150,120] \to [240,150,150] \to [240,240,180]$$
$$\to [240,240,240]$$

8.3.4 Time schedules

With the travelling time in the 3-region railway line, the headway sequences for trains to update the nominal headway from 120sec to 240sec (see Section 8.3.2) are given in Table 8.3.

Table 8.3 – Time schedule for trains to change the headway from 120sec to 240sec using 8 sets of headway-sequence along a 3-region railway line

		Departure time at station '1'	Region 3→2	Region 2→1	Region 1→0
Train number	Headway (sec)	Sequence 1: 120sec→150sec→180sec→210sec→240sec			
5	120	600	952	1393	1774
15	120→150	1830	2182	2623	3004
18	150→180	2310	2662	3103	3484
20	180→210	2700	3052	3493	3874
22	210→240	3150	3502	3943	4324
		Sequence 2: 120sec→150sec→180sec→240sec			
		Time schedule for trains to change the headway of 120sec to 180sec through 150sec is similar to sequence 1			
20	180→240	2730	3082	3523	3904
		Sequence 3: 120sec→150sec→210sec→240sec			
		Time schedule for trains to change the headway of 120sec to 150sec is similar to sequence 1			
18	150→210	2340	2692	3133	3514
20	210→240	2790	3142	3583	3964
		Sequence 4: 120sec→150sec→240sec			
		Time schedule for trains to change the headway of 120sec to 150sec is similar to sequence 1			
18	150→240	2370	2722	3163	3544
		Sequence 5: 120sec→180sec→210sec→240sec			
15	120→180	1860	2212	2653	3034
17	180→210	2250	2602	3043	3424
19	210→240	2700	3052	3493	3874

To be continued

		Sequence 6: 120sec→180sec→240sec			
		Time schedule for trains to change the headway of 120sec to 180sec is similar to sequence 5			
17	180→240	2280	2632	3073	3454
		Sequence 7: 120sec→210sec→240sec			
15	120→210	1890	2242	2683	3064
17	210→240	2340	2692	3133	3514
		Sequence 8: 120sec→240sec			
15	120→240	1920	2272	2713	3094

1. Stations 4 and 8 are the overlapping stations of regions 3 and 2 and regions 2 and 1 respectively in this application.
2. Trains terminate at station 11 of region 1.
3. Headway is initially updated on train '15' in all sequences.

An overview of the traffic flow model to represent the headway change as stated in Table 8.3 is illustrated in Fig. 8.18. The operating headway for each region in a line is changed in stages under the corresponding headway-sequence.

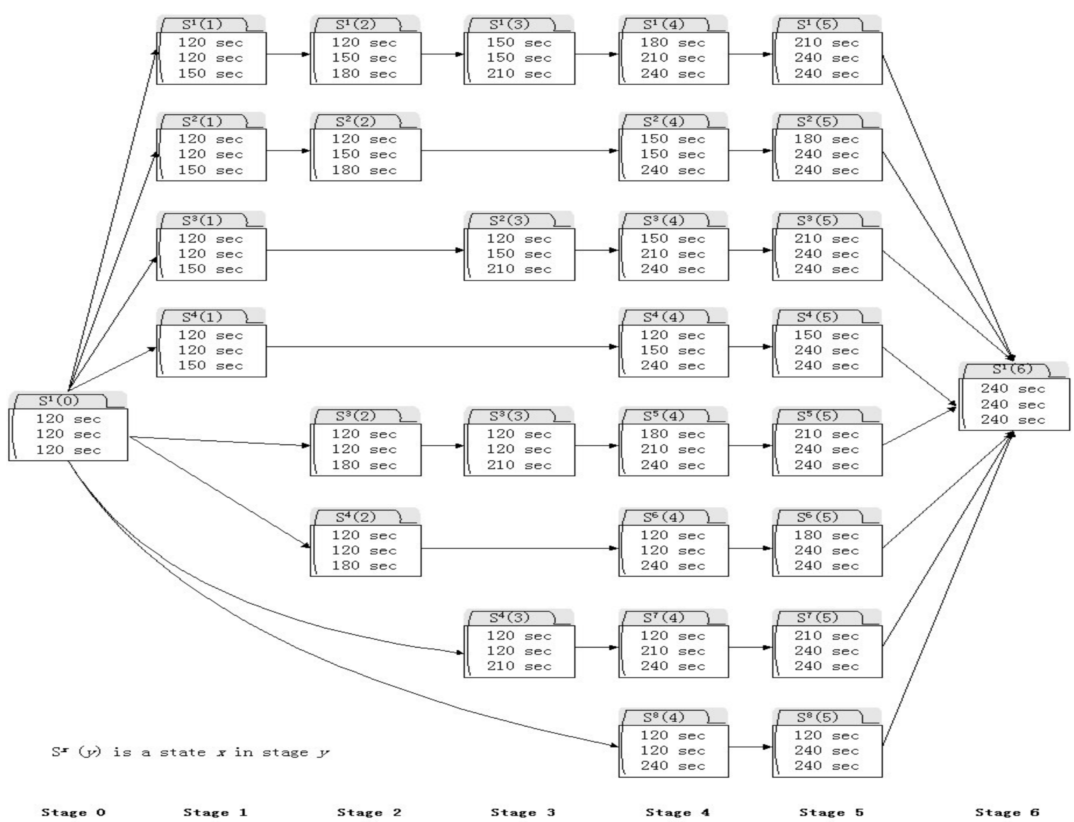


Fig. 8.18 Traffic flow model to represent the change of nominal headway from 120sec to 240sec with $\sigma_{Hd} = 30\sec$

8.4 Simplified traffic flow model

Fig. 8.18 shows that all the possible paths (i.e., headway-sequences) for reaching the headway of 240sec from 120sec. Each path has different numbers of step changes and various size of headway step. For example, the step changes of headway are 30sec, 60sec and 30sec in the sequence of 120sec→150sec→210sec→240sec; whilst they are 60sec and 60sec in the sequence of 120sec→180sec→240sec. With different levels and numbers of step changes of headway in the 8 sequences, the total number of states are different in the sequences. To enable a better pictorial view of the multi-train operation in successive regions, some 'dummy' states are introduced.

Fig. 8.19 shows the introduction of 'dummy' states in 3 out of the 8 sequences in a 3-region railway line. $S^3(2)$ and $S^2(3)$ are the two corresponding 'dummy' states in stage 2 and 3 of sequence 2 and 3, and the headway distribution of $S^3(2)$ is identical to that of $S^3(1)$ while $S^2(3)$ is a duplicate of $S^2(2)$, to provide the same number of stage transformations in the given 3 regional-headway sequences.

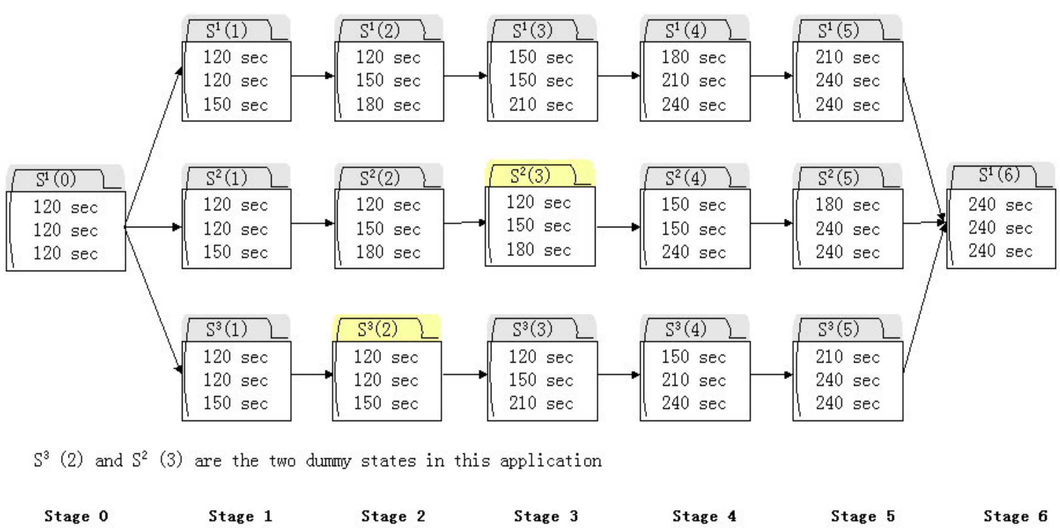


Fig. 8.19 Dummy state representation

Even though the complexity of the traffic flow model can be reduced substantially by adopting a moderate value for the headway steps at the system level of control, state simplification is still desired to reduce the computational effort and storage memory. Similar to the traffic model employed in ***RTC***, the number of states in each stage can be minimised by state grouping. At the system level of control, the objective of state grouping is to combine some of the states into a single state at a particular stage when they have the same service headway of trains in the corresponding regions. Fig. 8.20 depicts the possible state reduction for the given 3 regional-headway sequences. States $S^1(1)$, $S^2(1)$, and $S^3(1)$ are merged together in stage 1; $S^1(2)$ and $S^2(2)$ are combined into a new state $S^1(2)$ in stage 2 while $S^1(5)$ and $S^3(5)$ are grouped into a new state, $S^1(5)$ in the simplified traffic flow model. Fig. 8.21 shows an overview of the simplified traffic flow model.

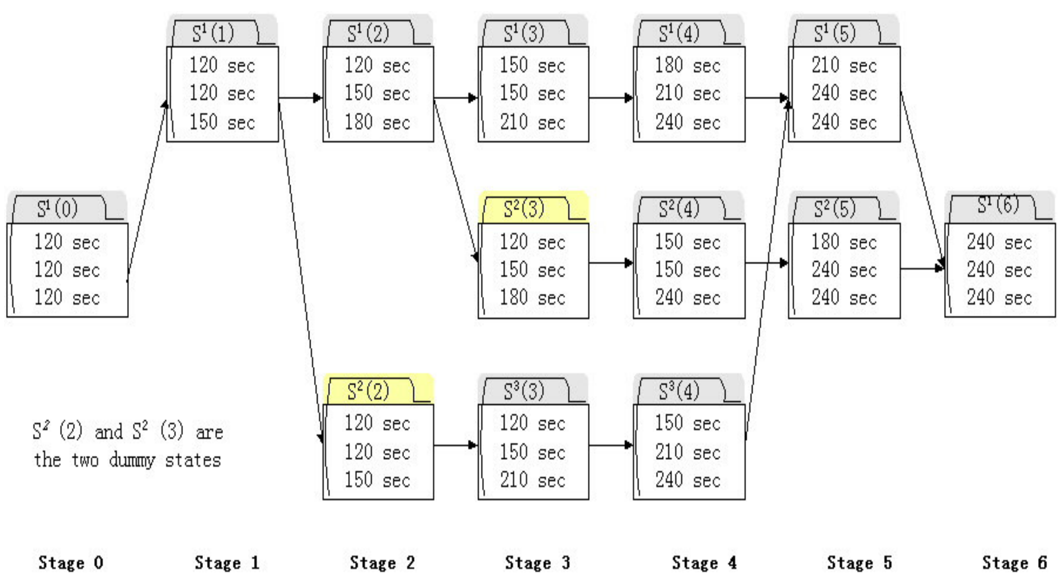


Fig. 8.20 State grouping

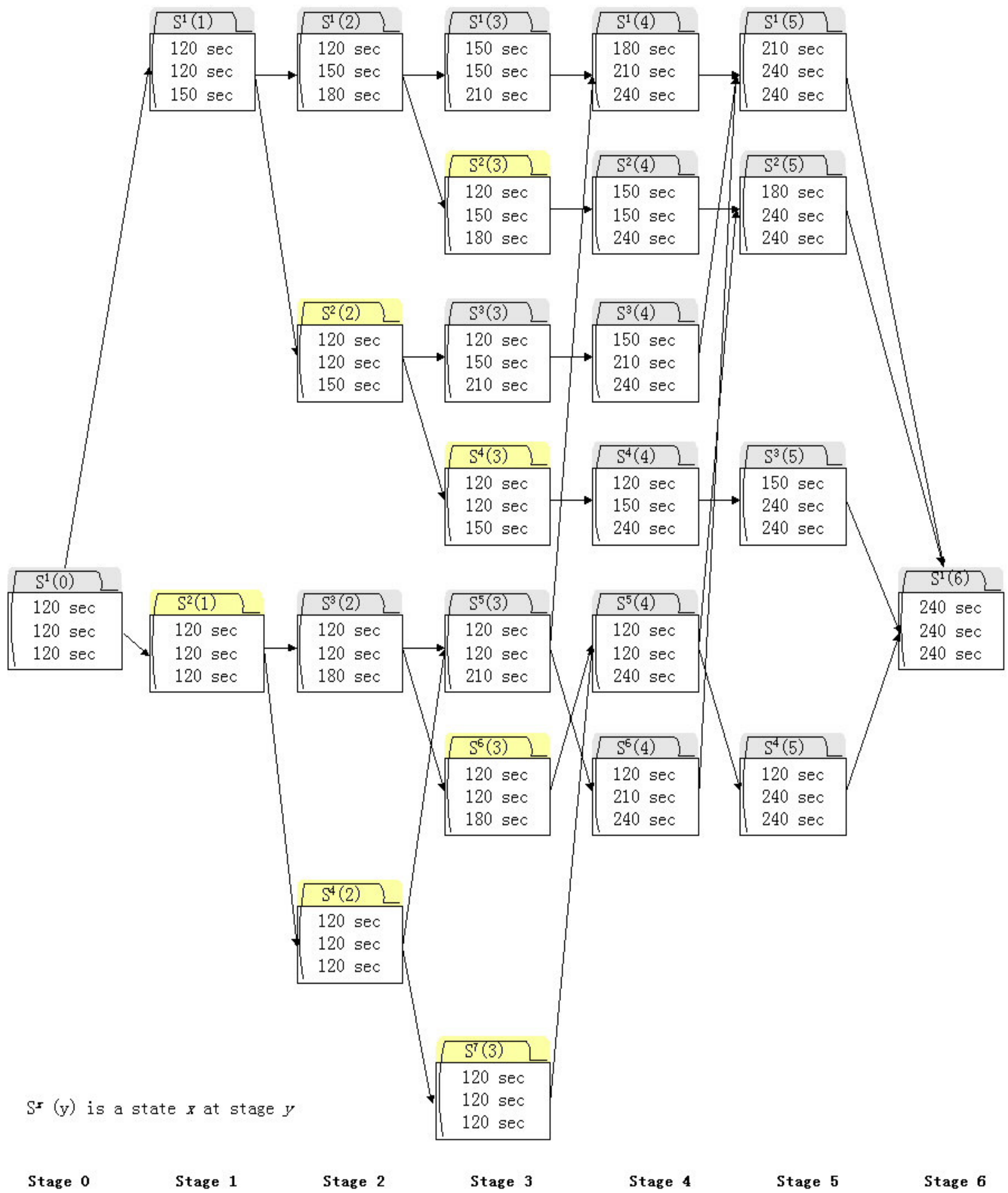


Fig. 8.21 Simplified traffic flow model in a 3-region railway line

8.5 Cost function

To determine the additional level of kVA demand on the power supply system with respect to the specific threshold on each state transition, the following cost function is adopted.

$$Cost = 0 \qquad \text{when} \quad PPD_k \le P_t \ or \ S_j(k) = dummy\ state$$

$$Cost = \frac{PPD_k - P_t}{P_t} \ \text{when} \quad PPD_k > P_t \tag{8.3}$$

where PPD_k is the actual peak power demand on the supply system in stage *k*, and P_t is the kVA threshold imposed by the power utility. $S_j(k)$ is the dummy state *i* in stage *k*. With Eqn 8.3, the function will produce a zero value in stage *k* when $PPD_k \le P_t$. A lower cost implies a lower additional kVA demand in a stage.

8.6 Optimisation with DP

With reference to Fig. 8.21, the power demand required for changing the headway from one state to another between every two stages for the example in Section 8.3 is obtained by a railway power system simulator and is shown in Table 8.4. Suppose the threshold imposed by the power supply utility is set at 2500kVA, the cost of reaching one state can be calculated by Eqn. 8.3. For example, the cost of reaching state $S^1(1)$ is defined as:

$$Cost = \frac{2676 - 2500}{2500} = 0.07$$

Details of the cost of reaching each state in Fig. 8.21 are given in Table 8.4.

Table 8.4 – Cost value of peak power demand at states for the 8 sets of regional-headway sequence

State	120→150→180 →210→240	kVA	Cost	State	120→150→180 →240	kVA	Cost
$S^1(0)$	[120,120,120]	3224	0.29	$S^1(0)$	[120,120,120]	3224	0.29
$S^1(1)$	[150,120,120]	2676	0.07	$S^1(1)$	[150,120,120]	2676	0.07
$S^1(2)$	[180,150,120]	2510	0.004	$S^1(2)$	[180,150,120]	2510	0.004
$S^1(3)$	[210,150,150]*	2038	0	$S^2(3)$	[180,150,120]**	-	0
$S^1(4)$	[240,210,180]	2086	0	$S^2(4)$	[240,150,150]*	2000	0
$S^1(5)$	[240,240,210]	1870	0	$S^2(5)$	[240,240,180]	1762	0
$S^1(6)$	[240,240,240]	1746	0	$S^1(6)$	[240,240,240]	1746	0
State	120→150→210 →240	kVA	Cost	State	120→150→240	kVA	Cost
$S^1(0)$	[120,120,120]	3224	0.29	$S^1(0)$	[120,120,120]	3224	0.29
$S^1(1)$	[150,120,120]	2676	0.07	$S^1(1)$	[150,120,120]	2676	0.07
$S^2(2)$	[150,120,120]**	-	0	$S^2(2)$	[150,120,120]**	-	0
$S^3(3)$	[210,150,120]	2510	0.004	$S^4(3)$	[150,120,120]**	-	0
$S^3(4)$	[240,210,150]	2137	0	$S^4(4)$	[240,150,120]	2581	0.032
$S^1(5)$	[240,240,210]	1863	0	$S^3(5)$	[240,240,150]	1898	0
$S^1(6)$	[240,240,240]	1746	0	$S^1(6)$	[240,240,240]	1746	0
State	120→180→210 →240	kVA	Cost	State	120→180→240	kVA	Cost
$S^1(0)$	[120,120,120]	3224	0.29	$S^1(0)$	[120,120,120]	3224	0.29
$S^2(1)$	[120,120,120]**	-	0	$S^2(1)$	[120,120,120]**	-	0
$S^3(2)$	[180,120,120]	2694	0.078	$S^3(2)$	[180,120,120]	2694	0.078
$S^5(3)$	[210,120,120]*	2630	0.052	$S^6(3)$	[180,120,120]**	-	0
$S^1(4)$	[240,210,180]	2086	0	$S^5(4)$	[240,120,120]*	2193	0
$S^1(5)$	[240,240,210]	1870	0	$S^2(5)$	[240,240,180]	1800	0
$S^1(6)$	[240,240,240]	1746	0	$S^1(6)$	[240,240,240]	1746	0
State	120→210→240	kVA	Cost	State	120→240	kVA	Cost
$S^1(0)$	[120,120,120]	3224	0.29	$S^1(0)$	[120,120,120]	3224	0.29
$S^2(1)$	[120,120,120]**	-	0	$S^2(1)$	[120,120,120]**	-	0
$S^4(2)$	[120,120,120]**	-	0	$S^4(2)$	[120,120,120]**	-	0
$S^5(3)$	[210,120,120]	2902	0.161	$S^7(3)$	[120,120,120]**	-	0
$S^6(4)$	[240,210,120]	2624	0.05	$S^5(4)$	[240,120,120]	2842	0.137
$S^1(5)$	[240,240,210]	1857	0	$S^4(5)$	[240,240,120]	2298	0
$S^1(6)$	[240,240,240]	1746	0	$S^1(6)$	[240,240,240]	1746	0

* The headway distribution over the line in a state is modified in accordance with the track layout.

** A zero cost value is given to dummy states.

To demonstrate how to obtain the optimal solution by DP under the traffic condition as depicted in Fig. 8.21, the process to reach states in each stage in the model is illustrated in Table 8.5.

Table 8.5 – Optimisation at each stage with respect to Fig. 8.20

Stage	State	Possible cost of reaching the corresponding states	Optimal path of reaching the corresponding states
1	$S^1(1)$ $S^2(1)$	$0.29+0.07=0.36$ $0.29+0=0.29$	$S^1(0)\rightarrow S^1(1)$ $S^1(0)\rightarrow S^1(2)$
2	$S^1(2)$ $S^2(2)$ $S^3(2)$ $S^4(2)$	$0.36+0.004=0.364$ $0.36+0=0.36$ $0.29+0.078=0.368$ $0.29+0=0.29$	$S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^4(2)$
3	$S^1(3)$ $S^2(3)$ $S^3(3)$ $S^4(3)$ $S^5(3)$ $S^6(3)$ $S^7(3)$	$0.364+0=0.364$ – $0.364+0=0.364$ – $0.36+0.004=0.364$ – $0.36+0=0.36$ – $0.368+0.052=0.42$ $0.29+0.161=0.451$ $0.368+0=0.368$ – $0.29+0=0.29$ –	$S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^1(3)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^2(3)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^3(3)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^4(3)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)\rightarrow S^5(3)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)\rightarrow S^6(3)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^4(2)\rightarrow S^7(3)$
4	$S^1(4)$ $S^2(4)$ $S^3(4)$ $S^4(4)$ $S^5(4)$ $S^6(4)$	$0.364+0=0.364$ $0.42+0=0.42$ $0.364+0=0.364$ – $0.364+0=0.364$ – $0.36+0.032=0.392$ – $0.42+0.05=0.47$ – $0.368+0=0.368$ $0.29+0.137=0.427$	$S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^1(3)\rightarrow S^1(4)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^2(3)\rightarrow S^2(4)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^3(3)\rightarrow S^3(4)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^4(3)\rightarrow S^4(4)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)\rightarrow S^6(3)\rightarrow S^5(4)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)\rightarrow S^5(3)\rightarrow S^6(4)$
5	$S^1(5)$ $S^2(5)$ $S^3(5)$ $S^4(5)$	$0.364+0=0.364$ $0.364+0=0.364$ $0.47+0=0.47$ – $0.364+0=0.364$ $0.368+0=0.368$ $0.392+0=0.392$ – $0.368+0=0.368$ –	$S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^1(3)\rightarrow S^1(4)\rightarrow S^1(5)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^3(3)\rightarrow S^3(4)\rightarrow S^1(5)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^2(3)\rightarrow S^2(4)\rightarrow S^2(5)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^4(3)\rightarrow S^4(4)\rightarrow S^3(5)$ $S^1(0)\rightarrow S^2(1)\rightarrow S^3(2)\rightarrow S^6(3)\rightarrow S^5(4)\rightarrow S^4(5)$
6	$S^1(6)$	$0.364+0=0.364$ $0.364+0=0.364$ $0.364+0=0.364$ $0.392+0=0.392$ $0.368+0=0.368$	$S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^1(3)\rightarrow S^1(4)\rightarrow S^1(5)\rightarrow S^1(6)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^2(3)\rightarrow S^2(4)\rightarrow S^2(5)\rightarrow S^1(6)$ $S^1(0)\rightarrow S^1(1)\rightarrow S^2(2)\rightarrow S^3(3)\rightarrow S^3(4)\rightarrow S^1(5)\rightarrow S^1(6)$

With reference to Table 8.5, three different sequences in this application arrive at $S^1(6)$ with the same cost of 0.364. To determine which sequence is preferred for service regulation, the total transition times for the 3 sequences $S^1(0)$ to $S^1(6)$ are devised as below:

Sequence I: $S^1(0)\rightarrow S^1(1)\rightarrow S^1(2)\rightarrow S^1(3)\rightarrow S^1(4)\rightarrow S^1(5)\rightarrow S^1(6)$,

$$TL_{Hd^E}^{Hd^s} = 4324-1830=2494\,\text{sec}$$

Sequence II: $S^1(0) \to S^1(1) \to S^1(2) \to S^2(3) \to S^2(4) \to S^2(5) \to S^1(6)$,

$$TL_{Hd^E}^{Hd^s} = 3904 - 1830 = 2074\,\text{sec}$$

Sequence III: $S^1(0) \to S^1(1) \to S^2(2) \to S^3(3) \to S^3(4) \to S^1(5) \to S^1(6)$

$$TL_{Hd^E}^{Hd^s} = 3964 - 1830 = 2134\,\text{sec}$$

From the operational point of view, sequence II is optimal because the headway change takes the least time to complete. To reflect the importance of the time required to change one headway level to another, the following simple rule is adopted.

If *two or more optimal headway-sequences are attained with DP*

Then *choose the one with the shorter updating time period*

8.7 Software implementation

To carry out train operation for headway regulation at the system level of control in *HTRC*, two operational modules are integrated into the *CTC* and they are:

1. An event-based traffic flow module
2. An optimal headway-sequence generation module.

With the traffic flow module, the maximum number of step changes of headway is first computed for the given operational constraints of σ_{Hd}, Hd_E and Hd_S, followed by the introduction of dummy states, if possible. A simplified traffic flow model can then be attained by state grouping.

Once the traffic flow model is established, the optimal headway-sequence generator starts to seek for the appropriate number of steps to change the headway and the corresponding time required. This module determines a set of costs to reach the states in each stage with the possible headway step(s). The optimal solution and the corresponding control actions in successive stage transformations are attained by DP. The structure of the system controller is illustrated in Fig. 8.22.

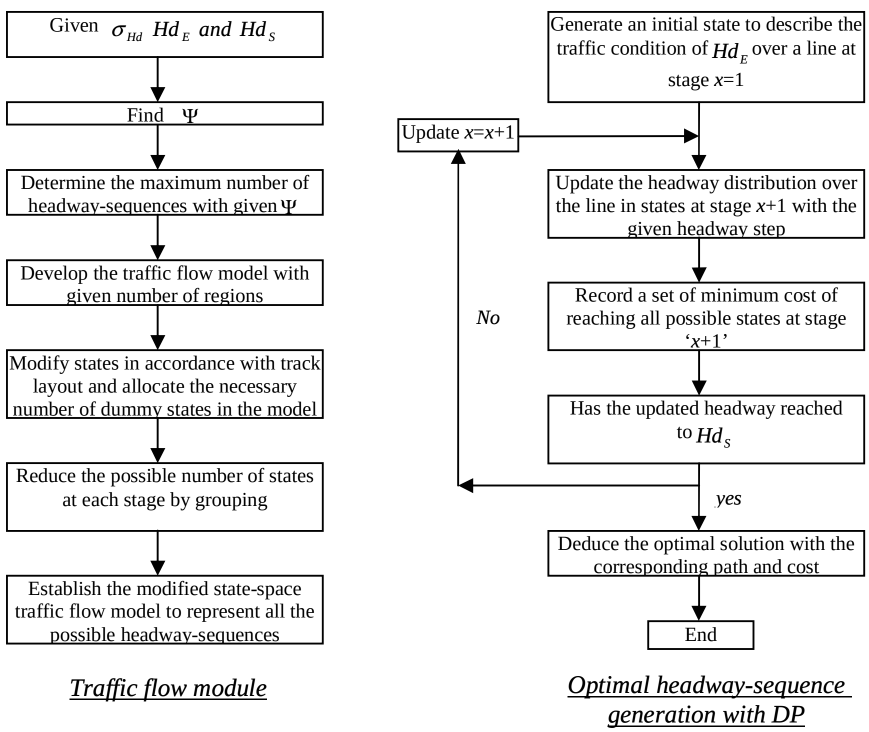


Fig. 8.22 Structure of the *CTC*

8.8 Practical considerations

In practice, the C*TC* can be integrated into the central control centre. The C*TC* determines the necessary service headways for the *RTC*s in accordance with the timetable. With the given headway in a certain period of time, the *RTC* then provides an appropriate set of dwell-times and run-times for trains and forwards the recommended control actions to the on-board *TBC*s. The control actions requested by the C*TC* can be supervised by the ATP to ensure train operation safety.

CTC requires three sets of data to devise the necessary headway instructions:

1. System-related data – number of feeder substations (i.e., the number of regions in *HTRC*); location of the 'overlapping' station for each region.
2. Operation-dependent data – the travelling times in each region; additional kVA demand of each step change of headway; and the specific kVA threshold imposed by the power utility.
3. Control variable – resolution of step-change of headway under given Hd_E and Hd_S.

At the system level of control, the complexity of the traffic flow model varies as a function of σ_{Hd} for given Hd_E and Hd_S. A small step-change of headway is not appropriate in reality because of the time concern. Although the solution space of the headway control is relatively small, a fast microprocessor is still preferred to carry out the multi-train operation for real-time applications. Similar to the *TBC* and *RTC*, a duplicate and hot standby hardware and software are highly preferred to maintain the availability and reliability of the system level control.

Chapter 9
Performance Analysis of Central Train Controller

This chapter presents the results regarding the feasibility and performance of the central train controller (*CTC*). With different track topologies and operational schedules, the central train controller is capable of providing the appropriate set of step changes of headway to reduce the peak power demand. Optimality and flexibility of the control system are the two key performance indicators.

9.1 Simulation setup

To investigate the *CTC*'s flexibility and performance, three case studies are examined in this chapter. With the aid of the event-based traffic flow model, various track topologies and operational schedules are set up in the studies, for 3-region railway line with a total track length of over 10 kilometres.

The first study investigates the train service regulation from rush hour service to off-peak service, where the nominal headway is changed from 120sec to 240sec. The number of stations in each region and the corresponding inter-region runs over a line are illustrated in Table 9.1. The kVA threshold imposed by the power utility and the minimum resolution on the headway control variable, σ_{Hd} , are set at 2100kVA and 30sec respectively. The objective of this study is to verify the capability of the controller in providing a feasible solution for daily train headway regulation.

Table 9.1 – Traffic conditions in a 3-region railway line of *HTRC*

Region	Stations	Number of inter-station runs in regions	*Travelling time (sec)
3	1→4	3	352
2	4→7	3	382
1	7→11	4	440

*The travelling time comprises a series of dwell-times and run-times of a train in a region.

The second study looks into the impact of the kVA threshold level on the decision regarding the headway-sequence to minimise the peak kVA demand on the supply system, under the same traffic condition as stated in study 1. Two kVA threshold levels, 2300kVA and 1800kVA, are used for comparison in the test.

To further illustrate the feasibility of headway control in railway operations, a 3-region railway line, in which the number of stations on each of the 3 regions and their corresponding inter-region travelling times are shown in Table 9.2, is set up for analysis in the last study. The nominal headway is changed to 150sec from 300sec, with a power threshold of 1500kVA and the minimum resolution of headway step remains 30sec. The objective of the study is to investigate whether the controller is capable of defining the appropriate control actions for trains with different change-over conditions.

To determine the cost of reaching every state in each stage within the traffic flow model in each of the 3 studies, the peak power demand of train operation with the specific headway distribution attained in each state was determined by a multi-train simulator.

Table 9.2 – Travelling time of a train in each region

Region	Stations	Number of inter-station runs in regions	*Travelling time (sec)
3	1→4	3	352
2	4→8	4	441
1	8→11	3	381

*The travelling time comprises a series of dwell-times and run-times of a train in a region.

9.2 Results and discussions

9.2.1 Case study I

With the traffic conditions shown in Table 9.1 and the operational constraints of σ_{Hd}, Hd_E and Hd_S, the traffic flow model for headway regulation in a 3-region railway line is shown in Fig. 9.1. Details of the possible sequences to change the headway from 120 to 240sec are given in Table B1 of Appendix B. Within the model, 8 dummy states are introduced and no state reduction at the intermediate stages is achieved in this application. The number of states evolved in stages 3 and 4 is higher, as shown in Table 9.3, when compared with that in Section 8.3. It clearly pinpoints that the complexity of the traffic flow model is different from one track layout to another because of different inter-region runs over a line.

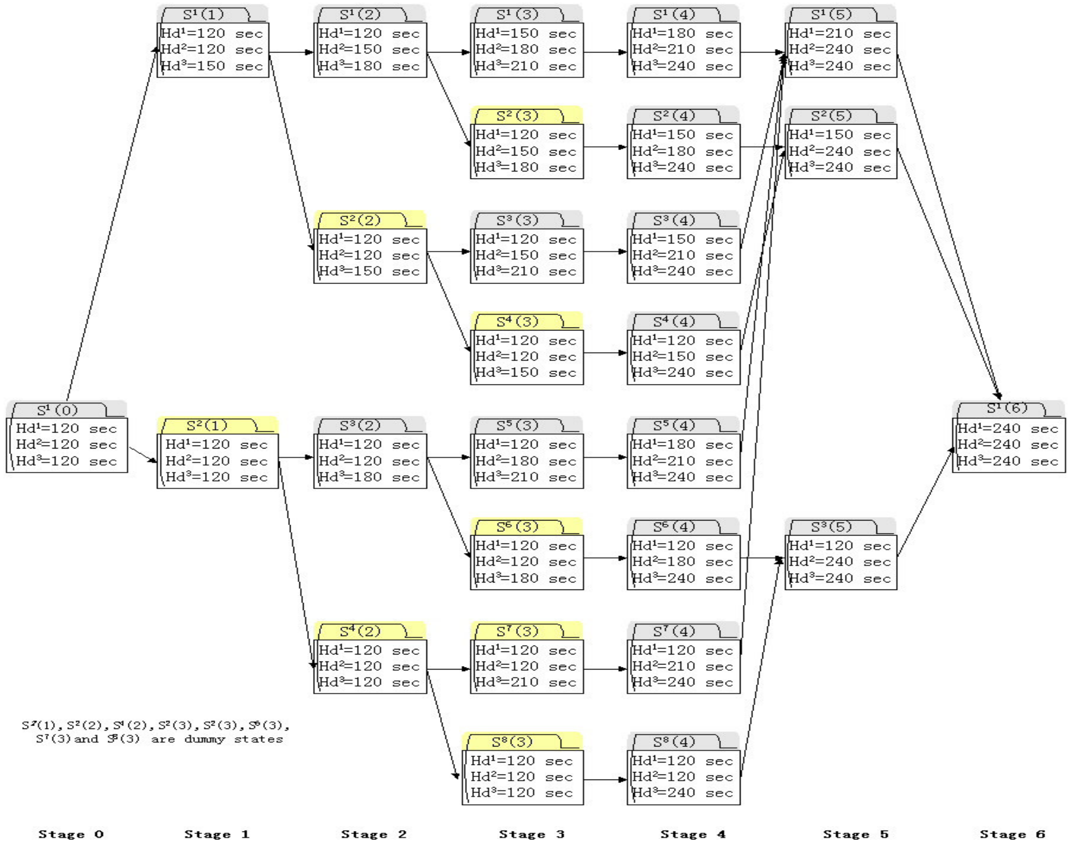


Fig. 9.1 Traffic flow model to represent the change of headway from 120sec to 240sec

Table 9.3 – Number of states at stages

Case I*							
	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Number of states	1	2	4	7	6	4	1
Case II**							
	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Number of states	1	2	4	8	8	3	1

* The numbers of inter-station runs on the 3 regions are 3, 4 and 3 respectively and the model is described in Fig. 8.14.
** The numbers of inter-station runs on the 3 regions are 3, 3 and 4 respectively.

With reference to Table 9.4, the optimal cost of 0.461 is obtained, in which the regional-headway sequence of $S^1(0)\rightarrow S^2(1)\rightarrow S^4(2)\rightarrow S^7(3)\rightarrow S^7(4)\rightarrow S^1(5)\rightarrow S^1(6)$ is deduced in this application. Nevertheless, states $S^2(1)$, $S^4(2)$ and $S^7(3)$ are the dummy states and the sequence [120,120,120]→[120,120,240]→[120,240,240]→

[240,240,240] is therefore the ultimate outcome and the time required to change the headway between the two nominal levels is

$$TL_{Hd^E}^{Hd^s} = 3094 - 1920 = 1174\,\text{sec}$$

where the corresponding time schedule for trains with the given headway-sequence is shown in Table B1 of Appendix B.

Table 9.4 – Optimal solution

Cost*	0.461
Optimal headway sequence	$S^1(0)\rightarrow S^2(1)\rightarrow S^4(2)\rightarrow S^7(3)\rightarrow S^7(4)\rightarrow S^1(5)\rightarrow S^1(6)$
Actual sequence	120sec→240sec

*Details on the set of costs to reach each state in the model and their corresponding additional kVA demand are illustrated in the Table B2 of Appendix B

9.2.2 Case study II

Given the two thresholds of 2300kVA and 1800kVA, Table 9.5a and 9.5b show that the costs of 0.245 and 0.996 are achieved respectively when the headway of 240sec is fully spread over the line, under the same traffic condition as given in the case study I. With 2300kVA threshold, two optimal sequences with the same cost are attained. By the rule stated in Section 8.3.6, the time required to change the headway from 120sec to 240sec is then the decisive factor. The two corresponding time required are calculated as:

Sequence I: $TL_{Hd^E}^{Hd^s} = 3874 - 1860 = 2014\,\text{sec}$

Sequence II: $TL_{Hd^E}^{Hd^s} = 3454 - 1860 = 1594\,\text{sec}$

Because of the shorter time required to change the headway, sequence II is chosen as the optimal sequence.

Table 9.5a – Optimal solution with 2300kVA threshold

Cost*	0.245
Optimal headway sequence	$S^1(0) \rightarrow S^2(1) \rightarrow S^3(2) \rightarrow S^5(3) \rightarrow S^5(4) \rightarrow S^1(5) \rightarrow S^1(6)$
Actual sequence	120sec→180sec→210sec→240sec
Optimal headway sequence	$S^1(0) \rightarrow S^2(1) \rightarrow S^3(2) \rightarrow S^6(3) \rightarrow S^6(4) \rightarrow S^4(5) \rightarrow S^1(6)$
Actual sequence	120sec→180sec→240sec

Table 9.5b – Optimal solution with 1800kVA threshold

Cost*	0.996
Optimal headway sequence	$S^1(0) \rightarrow S^2(1) \rightarrow S^4(2) \rightarrow S^7(3) \rightarrow S^7(4) \rightarrow S^1(5) \rightarrow S^1(6)$
Actual sequence	120sec→210sec→240sec

*Details on the set of costs to reach each state with different levels of kVA threshold and their corresponding peak kVA demand are presented in Tables B3 and B4 of Appendix B

On the other hand, when the peak power threshold is set at 1800kVA, the optimal headway-sequence, which requires 3514-1890=1624sec, is chosen. Since the threshold is set at a lower level, the difference between the peak kVA demand and the specified threshold is larger and hence the cost is increased, compared with 2300 kVA. Therefore, different headway-sequences are usually deduced for different levels of kVA threshold, even under the same system constraints and operational requirements.

9.2.3 Case study III

Given the track topology and inter-region travelling times of Table 9.2, the sequence to change the nominal headway from 300sec to 150sec in a 3-region railway line is calculated and shown in Fig. 9.2. Details of the possible sequences are given in Table B5 of Appendix B. Fig. 9.2 shows that more stages are in the model when the number of step changes of headway (i.e., $\Psi = 4$ in this application) to achieve the subsequent nominal headway increases. Indeed, the number of states in the model is of the order of $2^{\psi-1}$. Moreover, more dummy states are required to ensure that each sequence will have the same number of stage transformations for DP. Further, more states, having the same headway distribution, are to be grouped at the intermediate stages when higher resolution on headway steps is adopted.

Under peak demand threshold of 1500kVA, the optimal cost of 0.282 is attained in two different headway-sequences, as given in Table 9.6. A shorter time required (i.e., 1174sec) is achieved in the sequence of 300sec→150sec, and hence one-step change in headway is more appropriate in this study.

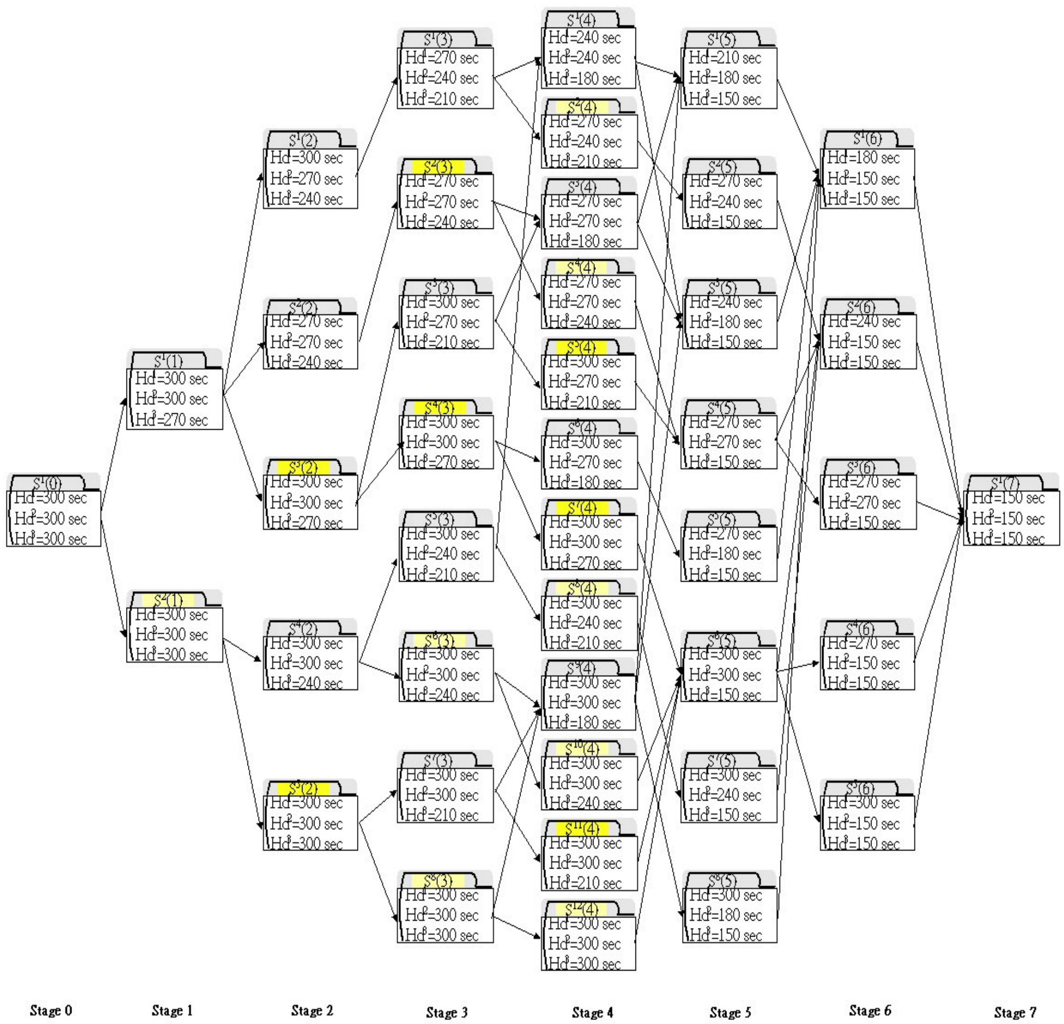


Fig. 9.2 Traffic flow model to represent the change of headway from 300sec to 150sec

Table 9.6 – Optimal solution with 1500kVA threshold

Cost*	0.282
Optimal headway sequence	$S^1(0) \rightarrow S^1(1) \rightarrow S^3(2) \rightarrow S^4(3) \rightarrow S^7(4) \rightarrow S^6(5) \rightarrow S^4(6) \rightarrow S^1(7)$
Actual sequence	300sec→270sec→150sec
Optimal headway sequence	$S^1(0) \rightarrow S^2(1) \rightarrow S^5(2) \rightarrow S^8(3) \rightarrow S^{12}(4) \rightarrow S^6(5) \rightarrow S^4(6) \rightarrow S^1(7)$
Actual sequence	300sec→150sec

*Details on the set of costs to reach each state in the model and their corresponding additional kVA demand are illustrated in the Table B6 of Appendix B

9.3 Concluding remarks

The mechanism of the central train controller to minimise the peak power demand on the supply system during ‘change-over’ between the two levels of nominal headway has been presented, where a dynamic programming technique has been adopted. The studies show that the controller is capable of providing the appropriate sequence of changes of headway. In addition, the level of kVA threshold is one of the major parameters in obtaining the solution, as different headway-sequences are usually deduced for different levels of kVA threshold.

The complexity and size of the problem depends on the number of regions on a railway line and the resolution of headway steps. Though a more precise train operation and control can be performed with a smaller headway step, the demand in terms of computational effort and memory storage inevitably inflates. Moreover, such small headway steps are not preferable in reality because of the time required on computation.

Chapter 10
Conclusions and Further Work

10.1 Conclusions

10.1.1 Achievements

A study to improve train operation and control for DC metro systems through advanced intelligent techniques and system design has been presented. Hierarchical train regulation and control (*HTRC*) has been adopted to enhance the capability of train control through 3-level of coordination, taking into account the consideration of traffic conditions, tariff and computational demands. Each level adjusts train operation within its own specific sphere in accordance with given operational constraints and requirements. The proposed *HTRC* system strikes an effective balance between the operational cost and the quality of train service for real-time applications.

Within the three levels of train control in *HTRC*, the central train controller (*CTC*) determines the appropriate sequence of headway-step changes in regions along a line to minimise the overall peak power demand on the power supply at the system level. Simulation results reveal that different optimal headway-sequences are obtained to meet the subsequent service demand under different headway transitions. Resolution of headway-step and the specific kVA threshold are the two key decisive factors. The main drawback is that the complexity of the event-based model to represent the headway sequence increases with decreasing size of headway-step. From the operational point of view, the complexity of the model can be limited with a moderate resolution, which is more appropriate in real life. A quick search in finding the optimal headway-sequence is achieved by dynamic programming (DP).

A regional train controller (*RTC*) has been developed to provide the optimal set of dwell-times and run-times of trains at regional level. Simulation results show that the *RTC* enables train coordination and reduces the energy demand placed on the supply system under a given level of change of headway from the nominal value. A higher efficiency and flexibility of train coordination can be achieved by run-time control as compared with that by dwell-time and mixed control. Further, inter-station run-time is usually longer than dwell time at stations and hence a larger control space is available.

As there are a number of regions on a line, exchange of arrival-times sets for trains at the 'overlapping station' between regions is required to enable the train coordination for recovery of train service from disturbance when time delays on trains at stations are substantial and/or the disturbed train is located near the end of a region. Results reveal that the disturbance is gradually reduced when the delayed train is going through successive regions. Successful train coordination among regions to recover the service is thereby achieved.

The complexity of the event-based model to represent dwell-time and run-time coordination escalates if the number of inter-station runs and/or the number of possible sets of control actions increase. Computational demand and memory storage are thus the two major performance indicators in this level of control. Complexity of the model in the *RTC* is more substantial compared with that of the

CTC because the exhaustive details of train description (i.e., dwell-time and run-time of each train) are required. The number of states at the immediate stage, however, can be significantly reduced by grouping. In addition, it is common to have separation of a few kilometres between regions in metro systems. When there are fewer stations in a region, the capability of dwell-time and run-time control for real-time railway applications is enhanced.

At train level of control, coast control is adopted to regulate the speed of a train during inter-station runs to meet the given run-time constraints imposed by the *RTC*. Since the details of train activities are required, a time-based simulation model is appropriate to carry out individual train movement to aid the train-based controller. The solution space at the train level of control varies with the inter-station distance and the minimum coasting speed to ensure sufficient momentum for trains to go on. With multiple coasting-point control, the solution space for the next coasting point varies with the location of the previous coasting point. In other words, the location of the first coasting point will affect that of the second and so on.

A number of studies has been undertaken to locate the necessary coasting point(s) with different traffic conditions and various searching methods. The studies show that the classical searching method provides a smaller average number of iterations in single coasting point control. However, when multiple coasting-points are adopted, heuristic approach is preferred. It has been proven that inter-station distance and track layout are the two major factors in choosing the appropriate searching method for coasting control. In addition, single coasting point control is more suitable in metro operations because there is not enough room for multiple coasting-point operation with the short inter-station distance.

The author studies the feasibility of applying *HTRC* on railway operation problem at three levels of control. *HTRC* ensures manageable multi-train control and provides appropriate control actions for train coordination at each level. The recommended control actions at the system, region and train levels enable a flexible and efficient train control. By streamlining computation demand, the computational effort on each level of train control is not excessive. *HTRC* therefore has the potential for real-time train control in metro operations. However, with *HTRC*, the functions of each controller must be well defined and specified to avoid overlaps or gaps in the control.

10.1.2 Discussions

Due to the need of a frequent train service in metro operations, the density and complexity of railway traffic increases. Even though the traffic can still be maintained with the predetermined schedule in existing railway systems, flexibility and efficiency of the train operation may not be attained at the highest extent. Further, with the expansion of train service demand, traditional approaches used in train coordination are confined from the operational viewpoint.

An attempt to improve train control with various intelligent approaches and modelling techniques applying on railway operation problems have been proven successful for service regulation in the thesis. Numerous systems (i.e., *CTC*, *RTC* and *TBC*) with different levels of functionality in *HTRC* are highlighted to devise the proper train instructions, which are safeguarded by the *ATP*, for real-time applications according to the time-varying service demand. The results reveal that a high capability of train control is achieved to maintain the service with the application of intelligent approaches. The proposed train controllers have therefore the potential to perform the functions of *ATS* and *ATO* in railway operations.

Further, a stringent real-time monitoring and control is desired in railway system because of the demanding safety standards. Traditional train control is usually determined by the operators' and drivers' knowledge and experience or uses some rigid control system rules to meet the specific traffic conditions, but not the current service demand. PID control is one of the options to regulate train speed in inter-stations, where the main drawbacks include poor ride comfort for passengers and higher maintenance cost of traction drives. As technology continues to advance, intelligent approaches to obtain a fully automated and driverless operation become feasible, where reliable and rigorous train coordination is further enhanced to provide optimal usage of the track. In addition, a strong commitment to maintain high levels of safety management can also be achieved because of the continuous supervision in traffic. However, when a significant saving on the train operational cost is not expected by the application of intelligent techniques, it may not be worthwhile to upgrade the existing control systems, considering the heavy capital and maintenance cost, which is another decisive factor.

10.2 Further works

From the operational point of view, the two traffic flow models adopted in *RTC* and *CTC* can be extended to represent the traffic on double-track lines. On a double-track line, trains run on one line with certain headway while the other line serves passengers with a different headway in opposite direction. Since a bi-directional traffic condition is presented on a double-track line, representation of the bi-directional traffic within the states in the two corresponding models is needed. For example, in the model of *CTC*, a state represents the corresponding service headway in each region of the line and each region contains one headway of ‘up’ line and the other of ‘down’ line. On a double-track line, the complexity of the problem in either regional or system level of control inevitably inflates and hence state reduction techniques to simplify the problem is the key concern for real-time railway applications.

At the train-based level, coast-control has been adopted for trade-off between run time and energy consumption of train operation. Energy saving is optimal in accordance with the specific inter-station run-time requirement imposed by *RTC*. The advantage of coast-control, however, is attained only in minimising energy consumption for individual train movement, but not in the coordination with the trains nearby (i.e., other trains in front and behind) to reduce peak demand on the supply system. To achieve higher efficiency in train operation, one viable option is to determine the coasting solution with coordination of the trains nearby. For instance, a train is recommended to coast when the trains nearby are operating in powering mode, and to motor and then coast again to meet the desired schedule when the trains nearby are no longer in powering mode. Peak demand on the supply system is reduced because of the avoidance of trains operating at powering mode simultaneously.

In reality, speed profiles of the trains in adjacent inter-station runs are recorded, and taken as the decisive constraint in finding the coasting solution in a particular run. Since details of the trains' operation nearby are known, the peak demand of train operation can be calculated under the specific coasting solution. A balance between energy consumption and peak demand on the supply system is the key concern in this application. Therefore, an advisory system to find out the appropriate coasting strategy in a run with the coordination of the trains nearby is worth investigating.

Regenerative braking is another alternative to reduce energy and power demand of train operation in railway applications. To ensure the energy recovery among motoring and braking trains, efficient train regulation is required. A high efficiency of energy recovery from train movement is also important. As compared with coast-control, a lower energy and peak power demand is the result of the application of regeneration, because the powering trains obtain energy from the braking trains nearby, whereas in coast-control, the energy for trains to accelerate is provided by the power supply system. However, the railway equipment has to be enhanced to cope with the reversal of current and rise of voltage in practice. Hence, the extents of energy saving of train operation by regeneration and its limitations on railway application are worth further study.

Appendix A

<u>Operation factors of the Nelder and Mead method</u>

The effects of the three operation factors of the Nelder & Mead method on the searching performance, in terms the average number of iterations and fitness (cost), are investigated here. There is no specific rule on assigning these three operation factors and their specific range are listed as follows: -

1. Reflection factor, $\alpha > 0$;
2. Expansion factor, $\gamma > 1$;
3. Contraction factor, $0 < \beta < 1$

In order to define an appropriate range of value of the operation factors for coasting points search, 4 sets of operation factors have been chosen at regular intervals as shown in Table A. Five tests have been carried out to reach the specific cost value in each case under the same track layout conditions as in Tables 5.1 and 5.2.

Table A Average number of iterations with different sets of operation factors

	Nelder and Mead											
	α	γ	β	α	γ	β	α	γ	β	α	γ	β
	0.2	1.2	0.2	0.4	1.4	0.4	0.6	1.6	0.6	0.8	1.8	0.8
Average number of iterations	10.3			11.3			8.7			12*		
Fitness	0.0098			0.016			0.02			0.018		

* A specific fitness value cannot be attained in some tests up to the maximum number of iterations.

The result shows that a lower cost and reasonable average number of iterations can be attained with smaller values on operation factors. Though a smaller average number of iterations can be reached with larger values of the operation factors, the possibility of finding the optimal solution is limited as the generated solution is more likely to be trapped out from the expected solution when the updated solution is close to the optimal point. Furthermore, the generated solution may be out of the boundary of the solution space if the three operation factors are set to a larger value. Hence, smaller values of operation factors are necessary to trade off between the number of iterations and the cost. As a result, the three operation factors α, γ and β are set at 0.2, 1.2 and 0.2 respectively

Appendix B

Study 1

Table B1 – Time schedule for trains to change the headway from 120sec to 240sec as depicted in Fig. 9.1

		Departure time at station '1'	Region 3→2	Region 2→1	Region 1→0
Train number	Headway (sec)	Sequence 1: 120→150→180→210→240			
5	120	600	952	1334	1774
15	120→150	1830	2182	2564	3004
18	150→180	2310	2662	3044	3484
20	180→210	2700	3052	3434	3874
22	210→240	3150	3502	3884	4324
		Sequence 2: 120→150→180→240 Time schedule for trains to change the headway from 120sec to 180sec through 150sec is similar to sequence 1			
20	180→240	2730	3082	3464	3904
		Sequence 3: 120→150→210→240 Time schedule for trains to change the headway from 120sec to 150sec is similar to sequence 1			
18	150→210	2340	2692	3074	3514
20	210→240	2790	3142	3524	3964
		Sequence 4: 120→150→240 Time schedule for trains to change the headway from 120sec to 150sec is similar to sequence 1			
18	150→240	2370	2722	3104	3544
		Sequence 5: 120→180→210→240			
15	120→180	1860	2212	2594	3034
17	180→210	2250	2602	2984	3424
19	210→240	2700	3052	3434	3874
		Sequence 6: 120→180→240 Time schedule for trains to change the headway from 120sec to 180sec is similar to sequence 5			
17	180→240	2280	2632	3014	3454
		Sequence 7: 120→210→240			
15	120→210	1890	2242	2624	3064
17	210→240	2340	2692	3074	3514
		Sequence 8: 120→240			
15	120→240	1920	2272	2654	3094

1. Stations 4 and 7 are the overlapping stations of regions 3 and 2 and regions 2 and 1 respectively in this application.
2. Trains terminate at station 11 of region 1.
3. Headway is initially changed on train '15' in all sequences.

Table B2 – A set of costs of reaching all states in each stage as shown in Fig. 9.1 with 2100 kVA threshold

Stage 0 to 1	Peak power demand	Additional kVA demand	Cost	Stage 1 to 2	Peak power demand	Additional kVA demand	Cost
$S^1(0) \rightarrow S^1(1)$	2800	700	0.333	$S^1(1) \rightarrow S^1(2)$	2307	207	0.099
$S^1(0) \rightarrow S^2(1)$*	-	-	0	$S^1(1) \rightarrow S^2(2)$*	-	-	0
				$S^2(1)$*$\rightarrow S^3(2)$	2377	277	0.132
				$S^2(1) \rightarrow S^4(2)$*	-	-	0
Stage 2 to 3	**Peak power demand**	**Additional kVA demand**	**Cost**	**Stage 3 to 4**	**Peak power demand**	**Additional kVA demand**	**Cost**
$S^1(2) \rightarrow S^1(3)$	2115	15	0.007	$S^1(3) \rightarrow S^1(4)$	1804	0	0
$S^1(2) \rightarrow S^2(3)$*	-	-	0	$S^2(3)$*$\rightarrow S^2(4)$	1785	0	0
$S^2(2)$*$\rightarrow S^3(3)$	2151	51	0.024	$S^3(3) \rightarrow S^3(4)$	1667	0	0
$S^2(2)$*$\rightarrow S^4(3)$*	-	-	0	$S^4(3)$*$\rightarrow S^4(4)$	2151	51	0.024
$S^3(2) \rightarrow S^5(3)$	2074	74	0.035	$S^5(3) \rightarrow S^5(4)$	1799	0	0
$S^3(2) \rightarrow S^6(3)$*	-	-	0	$S^6(3)$*$\rightarrow S^6(4)$	2196	96	0.046
$S^4(2) \rightarrow S^7(3)$*	2379	279	0.133	$S^7(3)$*$\rightarrow S^7(4)$	1725	0	0
$S^4(2) \rightarrow S^8(3)$*	-	-	0	$S^8(3)$*$\rightarrow S^8(4)$	2540	440	0.21
Stage 4 to 5	**Peak power demand**	**Additional kVA demand**	**Cost**	**Stage 5 to 6**	**Peak power demand**	**Additional kVA demand**	**Cost**
$S^1(4) \rightarrow S^1(5)$	1695	0	0	$S^1(5) \rightarrow S^1(6)$	2025	0	0
$S^2(4) \rightarrow S^2(5)$	1973	0	0	$S^2(5) \rightarrow S^1(6)$	2025	0	0
$S^3(4) \rightarrow S^1(5)$	1695	0	0	$S^3(5) \rightarrow S^1(6)$	2264	164	0.078
$S^4(4) \rightarrow S^2(5)$	1901	0	0				
$S^5(4) \rightarrow S^1(5)$	1695	0	0				
$S^6(4) \rightarrow S^2(5)$	1985	0	0				
$S^7(4) \rightarrow S^1(5)$	1704	0	0				
$S^8(4) \rightarrow S^3(5)$	2338	238	0.113				

1. The headway distribution over the line in a state is modified in accordance with the track layout.
2. ‘*’ represents a dummy state in which a zero cost value is given of reaching such state.

Study 2

Table B3 – A set of costs of reaching all states in each stage as shown in Fig. 9.1 with 2300 kVA threshold

Stage 0 to 1	Peak power demand	Additional kVA demand	Cost	Stage 1 to 2	Peak power demand	Additional kVA demand	Cost
$S^1(0) \rightarrow S^1(1)$	2800	500	0.217	$S^1(1) \rightarrow S^1(2)$	2307	7	0.003
$S^1(0) \rightarrow S^2(1)$*	-	-	0	$S^1(1) \rightarrow S^2(2)$*	-	-	0
				$S^2(1)$*$\rightarrow S^3(2)$	2377	77	0.033
				$S^2(1) \rightarrow S^4(2)$*	-	-	0
Stage 2 to 3	Peak power demand	Additional kVA demand	Cost	Stage 3 to 4	Peak power demand	Additional kVA demand	Cost
$S^1(2) \rightarrow S^1(3)$	2115	0	0	$S^1(3) \rightarrow S^1(4)$	1804	0	0
$S^1(2) \rightarrow S^2(3)$*	-	-	0	$S^2(3)$*$\rightarrow S^2(4)$	1785	0	0
$S^2(2)$*$\rightarrow S^3(3)$	2151	0	0	$S^3(3) \rightarrow S^3(4)$	1667	0	0
$S^2(2)$*$\rightarrow S^4(3)$*	-	-	0	$S^4(3)$*$\rightarrow S^4(4)$	2151	0	0
$S^3(2) \rightarrow S^5(3)$	2074	0	0	$S^5(3) \rightarrow S^5(4)$	1799	0	0
$S^3(2) \rightarrow S^6(3)$*	-	-	0	$S^6(3)$*$\rightarrow S^6(4)$	2196	0	0
$S^4(2) \rightarrow S^7(3)$*	2379	79	0.034	$S^7(3)$*$\rightarrow S^7(4)$	1725	0	0
$S^4(2) \rightarrow S^8(3)$*	-	-	0	$S^8(3)$*$\rightarrow S^8(4)$	2540	240	0.104
Stage 4 to 5	Peak power demand	Additional kVA demand	Cost	Stage 5 to 6	Peak power demand	Additional kVA demand	Cost
$S^1(4) \rightarrow S^1(5)$	1695	0	0	$S^1(5) \rightarrow S^1(6)$	2025	0	0
$S^2(4) \rightarrow S^2(5)$	1973	0	0	$S^2(5) \rightarrow S^1(6)$	2025	0	0
$S^3(4) \rightarrow S^1(5)$	1695	0	0	$S^3(5) \rightarrow S^1(6)$	2264	0	0
$S^4(4) \rightarrow S^2(5)$	1901	0	0				
$S^5(4) \rightarrow S^1(5)$	1695	0	0				
$S^6(4) \rightarrow S^2(5)$	1985	0	0				
$S^7(4) \rightarrow S^1(5)$	1704	0	0				
$S^8(4) \rightarrow S^3(5)$	2338	38	0.017				

1. The headway distribution over the line in a state is modified in accordance with the track layout.
2. '*' represents a dummy state in which a zero cost value is given of reaching such state.

Table B4 – A set of costs of reaching all states in each stage as shown in Fig. 9.1 with 1800 kVA threshold

Stage 0 to 1	Peak power demand	Additional kVA demand	Cost	Stage 1 to 2	Peak power demand	Additional kVA demand	Cost
$S^1(0)\rightarrow S^1(1)$	2800	1000	0.56	$S^1(1)\rightarrow S^1(2)$	2307	507	0.282
$S^1(0)\rightarrow S^2(1)$*	-	-	0	$S^1(1)\rightarrow S^2(2)$*	-	-	0
				$S^2(1)$*$\rightarrow S^3(2)$	2377	577	0.321
				$S^2(1)\rightarrow S^4(2)$*	-	-	0
Stage 2 to 3	**Peak power demand**	**Additional kVA demand**	**Cost**	**Stage 3 to 4**	**Peak power demand**	**Additional kVA demand**	**Cost**
$S^1(2)\rightarrow S^1(3)$	2115	315	0.175	$S^1(3)\rightarrow S^1(4)$	1804	4	0.002
$S^1(2)\rightarrow S^2(3)$*	-	-	0	$S^2(3)$*$\rightarrow S^2(4)$	1785	0	0
$S^2(2)$*$\rightarrow S^3(3)$	2151	351	0.195	$S^3(3)\rightarrow S^3(4)$	1667	0	0
$S^2(2)$*$\rightarrow S^4(3)$*	-	-	0	$S^4(3)$*$\rightarrow S^4(4)$	2151	351	0.195
$S^3(2)\rightarrow S^5(3)$	2074	274	0.152	$S^5(3)\rightarrow S^5(4)$	1799	0	0
$S^3(2)\rightarrow S^6(3)$*	-	-	0	$S^6(3)$*$\rightarrow S^6(4)$	2196	396	0.22
$S^4(2)\rightarrow S^7(3)$*	2379	579	0.322	$S^7(3)$*$\rightarrow S^7(4)$	1725	0	0
$S^4(2)\rightarrow S^8(3)$*	-	-	0	$S^8(3)$*$\rightarrow S^8(4)$	2540	740	0.411
Stage 4 to 5	**Peak power demand**	**Additional kVA demand**	**Cost**	**Stage 5 to 6**	**Peak power demand**	**Additional kVA demand**	**Cost**
$S^1(4)\rightarrow S^1(5)$	1695	0	0	$S^1(5)\rightarrow S^1(6)$	2025	225	0.125
$S^2(4)\rightarrow S^2(5)$	1973	173	0.096	$S^2(5)\rightarrow S^1(6)$	2025	225	0.125
$S^3(4)\rightarrow S^1(5)$	1695	0	0	$S^3(5)\rightarrow S^1(6)$	2264	464	0.258
$S^4(4)\rightarrow S^2(5)$	1901	101	0.056				
$S^5(4)\rightarrow S^1(5)$	1695	0	0				
$S^6(4)\rightarrow S^2(5)$	1985	185	0.103				
$S^7(4)\rightarrow S^1(5)$	1704	0	0				
$S^8(4)\rightarrow S^3(5)$	2338	538	0.3				

1. The headway distribution over the line in a state is modified in accordance with the track layout.
2. '*' represents a dummy state in which a zero cost value is given of reaching such state.

Study 3

Table B5 – Time schedule for trains to change the headway from 300sec to 150sec as depicted in Fig. 9.2

		Departure time at station '1'	Region 3→2	Region 2→1	Region 1→0
Train number	Headway (sec)	Sequence 1: 300→270→240→210→180→150			
5	300	1500	1852	2293	2674
10	300→270	2970	3322	3763	4144
12	270→240	3480	3832	4273	4654
14	240→210	3930	4282	4723	5104
16	210→180	4320	4672	5113	5494
19	180→150	4830	5182	5623	6004
		Sequence 2: 300→270→240→210→150 Time schedule for trains to change the headway from 300sec to 210sec through 270 and 240sec is similar to sequence 1			
16	210→150	4290	4642	5083	5464
		Sequence 3: 300→270→240→180→150 Time schedule for trains to change the headway from 300sec to 240sec through 270sec is similar to sequence 1			
14	240→180	3900	4252	4693	5074
17	180→150	4410	4762	5203	5584
		Sequence 4: 300→270→240→150 Time schedule for trains to change the headway from 300sec to 240sec through 270sec is similar to sequence 1			
14	240→150	3870	4222	4663	5044
		Sequence 5: 300→270→210→180→150 Time schedule for trains to change the headway from 300sec to 270sec is similar to sequence 1			
12	270→210	3450	3802	4243	4624
14	210→180	3840	4192	4633	5014
17	180→150	4350	4702	5143	5524
		Sequence 6: 300→270→210→150 Time schedule for trains to change the headway from 300sec to 210sec through 270sec is similar to sequence 5			
14	210→150	3810	4162	4603	4984
		Sequence 7: 300→270→180→150 Time schedule for trains to change the headway from 300sec to 270sec is similar to sequence 1			
12	270→180	3420	3772	4213	4594
15	180→150	3930	4282	4723	5104
		Sequence 8: 300→270→150 Time schedule for trains to change the headway from 300sec to 270sec is similar to sequence 1			
12	270→150	3390	3742	4183	4564
		Sequence 9: 300→240→210→180→150			
10	300→240	2940	3293	3733	4114
12	240→210	3390	3742	4183	4564

14	210→180	3780	4132	4573	4954
17	180→150	4290	4642	5083	5464
		Sequence 10: 300→240→210→150 Time schedule for trains to change the headway from 300sec to 210sec through 240sec is similar to sequence 9			
14	210→150	3750	4102	4543	4924
		Sequence 11: 300→240→180→150 Time schedule for trains to change the headway from 300sec to 240sec is similar to sequence 9			
12	240→180	3360	3712	4153	4534
15	180→150	3870	4222	4663	5044
		Sequence 12: 300→240→150 Time schedule for trains to change the headway from 300sec to 240sec is similar to sequence 9			
12	240→150	3330	3682	4123	4504
		Sequence 13: 300→210→180→150			
10	300→210	2910	3262	3703	4084
12	210→180	3300	3652	4093	4474
15	180→150	3810	4162	4603	4984
		Sequence 14: 300→210→150 Time schedule for trains to change the headway from 300sec to 210sec t is similar to sequence 13			
12	210→150	3270	3622	4063	4444
		Sequence 15: 300→180→150			
10	300→180	2880	3232	3673	4054
13	180→150	3390	3742	4183	4564
		Sequence 16: 300→150			
10	300→150	2850	3202	3643	4024

1. Stations 4 and 8 are the overlapping stations of regions 3 and 2 and regions 2 and 1 respectively in this application.
2. Trains terminate at station 11 of region 1.
3. Headway is initially changed on train ‘10’ in all sequences.

Table B6 – A set of costs of reaching all states in each stage in Fig. 9.2 with 1500 kVA threshold

Stage 0 to 1	Peak power demand	Additional kVA demand	Cost	Stage 1 to 2	Peak power demand	Additional kVA demand	Cost
$S^1(0) \rightarrow S^1(1)$	1412	0	0	$S^1(1) \rightarrow S^1(2)$	1398	0	0
$S^1(0) \rightarrow S^2(1)$*	-	0	0	$S^1(1) \rightarrow S^2(2)$	1398	0	0
				$S^1(1) \rightarrow S^3(2)$*	-	0	0
				$S^2(1)$*$\rightarrow S^4(2)$	1694	194	0.129
				$S^2(1)$*$\rightarrow S^5(2)$*	-	0	0
Stage 2 to 3	Peak power demand	Additional kVA demand	Cost				
$S^1(2) \rightarrow S^1(3)$	1728	228	0.152				
$S^2(2) \rightarrow S^2(3)$*	-	0	0				
$S^3(2)$*$\rightarrow S^3(3)$	1328	0	0				

$S^3(2)^* \rightarrow S^4(3)^*$	-	0	0				
$S^4(2) \rightarrow S^5(3)$	1694	194	0.129				
$S^4(2) \rightarrow S^6(3)^*$	-	0	0				
$S^5(2)^* \rightarrow S^7(3)$	1290	0	0				
$S^5(2)^* \rightarrow S^8(3)^*$	-	0	0				
Stage 3 to 4	Peak power demand	Additional kVA demand	Cost	Stage 4 to 5	Peak power demand	Additional kVA demand	Cost
$S^1(3) \rightarrow S^1(4)$	1901	401	0.267	$S^1(4) \rightarrow S^1(5)$	1665	165	0.11
$S^1(3) \rightarrow S^2(4)^*$	-	0	0	$S^1(4) \rightarrow S^3(5)$	1665	165	0.11
$S^2(3)^* \rightarrow S^3(4)$	1792	292	0.195	$S^2(4)^* \rightarrow S^2(5)$	2328	828	0.552
$S^2(3)^* \rightarrow S^4(4)^*$	-	0	0	$S^3(4) \rightarrow S^1(5)$	1665	165	0.11
$S^3(3) \rightarrow S^3(4)$	1967	467	0.311	$S^3(4) \rightarrow S^3(5)$	1747	247	0.165
$S^3(3) \rightarrow S^5(4)^*$	-	0	0	$S^4(4)^* \rightarrow S^4(5)$	1787	287	0.191
$S^4(3)^* \rightarrow S^6(4)$	1373	0	0	$S^5(4)^* \rightarrow S^4(5)$	1921	421	0.28
$S^4(3)^* \rightarrow S^7(4)^*$	-	0	0	$S^6(4) \rightarrow S^5(5)$	1709	209	0.139
$S^5(3) \rightarrow S^1(4)$	1901	401	0.267	$S^7(4)^* \rightarrow S^6(5)$	1417	0	0
$S^5(3) \rightarrow S^8(4)^*$	-	0	0	$S^8(4)^* \rightarrow S^7(5)$	2328	828	0.552
$S^6(3)^* \rightarrow S^9(4)$	1792	292	0.195	$S^9(4) \rightarrow S^1(5)$	1665	165	0.11
$S^6(3)^* \rightarrow S^{10}(4)^*$	-	0	0	$S^9(4) \rightarrow S^3(5)$	1746	246	0.146
$S^7(3) \rightarrow S^9(4)$	1887	387	0.258	$S^9(4) \rightarrow S^8(5)$	1665	165	0.11
$S^7(3) \rightarrow S^{11}(4)^*$	-	0	0	$S^{10}(4)^* \rightarrow S^6(5)$	1787	287	0.191
$S^8(3)^* \rightarrow S^9(4)$	1431	0	0	$S^{11}(4)^* \rightarrow S^6(5)$	1833	333	0.222
$S^8(3)^* \rightarrow S^{12}(4)^*$	-	0	0	$S^8(12)^* \rightarrow S^6(5)$	1407	0	0
Stage 5 to 6	Peak power demand	Additional kVA demand	Cost	Stage 6 to 7	Peak power demand	Additional kVA demand	Cost
$S^1(5) \rightarrow S^1(6)$	1968	468	0.312	$S^1(6) \rightarrow S^1(7)$	1786	286	0.19
$S^2(5) \rightarrow S^2(6)$	1834	334	0.222	$S^2(6) \rightarrow S^1(7)$	1786	286	0.19
$S^3(5) \rightarrow S^1(6)$	1968	468	0.312	$S^3(6) \rightarrow S^1(7)$	1787	287	0.191
$S^4(5) \rightarrow S^2(6)$	1756	256	0.171	$S^4(6) \rightarrow S^1(7)$	1786	286	0.19
$S^4(5) \rightarrow S^3(6)$	1834	334	0.223	$S^5(6) \rightarrow S^1(7)$	1787	287	0.191
$S^5(5) \rightarrow S^1(6)$	1968	468	0.312				
$S^6(5) \rightarrow S^2(6)$	1756	256	0.171				
$S^6(5) \rightarrow S^4(6)$	1638	138	0.092				
$S^6(5) \rightarrow S^5(6)$	1834	334	0.223				
$S^7(5) \rightarrow S^2(6)$	1834	334	0.223				
$S^8(5) \rightarrow S^1(6)$	1968	468	0.312				

1. The headway distribution over the line in a state is modified in accordance with the track layout.
2. '*' represents a dummy state in which a zero cost value is given of reaching such state.

References

Chapter 1

[1] Mellitt B., 'The impact of electification systems and traction control on signalling and communications', IEE Vacation School on Railway Signalling and Control Systems – Lecture C6/7; July 1996.

[2] http://www.mtr.com.hk/eng/railway/railway_details.html#quality

[3] Chew K.L. and Pang J., 'An optimisation based approach to the train operator scheduling problem at Singapore MRT', Kluwer Academic Publishers – Annals of Operation Research, 108, pp.111-122; 2001.

[4] Wojtas B.J., 'Development on British Railways traction and rolling Stock', Power Engineering Journal, pp.95-102; March 1989.

[5] Hill R.J., 'Electric Railway Traction – Part 3 Traction power supplies', Power Engineering Journal, pp.275-286; December 1994.

[6] Hill R.J., 'Electric Railway Traction – Part 4 Signalling and interlocking', Power Engineering Journal, pp.201-206; August 1995.

[7] Banerjee M.K. and Hoda N.E., 'Review of the automatic train control system for Cairo Metro Line 2', Power Engineering Journal, pp.217-228; 1998.

[8] Chang C.S., 'Automatic train regulation system – A new concept in computerisation of train control system', IEEE Region 10 Conference on Computer and Communication Systems, pp.821-826 (Vol.2); September 1990.

[9] Pollack M.W., 'Train Control – Automating the World's railways for safety', IEEE Potentials, pp.8-12; February 1998.

[10] Flinders F., Mathew R. and Oghanna W., 'Energy savings through regenerative braking using retrofit converters', Proceedings of the 1995 IEEE/ASME Joint, pp.55-61; April 1995.

[11] Sujitjorn S., 'Coast control logic for automated transit system', MSc. thesis, University of Birmingham, UK, 1986.

[12] Takagi R. and Sone S., 'The development of intelligent power systems for railways', COMPRAIL' 92, pp.39-49; 1992.

[13] De Cuadra F., Fernandez A., De Juan J. and Herrero M.A., 'Energy-saving automatic optimization of train speed commands using direct search techniques', Computer in Railway V, pp.337-346 (Vol.1); 1996.

[14] Chui A., Li K.K. and Lau P.K., 'Traction energy management in KCR', IEE 2nd International Conference on APSCOM, pp.202-208; 1993.

Chapter 2

[15] Schneeweiss Ch., 'Hierarchical planning in organisations: elements of a general theory', International Journal of Production Economics, pp.547-556 (Vol.56); 1998.

[16] Rowe E., 'Los Angeles automated traffic surveillance and control (ATSAC) system', IEEE Transactions on Vehicular Technology, pp.16-20 (Vol.40, No.1); February 1992.

[17] Lo K.R., Chang C.J. and Shung C.B., 'A neural fuzzy resource manager for hierarchical cellular systems supporting multimedia services', IEEE Transactions on Vehicular Technology, pp.16-20 (Vol.52, No.5); September 2003.

[18] Yang X., Fang Y. and Lin Y.B., 'Hierarchical implicit deregistration with forced registrations in 3G wireless network', IEEE Transactions on Vehicular Technology, pp.16-20 (Vol 53: No.1); January 2004.

[19] Fukuda T. and Shibata T., 'Hierarchical control system in intelligent robotics mechatronics', Proceedings of the Industrial, Electronics, Control and Instrumentation, pp.33-38 (Vol.1); November 1993.

[20] Huang G. and Song K., 'A simple two stage optimisation algorithm for constrained power economic dispatch', IEEE Transactions on Power Systems, pp.1818-1824 (Vol.9, No.4); November 1994.

[21] Stoica P., Agrawal M. and Ahgren P., 'On the hierarchical least-squares algorithm', IEEE Communications Letters, pp.153-155 (Vol.6, No.4); April 2002.

[22] Yager R.R., 'On a hierarchical structure for fuzzy modelling and control', IEEE Transactions on Systems, Man and Cybernetics, pp.1189-1197 (Vol. 23, No. 4); July 1993.

[23] Castillo O., Lozano A. and Melin P., 'Hierarchical genetic algorithm for fuzzy system optimisation in intelligent control', IEEE Annual meeting – Fuzzy Information, 2004, Processing NAFIPS '04, pp.292-297; June 2004.

[24] Tang K.S., Man K. F., Liu Z. F. and Kwong S., 'Minimal fuzzy memberships and rules using hierarchical genetic algorithms', IEEE Transactions on Industrial Electronics, pp.162-169 (Vol.45, No.1); February 1998.

[25] Patel D., 'Centralised control on the central Line', IEE – International Conference on Public Transport Electronic Systems, pp.114-120; May 1996.

[26] Wooland D., 'Railway control philosophy – Part One', Institution of Railway Signal Engineers – News, pp.4-11; September 2004.

[27] Goddard E.O., 'Supervision and operation of mass transit systems', IEE Vacation School on Railway Signalling and Control Systems – Lecture D5, pp.1-36; July 1996.

Chapter 3

[28] Nock O.S. (Ed.), 'Railway signalling', A. & C. Black, London, 1990.

[29] Short R.C., 'Fundamentals of signalling and train control systems', IEE Vacation School on Railway Signalling and Control Systems – Lecture B2, pp.1-21; 1992.

[30] http://www.railway-technical.com

[31] Fujii M., 'New Tokaido line', IEEE Proceedings, pp.127-132 (Vol.56, No.4); 1986.

[32] Guilloux J.P, 'High speed signalling in France', International Conference on Main Line Railway Electrification, pp.397-402; September 1989.

[33] Pearson L.V., 'Moving block railway signalling', PhD thesis, Loughborough University of Technology; 1973.

[34] Lockyear M.J., 'Changing track – moving-block railway signalling', IEE Review, pp.21-25; January 1996.

[35] Lockyear M.J. and Norgrove K., 'Transmission based signalling systems', IEE Vacation School on Railway Signalling and Control Systems – Lecture C4, July 1996.

[36] Lockyear M.J., 'The application of a transmission based moving block automatic train control system on Docklands Light Railway', IEE – International Conference on Developments in Mass Transit Systems, pp.51-61; 1998.

[37] Cox G., ‘A study of the three present Linienzugbeeinflussung (LZB) Systems, their functionality and the possibility of defining a universal LZB system’, MEng thesis, University of Bath; 1993.

[38] Pessina G. and Giraudi P., ‘Comparison between traction electrical motors based on the working of a ten years old light rail transit’, IEE – International Conference on Developments in Mass Transit Systems, pp.31-35 (No.453); April 1998.

[39] Attaianese C., Locci N., Marongiu I. and Perfetto A., ‘A digitally controlled DC drive’, Electrotechnical Conference 1994, pp.1271-1274 (Vol.3); April 1994.

[40] Rashid H.M., ‘Power electronics – Circuits, Devices, and Applications, 2nd Edition’, Prentice-Hall International; 1993.

[41] White R.D., ‘Overview of A.C. electrification infrastructure’, 1th Residential Course on Railway Electrification Infrastructure Systems, B4: pp.1-40, Warwick, UK; 2003.

[42] Williment P.T., ‘East coast main line electrification – 25 kV AC system Design’, IEE – International Conference on Main Line Railway Electrification, pp.306-310; September 1989.

[43] Roussel H., ‘Power supply for the Atlantic TGV high speed line’, IEE – International Conference on Main Line Railway Electrification, pp.388-392, September 1989.

[44] Bahra K.S. and Batty P.G., ‘Earthing and bonding of electrified railways’, International Conference on Developments in Mass Transit Systems, pp.296-302 (No.543); April 1998.

[45] Firpo P. and Savio S., ‘Optimised train running curve for electrical energy saving in autotransformer supplies AC railway systems’, IEE – International Conference on Electric Railways in a United Europe, pp.23-27; March 1995.

[46] Yu J., ‘DC power system studies for Jubilee Line’, IEE Colloquium on Systems Engineering on Large Railway Projects, pp.3/1-3/5; May 1997.

[47] White R.D., ‘Overview of D.C. electrification infrastructure’, 1st Residential Course on Railway Electrification Infrastructure Systems, B1: pp.1-34; Warwick, UK; 2003.

[48] Goodman C.J., Mellitt B. and Rambukwella N.B., ‘CAE for the electrical design of urban rail transit system’, COMPRAIL’ 87, pp.173-193; 1987.

[49] Goodman C.J., Siu L.K. and Ho T.K., ‘A review of simulation models for railway system’, IEE – International Conference on Developments in Mass Transit Systems, pp.80-85 (No.543); 1998.

[50] Yasunobu S., Miyamoto S. and Ihara H., ‘Fuzzy control for automatic train operation system’, IFAC Control in Transportation Systems, pp.33-39; 1983.

[51] Goodman C.J. and Siu L.K., ‘DC railway power network solutions by diakoptics’, Proceedings of 1994 ASME/IEEE Joint Railroad Conference, pp.103-110; 1994.

[52] Mohammad T.I. and Nathalie P., ‘Predictive and reactive approaches to the train-scheduling problem: a knowledge management perspective’, IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews, pp.476-484 (Vol. 4); November 2001.

[53] Ho T.K., Goodman C.J., and Norton J.P., ‘An event-based traffic flow model for traffic control at railway junctions’, ASPECT’95, pp.163-170; 1995.

[54] Mellit B. and Goodman C.J., 'Simulator for studying operational and power-supply conditions in rapid-transit railways', IEE Proceedings, (Vol.125, No.4); 1978.

Chapter 4

[55] Mellitt B., Sujitjorn S., Goodman C.J. and Rambukwella N.B., 'Energy minimisation using an expert system for dynamic coast control in rapid transit trains', Conference on Railway Engineering, pp.48-52; 1987.

[56] Wong K.K., 'Optimisation of Run Time and Energy Consumption of Train Movement', MSc dissertation, Hong Kong Polytechnic University; 2001.

[57] Chang C.S. and Sim S.S., 'Optimising train movements through coast control using genetic algorithms', IEE Proceedings - Electrical Power Applications, pp.65-73 (Vol.144, No.1); 1997.

[58] Sell P. S., 'Expert systems: A Practical Introduction', Macmilliam; 1985.

[59] Sujitjorn S., Mellitt B. and Rambukwella N. B., 'Energy minimisation in transit systems using a train based expert system for dynamic coast control', COMPRAIL' 92, pp.303-316; 1992.

[60] Mellitt B, Goodman C.J. and Rambukwella N.B., 'Optimisation of chopper equipment for minimising energy consumption in rapid transit systems', IEE Conference of Railways in Electronic Age, pp.34-41; 1987.

[61] Box M.J., Davies D. and Swann W.H., 'Non-linear Optimisation Techniques', Oliver and Boyd; 1969.

[62] Winston W.L., 'Operations research: applications and algorithms', International Thomson; 1994.

[63] Reklaitis G.V., Ravindran A. and Ragsdell K.M., 'Engineering Optimisation – Methods and Applications', John Wiley and Sons; 1983.

[64] Shoup T. and Mistree F., 'Optimisation Methods with Applications for Personal Computers', Prentice Hall; 1987.

[65] Douglass J., 'Optimum Seeking Methods', Prentice Hall; 1964.

[66] Gottfried S. and Weisman J., 'Introduction to Optimisation Theory', Prentice Hall; 1973.

[67] Bunday B. D., 'Basic Optimisation Methods', Edward Arnold; 1984.

[68] Reisman A., 'On alternative strategies for doing research in the management and social science', IEEE Transaction of Engineering Management, pp.215-220 (Vol.35); 1988.

[69] Lee I., 'Artificial intelligence search methods for multi-machine two-stage scheduling with due date penalty, inventory and machine costs', Computer and Operations Research, pp.835-852 (Vol.28); 2001.

[70] Sheta A.F. and De Jong K., 'Time-series forecasting using GA-tuned radial basis functions', information science, pp.221-228 (Vol.133); 2001.

[71] Chang C.S., Jiang W.Q. and Elangovan S., 'Application of genetic algorithms to determine worst-case switching overvoltage of MRT systems', IEE Proceedings - Electrical Power Applications, pp.81-87 (Vol.146, No.1); 1999.

[72] Chang C.S. and Du D., 'Improved optimisation method using genetic algorithms for mass transit signalling block-layout design', IEE Proceedings - Electrical Power Applications, pp.266-272 (Vol.145, No.3); 1998.

[73] Ho T.K. and Yeung T.H., 'Railway junction conflict resolution by genetic algorithm', Electronics Letters, pp.771-772 (Vol.36, No.8); 2000.

[74] Ho T.K. and Yeung T.H., 'Railway junction traffic control by heuristic methods', IEE Proceedings - Electrical Power Application, pp.77-84 (Vol.148, No.1); 2001.

[75] Haupt R.L. and Haupt S.E., 'Practical Genetic Algorithm', Wiley; 1998.

[76] Goldberg D.E., 'Genetic Algorithms in Search, Optimisation and Machine Learning', Addison Wesley; 1989.

[77] Man K.F., Tang K.S., Kwong S. and Halang W.A., 'Genetic algorithms for control and signal processing', Springer; 1997.

[78] Chan S. H., Ngan H. W. and Rad A. B., 'Minimum-Allele-Reserve-Keeper (MARK): a fast and effective mutation scheme for genetic algorithm (GA)', GECCO' 99, Proceedings of Genetic and Evolutionary Computation Conference, pp.106-113; 1999.

[79] Taskin T., 'Design, Implementation and Application of a Rail Transit System Simulation Model', PhD thesis, University of Birmingham; 1990.

Chapter 6

[80] Cooper L. and Cooper M.W., 'Introduction to Dynamic Programming', Pergamon Press; 1981.

[81] Ho T.K., Norton J.P. and Goodman C.J., 'Optimal traffic control at junctions', IEE Proceedings - Electrical Power Applications, pp.140-148 (Vol.144, No.2); 1997.

[82] Chang C.S., Phoa Y.H., Wang W. and Thia B.S., 'Economy/Regularity fuzzy-logic control of DC railway systems using event-driven approach', IEE Proceedings - Electrical Power Applications, pp.9-17 (Vol.143, No.1); 1996.

[83] Dreyfus S.E. and Law A.M., ‘The Art and Theory of Dynamic Programming’, Academic Press; 1997.

[84] Winston W.L. and Venkataramanan M., ‘Introduction to Mathematical Programming’, Thomson Learning, Inc.; 2003.

[85] Swart B., Cashman M., Gustavson P. and Hollingworth J., ‘Borland C++ Builder 6 Developer’s Guide’, Sams Publishing; 2002.

[86] Herbert S., ‘Borland C++ Builder : The Complete Reference’, McGraw-Hill ; 2001.

Chapter 8

[87] Takeuchi H. and Goodman C.J., ‘Simulation study of peak demand reduction strategies when starting under moving block signalling’, COMPRAIL’96, pp.187-196 (Vol.2); 1996.

[88] Takeuchi H., Goodman C.J. and Sone S., ‘Moving block signalling dynamics: performance measures and re-starting queued electric trains’, IEE Proceedings - Electric Power Applications, pp. 483-492 (Vol.150, No.4); July 2003.

[89] Sanso B. and Girard P., ‘Train scheduling desynchronisation and power peak optimisation in a subway system’, Proceedings of IEEE/ASME Joint - Railroad Conference, pp.75-78; 1995.

[90] Kawakami T., ‘Dynamic power saving strategy and data system for future Shinkansen traffic control’, Proceedings of the 4th International Symposium on Autonomous Decentralized Systems, pp.18-23; March 1999.