Design and Operations of Indian Railways

Dual Degree Dissertation

Submitted in partial fulfillment of the requirements of degree of

Master of Technology

by

Hussain Bharmal

(10D170002)

Supervisor

Prof. Narayan Rangaraj, Dept of Industrial Engineering and Operations Research

Department of Mechanical Engineering

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

June 2015

Dedication

To Indian Railways,
all those train rides across the country have inspired me to contribute.

Approval Sheet

This is to certify that Hussain Bharmal has satisfactorily completed his Dual Degree dissertation on "Design and Operations of Indian Railways" during academic year 2014-15 and his report is approved for submission.

Examiners

Name: R. GOPALAKRISHNAN

Signature: 02/07/15

Name: ASHUTOSH MAHAJAN

Signature: 02/07/15

Supervisor

Name: NARAYAN RANGARAJ

Signature: 2/7/2015

Chairman

Name: SWATI PATANKAR

Signature:

Date: 2/7/2015

Place: Mumbai

ii

Abstract

India has a very high and growing demand for passenger as well as freight trains and there is a need to increase the capacity of existing railway sections to meet this demand. There are sections in the country which are already saturated with extremely congested lines. Better operations (if possible) of these sections is also imperative. This thesis is divided into the following parts:

Understanding different signaling technologies in railway systems and their impact on throughput of the section. The term 'capacity' of a section will be understood in detail and its definition shall be adopted differently in different cases.

Cab to cab signaling (also called just cab signaling) is a form of signaling in which the train driver has information about the location of the train around it in real time, much like traffic on a road. Strategies as to how this technology should be used if implemented in a real section will be devised, keeping different objectives like latency and headway in mind. In order to do this, an algorithm on how to position automatic signals in order to get minimum headway and the parallels of this with cab signaling shall be explained.

The tool used for designing test sections is the train simulator developed at IIT Bombay. Real-life complexities currently modeled by this tool shall be explained.

The Allahabad Mughalsarai section (ALD-MGS henceforth) is a highly congested part of the North Central Railways where passenger trains acquire average delays of the order of hours and freight trains take up to 10 hours to traverse through a 150 km section. We'll demonstrate that the bottleneck of the section was the ALD station and devise strategies for better operations of the section. We'll observe parallels between a railway section and the production line and use ideas from production line theory to devise strategies to determine the bottleneck in railway sections.

Contents

1 Introduction **1**

1.1 Motivation 1

1.1.1 Signaling technologies 1

1.1.2 Capacity of a railway section 2

1.1.3 IIT Bombay rail simulator 2

1.1.4 Allahabad Mughalsarai section exercise 2

1.2 Objectives 2

1.3 Chapter Overview 3

2 Literature Review **5**

2.1 Signaling Systems 5

2.1.1 Absolute block signaling 5

2.1.2 Intermediate block signaling 6

2.1.3 Automatic signaling 6

2.1.4 Cab Signaling 8

2.2 Train Simulator, IIT Bombay 8

2.3 On capacity and UIC 406 10

2.3.1 Number of trains 11

2.3.2 Heterogeneity 12

2.3.3 Average Speed 12

2.3.4 Stability 13

2.4 Optimal automatic signal spacing algorithm 14

2.5 On bottleneck and its relevance to railway systems 16

2.5.1 Production systems 16
2.5.2 Parallels between Railway sections and production systems 19
2.5.3 Limitations of Scott's formula 20
2.5.4 ALD-MGS section: brief description and bottleneck detection 21

3 Signaling technologies and impact on capacity 23
3.1 Modeling cab signaling in the simulator 23
3.2 Standard section exercise 24
3.2.1 Train and track settings 25
3.2.2 Settings for different simulations 25
3.2.3 Results of exercise 26
3.2.4 Standard section conclusions 28
3.2.5 Cab signaling and completely coupled trains 29
3.3 Secunderabad Wadi section exercise 32
3.4 Capacity analysis with mixed traffic and punctuality constraints 34
3.4.1 Results of the exercise 35

4 Automatic signal spacing algorithm and cab signaling 38
4.1 Signal spacing algorithm 38
4.2 Results for Western Railway (suburban corridor) 39
4.2.1 Parameter values for the exercise 39
4.2.2 Results for automatic block signaling 39
4.3 Cab signaling algorithm using brakefrom and T functions . . . 42
4.4 Results of cab signaling and comparing with auto blocks . . . 43
4.5 Takeaways and limitations 45

5 Constraints on train path parameters and implementation via IIT Bombay rail simulator 47
5.1 Block working time 47

5.2 Link velocity 48
5.3 Gradients 51
5.4 Loop/block velocity 53
5.5 Speed restrictions within block 54
5.6 Overtaking of freight trains 54
5.7 Overtaking of scheduled trains 55
5.8 Crossover movement 56
5.9 Parameters not implemented 60

6 Allahabad Mughalsarai section: Bottleneck alleviation 62
6.1 Section details and model 62
6.2 Visit to Allahabad and observations on the section 66
6.3 Detecting bottleneck 67
6.4 Better operating strategies and investments 69

7 Conclusion 73

8 Appendix 75

List of Figures

2.1 Working of 4 aspect automatic signaling . . . 7
2.2 Pillars of railway capacity . . . 11
2.3 Effects of heterogenity on capacity . . . 12
2.4 Capacity vs speed . . . 13
2.5 Punctuality vs capacity utilization . . . 14
2.6 Blocks, signals, sighting, overlap, and braking . . . 15
2.7 Linear production system with 4 steps . . . 16
2.8 Part of the ALD-MGS IB signals . . . 19
2.9 Graphical Interpretation of a railway section . . . 20

3.1 Absolute block signaling - standard section . . . 26
3.2 Intermediate block signaling - standard section . . . 26
3.3 Automatic signaling - standard section . . . 27
3.4 Cab signaling - standard section . . . 27
3.5 Velocity profile on SC - W division . . . 33
3.6 IB: Traversal time v/s number of trains . . . 35
3.7 Auto block: Traversal time v's number of trains . . . 36

4.1 Headway v/s inter-station distance in auto block signaling . . . 40
4.2 Optimal Block lengths sizes vs inter station distance . . . 41
4.3 headway vs booked speed in auto block signaling . . . 42
4.4 Headway v/s inter station length comparison for auto block and cab signaling 44
4.5 Headway v/s booked speed comparison for auto block and cab signaling . . . 45

5.1 Block working time 0 mins . . . 48
5.2 Block working time working . . . 48
5.3 Link velocity working with link length of 200 m . . . 49
5.4 Link velocity working with link length of 10 m . . . 50
5.5 Link length working with link length of 0 m . . . 51
5.6 no gradient effect . . . 52

5.7 Uniform up gradient result 52
5.8 Loop velocity restriction 53
5.9 Speed restriction in blocks 54
5.10 Overtaking of freight trains by passenger trains 55
5.11 Overtaking of scheduled trains 56
5.12 Crossover test section 57
5.13 crossover at Jigna 58
5.14 Result of crossover 59
5.15 No crossover result 59
5.16 Time-table comparison for crossover movements 60

6.1 Allahabad division 63
6.2 ALD-MGS section 63
6.3 ALD-MGS block info 64
6.4 ALD-MGS Loop info 65
6.5 ALD-MGS Model 65
6.6 Train going from CAR to ALD 67
6.7 Trains going from ALD to CAR 68
6.8 ALD-MGS single train path 69
6.9 Fast and slow trains moving consecutively 71
6.10 Platooning of trains 71

8.1 Simulator main window 76
8.2 GUI input main window 76
8.3 Station inputs 77
8.4 Loop inputs 78
8.5 Loop list 78
8.6 Block inputs 79
8.7 Parameter inputs 80
8.8 Gradient inputs 81
8.9 Train inputs 82
8.10 File selection window 83

List of Tables

3.1 Standard section results 28
3.2 comparing cab signaling policies 32
3.3 SC - W intermediate and auto/cab signaling 34
3.4 SC - W intermediate signaling iteration results 35
3.5 SC - W automatic signaling iterations results 36
3.6 SC - W cab signaling iteration results 37
3.7 Comparing automatic and intermediate signaling 37

4.1 Headway v/s inter station length comparison for auto block and cab signaling 44
4.2 Headway v/s booked speed comparison for auto block and cab signaling 45

6.1 SC - W intermediate and auto/cab signaling 68

8.1 gradient effects 81

Nomenclature/Terminology

Following are the basic terms used in railway operations useful for this report:

- Passenger train: Train used for carrying passengers.
- Freight train: Trains used for transporting cargo as opposed to human passengers.
- Block: Any part of a railway track with some definite starting and ending locations whose position is designated by light (R/B/Y) based signals.
- Station: A designated location in a railway section where the trains can halt so that the public may board/off-board or goods may be loaded/unloaded.
- Loop: A block in a station area is termed as loop.
- Link: The part of a track connecting two blocks or a block and a loop on same or different railway tracks.
- Uplink: A link which connects two blocks in up direction is termed an uplink.
- Downlink: A link which connects two blocks in down direction is termed as a downlink.
- Crossover: A link while connecting two blocks may happen to cross another railway-line at some block on that line. That link is called a crossover.
- Velocity Profile: It gives us information about the speeds of a particular train at various locations.
- Block occupancy: The various times for which a particular block is occupied by various trains
- Loop occupancy: The various times for which a particular loop is occupied by various trains.
- Up main line: The railway line designated for the transport of trains in the up direction. It is sometimes referred as simply the up line.

- Down main line: The railway line designated for the transport of trains in the down direction. It is sometimes referred as simply the down line.
- Common line: The railway line designated for the transport of trains in the up as well as down direction.
- Throughput of Section/capacity: The number of trains crossing a section in any direction in unit time. This meaning will be expanded in multiple ways in this report.
- Headway: The time difference between two consecutive trains to clear the section.
- Latency of section: The average time taken by a train to cross the section.
- Block overlap distance: The train's rear has to clear a certain distance away from the next signal before the previous signal turns its aspect again. This distance is called block overlap distance.
- sighting distance: The distance away from the signal from which the signal color can be observed.
- Direction switch: It is the event of switching the direction of the common line.
- Block delay: The time it takes for a block to be declared free after a train has cleared it. This is an information delay involved with signaling.

Following are the nomenclatures used often in this report:

- Y - yellow aspect
- YY - double yellow aspect
- G - green aspect
- R - red aspect
- IR - Indian Railways
- NCR - North Central Railways
- SCR - South Central Railways
- ALD - Allahabad
- MGS - Mughalsarai
- ALD-MGS - Allahabad Mughalsarai section

- SC - Secunderabad
- W - Wadi
- SC - W - Secunderabad Wadi section
- IB - Intermediate Block
- NSR - New Sketch Rail simulator
- W - halt time at station

Notations for chapter 4 and related literature:

- H_i - headway of block i
- H overall headway between 2 stations
- L_i - Length of block i
- s_i - sighting distance of signal i
- o_i - block overlap distance for block i
- l_t/l - length of train
- b_i - braking distance at signal i

Declaration

I declare that the submission of this report represents my ideas in my own words and I have adequately cited and referenced the original wherever I have used other's ideas and words. I also declare that I have adhered to all principles of academic honesty and integrity. I have not misrepresented or falsified or fabricated any idea/data/fact/source in my submission. I understand that any violation of above will lead to a disciplinary action being taken by the institute and evoke penal action from the sources which have not been properly cited or whom proper permission has not been taken.

Hussain Bharmal

10D170002

Signature

Date: 2nd July, 2015

Place: Mumbai.

Chapter 1

Introduction

1.1 Motivation

The broad areas of this study revolve around signaling technologies and capacity planning on a rail section and bottleneck alleviation in railway operations. The importance of each of these is discussed in turn:

1.1.1 Signaling technologies

Signaling is an important aspect of railway operations. Before the advent of signaling systems, trains were operated using systems like fixed timetable in which a designated time was given to each train at different parts of the track, without any form of clearance signaling involved [10]. The trains were expected to move according to schedule and appropriate slacks were provided. This is rather unsafe as there is no direct confirmation whether the track is clear or not and may lead to accidents in case the previous train had halted on the track for operational reasons. Then came absolute block signaling when telegraphic methods and others were used to transfer signal information. Intermediate block was an extension of it. As technology developed further, automatic signaling came into use and is the most advanced system used in Indian railways to date. The next upcoming (and most likely the best one possible) is cab signaling, in which all trains know where the train in front of it is in real time. It's the best one theoretically as it gives complete information which a driver may require from a signaling system: the exact location of a train at all points of time. (The description of these technologies are in the literature review chapter).
An important question to ask with a signaling system is the impact of it on the throughput of the section. This is one of the studies conducted.

1.1.2 Capacity of a railway section

Broadly speaking, capacity of a rail section means the number of trains it can handle in a day or a unit time. There are a lot of aspects on which capacity depends, like heterogeneity of traffic, punctuality of trains, bottlenecks in the section, operating strategies employed, signal spacing and average speeds, to name a few. It is important to understand how capacity of a section should be defined in light of these restrictions placed on the track. For example, capacity defined while only considering speeds as a criteria will be far from the reality on a track which has large speed differentials (high heterogeneity). Thus, understanding capacity and how should it be applied in different cases is considered over here.

1.1.3 IIT Bombay rail simulator

The IIT Bombay rail simulator tool has been used extensively in this study. It is used to obtain feasible train paths of scheduled trains and unscheduled freight trains and can be used to make capacity statements. The simulator has been developed over the last decade and part of its functionality which hasn't been explained in previous reports so far needs to be consolidated in terms of writing. This will be of help for future exercises to be conducted using this tool.

1.1.4 Allahabad Mughalsarai section exercise

ALD-MGS is an extremely congested section of North Central Railways, one of the presently 17 railway zones in India. Even high priority trains like Rajdhani get delayed by the order of an hour while freight trains take up to 10 hours to cross the 150 km section. This is nothing short of a 'crisis situation' as described by the personnel of the Allahabad division. This section is also on the main corridor connecting East India (Kolkata side) to West India (New Delhi side). Thus, suggesting better operating strategies on a critical, real section was the motivation to undertake this exercise.

1.2 Objectives

The objectives of the exercises conducted are closely related to the motivations suggested in the last section:

1. For a standard 10 km section with just two stations and no gradients or speed restrictions, estimate the impact of signaling technologies on the throughput of the section. After this, the same was done for a real section of Indian Railways. The Secunderabad Wadi section of the South Central railways was chosen for this. In both these cases,

the definition of capacity employs only one type of train which is available for firing as soon as the track is empty for it and punctuality isn't kept as a criteria.

2. For the Secunderabad Wadi section, estimate capacity as the number of freight trains that can be run on the section given the current scheduled timetable and while maintaining a certain amount of punctuality (the punctuality criteria used was the following: average running time of freight train given passenger train timetable should be less than twice the running time of a freight train which has a through/free path). This will be done for different signaling technologies and the increment in capacity as defined will be observed.

3. Understand the optimal automatic signaling spacing (optimal in the sense to minimize the headway given the constraint that trains have to move in a free fashion unaffected by other trains, or in other words while maintaining minimum latency). As part of the same exercise, estimate the minimum headway possible on the Western suburban corridor in Mumbai and also estimate the impact of train velocity and interstation length on headway. Then, develop an algorithm to estimate capacity with cab signaling on the same track and train combinations and the same strategy.

4. Understand the complexities observed in running trains on a section: different kinds of speed restrictions and protocols observed when paths of trains moving on the same section meet. Also understand how these are implemented within the IIT Bombay Rail simulator. For example, station entry velocity, gradients, loop velocities, link velocities and other parameters. As documentation for the working of these parameters' implementation hasn't been done in the simulator, this is another objective for this part of the report.

5. Study the ALD-MGS section: station layouts, signal positions, working time-table. Model the section within the simulator. Estimate the bottleneck of this section and suggest ways to alleviate the bottleneck and the operations of the whole section in general to reduce headway and latency using the simulation model and other industrial engineering principles.

1.3 Chapter Overview

Chapter 2 contains the literature review for the studies conducted. It follows chronological order in which work has been done over the last year. The description given is kept as simple as possible and material has been suggested for further reading where required. Chapter 3 includes the work done on signaling technologies and their impacts on capacity. Objectives

1 and 2 are addressed in this chapter. Chapter 4/5/6 contain the work done trying to meet objectives 3/4/5 respectively. After this, we discuss the key conclusions of the work done and define problems that may be worked upon in the future to expand these exercises in Chapter 7. The Appendix at the end includes the manual for the form based GUI version of the IIT Bombay simulator which has been developed to make the tool more user friendly.

Chapter 2

Literature Review

2.1 Signaling Systems

The first part of this project revolved around studying the impact of signaling systems on throughput of the section (ceteris paribus). In order to do this work, we'll understand the different types of signaling systems used in the modern day railway systems. The primary interest for this study is in abstracting these systems for the simulator purposes rather than how they are technologically implemented.

2.1.1 Absolute block signaling

This is a signaling scheme designed to ensure the safe operation of a railway by allowing only one train to occupy a defined section of track, also known as a block, at a time. The block working is manual, using railway employees for the same. The information is communicated between consecutive 'signalmen' via signal codes. There is just one block between consecutive stations in this method, which means that there is only up to one train between consecutive stations. [1] The advantages and disadvantages of this system are:

- Advantages: This is an extremely safe method of operation as one can imagine. The next train is not allowed to even enter a block till the previous one's tail has gone out of the block in the block working method (true for other signaling systems which employ the block working method). Also, technical failures like signal failures which are seen with automatic signaling won't be seen here as operations are manual.

- Disadvantages: It seriously limits the capacity of a railway section: as only one train occupies a given section, the time difference between two consecutive trains, also called the headway, is very high. The larger the size of the block, higher is the expected headway. The time response of the setup is also high as it is manually operated.

Nonetheless, this is historically the most important type of signaling system used and is still currently used on sections where better technologies haven't been employed yet. ex. Dagmagpur to Jhingura (15.5 km, 1 intermediate station which is Pahara) in the ALD-MGS (Allahabad Mughalsaraidivision.

2.1.2 Intermediate block signaling

As the name suggests, in this system, there is an intermediate signal placed in the middle of an existing block, thus dividing it into 2 smaller blocks. [1] One can treat this system as Absolute Block Signaling in principle, but with more number of blocks on the same section. The intermediate block signals are also manned in India via block huts, which act as pseudo-stations to pass signals to trains. The pros and cons of this system are:

- Advantages: Provides the same type of safety as Absolute Block Signaling. Also, there can be 2 trains in the same space where there was just one previously as there is an extra block to accommodate the extra train. This leads to a higher throughput of the section as our analysis will show. The basic reason for the same is that there can be 2 trains in the same region where previously there was only 1, without adversely affecting the speed of each other.

- Disadvantages: In principle, it is still the same as absolute block signaling, with block lengths quite large. Thus the same points for disadvantages hold over here too. Also, the latency of section can increase in this system over the absolute block signaling as the train following it might have to wait if the block ahead of it has a train present in it, thus increasing the former one's traversal time than the absolute signaling case, in which the train would have just passed through. The same shall be discussed more in later sections.

2.1.3 Automatic signaling

Movement of the trains is controlled by the automatic stop signals in this method. Signals change color/aspect automatically by the passage of the train. Basics of this system: [2]

The line is divided into a series of automatic signaling sections each of which is governed by an Automatic Stop Signal with generally 3 or 4 aspects. 3 aspect signaling has the colors Red (Stop), Yellow (Caution) and Green (proceed freely), while 4 aspects will have an additional double yellow (Attention). 4 aspect signaling is currently being used in Mumbai suburban traffic and most parts of the ALD-MGS section among others in the country. Working of a 4 aspect automatic signaling is described in the figure below. Theoretically, the aspects can be kept at any positive integer greater than 1.This is important for modeling cab signaling (explained next) using automatic signaling with large (theoretically infinite) number of colors as will be discussed in chapter 3.

Overlap distance: No Automatic Signal assumes 'OFF' (not red) unless the line is clear not only up to the stop signal ahead, but also an adequate distance beyond it called the block overlap distance. For 4 aspect signaling, the block overlap is kept at 120m in western suburban section as well as the ALD-MGS section where automatic signaling has been implemented.

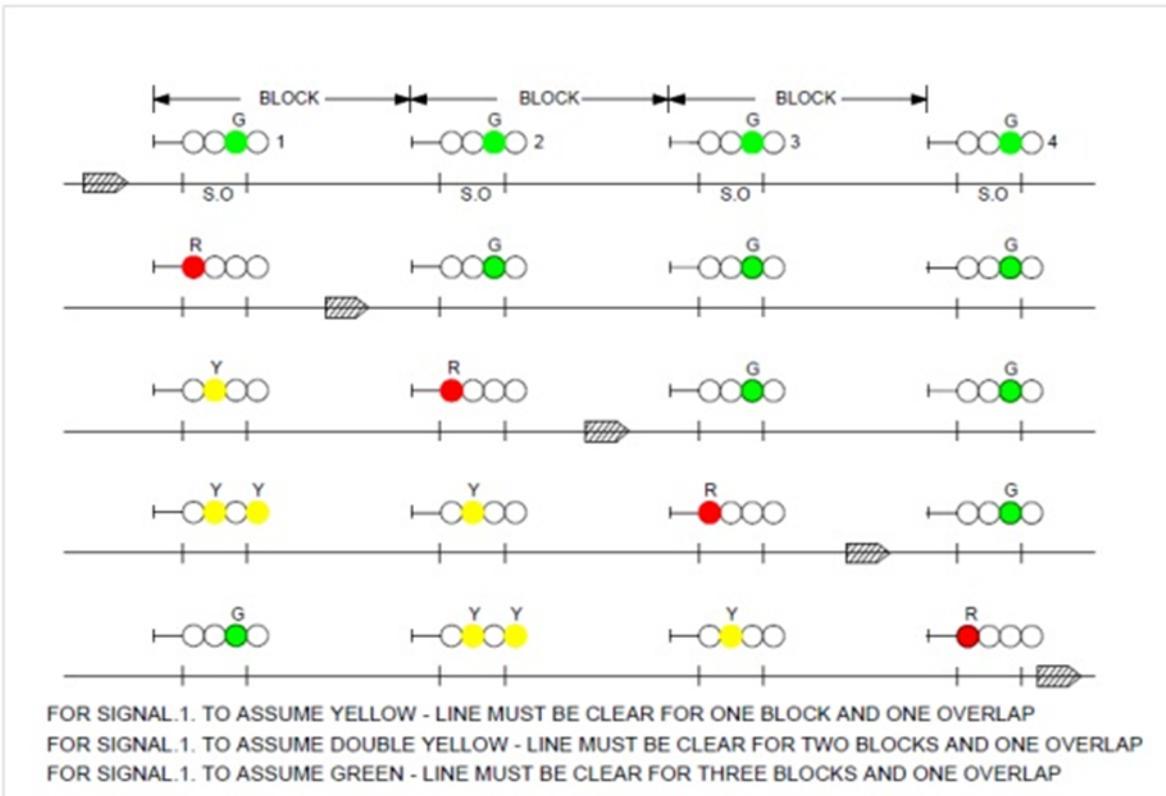

Figure 2.1: Working of 4 aspect automatic signaling

Image source: [2]

The railway simulator uses automatic signaling method in its logic, but with zero overlap distance. Other signaling systems have to be suitably abstracted to make them amenable to implementation via the simulator logic. The advantages and disadvantages of automatic signaling are as follows:

- Advantages: As manual working is removed, there is a much faster time response due to automation. Signals can be placed quite close to give smaller block size (it's impracticle to put manned block huts at every km, but possible to put an automated signal after every km of the track) and thus higher throughput. (discussed in later sections)

- Disadvantages: Technological failures. Example: There have been cases of this system failing in monsoon conditions in Mumbai leading to jams. The fail-safe protocol of an automatic signal is that it will always show red upon failure. So the train will move with extreme caution (assuming there is a red at the next signal too) if it sees red for a long time and no train ahead of it. This is a protocol observed in Indian railways and was observed in working at a signal near ALD. This increases the traversal time in case of failure.

2.1.4 Cab Signaling

Cab here means the crew compartment or the driver's compartment. As mentioned earlier, the train has data about the location and movements of the train ahead of it in (near) continuous time. This signaling system gives the maximum capacity possible compared to all other signaling system (ceteris paribus) as every train has complete information about the location about the driver in front of it. In other systems with block working, the position of the train in front of it (position of a train refers to the position of the rear of the train) can only be known within a block. Thus, there are no artificial blocks stopping trains for safety issues. The train will clear in continuous time, the part of the track which is some fixed safety distance behind its rear. This distance between two trains is always maintained so that in no case will two consecutive trains come closer than this distance, thus ensuring safety. This is much like traffic on a road, where car drivers visually know the location of the car in front of them all the time and maintain a safety distance while driving their vehicles. Cab signaling is used in North America and news articles have mentioned the interest of railway authorities in using this technology for the Harbour line of Mumbai suburban railways.

- Advantages: Maximum capacity can be achieved with this system while maintaining safety. No reliance on line side signals, either automated or manned.

- Disadvantage: Unless all trains in a given section are equipped with the technology, cab signaling can't be used.

2.2 Train Simulator, IIT Bombay

The rail traffic simulator is a JAVA based tool developed over the years on the IIT-B campus for study and analysis of rail operations. It computes valid paths of trains given inputs using block signaling method. It handles train scheduling on a linear section and generates a conflict free, feasible schedule along with time space graphs of trains and block occupancy charts. The functionality of the algorithm is described via the input-output description of the tool: (Chapter 2 of [3] has the inputs and outputs in further details)

Inputs:

1. Stations and loops: Start and end locations of stations. Loop configurations at the station. Loops can be up, down or common. Loops are linked with blocks and crossovers are also handled. Link priorities, link lengths and velocity restrictions and maximum velocity in loops are also considered.

2. Blocks: Start and end locations, directionality, maximum velocity within a block, further specific speed restrictions (ex. permanent speed restrictions on tracked curves), links with loops or other loops are considered.

3. Gradient and gradient effects: The ground may or may not be perfectly level throughout the section. The natural ups and downs with respect to the line of gravity have a considerable effect on the acceleration and deceleration of trains. These inputs basically take care of the gradients occurring on the tracks and how much they affect the acceleration/deceleration of the train depending on the gradient.

4. Scheduled trains: Inputs include start location and time, scheduled halt locations (locations are specified at loop level), arrival and departure times on these halt locations, length of train, max velocity, acceleration, deceleration, priorities (high priority trains will get to go first. There are priority up to 10 levels based on different versions of the simulator). Note that these inputs are for the desired time-table. Simulator will generate a feasible schedule based on given inputs.

5. Unscheduled trains: They are low priority by default. Train characteristic inputs are same as those of scheduled trains. Only starting location, starting time and final destination are mentioned due to the low priority nature of freight trains: No desired final times are mentioned or even where it is supposed to stop in the middle. It may stop at multiple locations depending on the availability of free paths to it.

6. A ('Global') parameter file: Contains the simulation time (duration of simulation in number of days), block working time (time for information exchange between the clearing of a block and the setting of the signal for the next train to proceed) and number of signal colors.

Output: Conflict free, feasible schedules of as many input trains as possible in the given simulation time along with space time graphs of all trains and their block reservations and occupancies. Note that block occupancy is different from block reservations. A train may not have occupied a block yet but it may have reserved the block for itself if there is no other train which can occupy the considered block before the said train.

The simulator's output can also be used to make capacity statements. For example: Running a single train on the designed section, the time it takes to traverse each block, loop can be obtained from the output, and using the time it takes to travel through block/loop which takes maximum time, capacity statements using Scott's formula (refer 2.3.1.1 of this report) can be made. Generalized capacity analysis with mixed traffic and punctuality constraints can also be done with iterative methods as suggested in 3.4 of this report.

There are 3 different versions of the simulator which have been used for this project. Please refer to the appendix for further information on these.

2.3 On capacity and UIC 406

Capacity is loosely defined as the number of trains that can be handled on a given section in a day. This definition is too simplistic and needs to be expanded in the light of the fact that certain constraints are placed on the operations of railway tracks. To state a few:

- How many trains can be handled on the given track while maintaining some level of punctuality?
- What is the capacity of the section if trains should run only on a YY signal by design? i.e. Trains should run at a headway in such a way that they always see either a YY or green when they are passing a signal so that they can move more or less freely.
- What is the capacity of a section which is supposed to handle mixed traffic with speed differentials between trains?
- Indian generates maximum revenues by freight trains only. So, in this context, the maximum number of freight trains which a track can handle given a time-table of scheduled trains on the same track will be the capacity of the said track. Thus, in this case, the capacity is just reduced to counting the number of freight trains on the section. One may even desire these freight trains should pass the section in a maximum time. Thus, the capacity with the added constraint will change further.

These are all different definitions of capacity while constraining the system to some criteria. Different railways divisions may employ different definitions of capacity based on their modus operandi. UIC 406 is a set of guidelines to generalize and estimate rail section capacity. This estimation is based on a train path compression algorithm, in which scheduled train paths on a section are pushed as close to each other as feasibly possible to operate (without violating the constraints like block occupancies, halt time and other user defined constraints which may be desired.) to create extra capacity. Based on the extra paths created after the compression, it calculates the current capacity utilization of the given time-table. The detailed working of the same can be found in [4] and [6]. UIC 406 answers the question of generalizing the idea of capacity and describes that capacity of a section depends on these parameters:

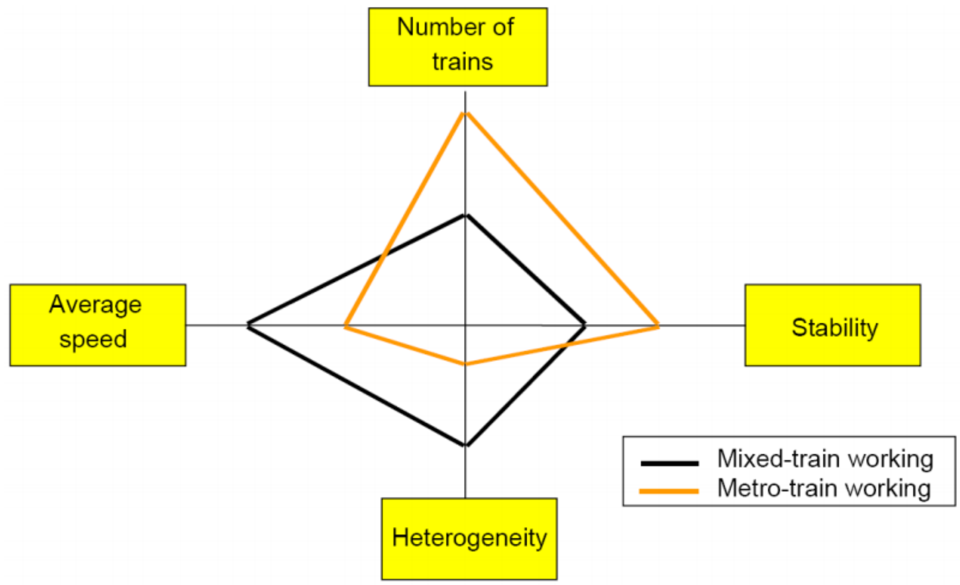

Figure 2.2: Pillars of railway capacity

Image source: [4]

Capacity is a balance of these 4 factors. For instance, it is possible to achieve a high average speed on a railway network and have a high heterogeneity – a mix of fast express trains and slower regional trains serving all stations. However, the cost of having high average speed with a high heterogeneity is that it is not possible to run as many trains with a high stability (punctuality) than if all trains ran with the same high speed i.e. a homogeneous train mix. 4 factors are examined here as they find use in this study:

2.3.1 Number of trains

This is arguably the most important pillar of capacity and is the basis on which the Scott's formula is derived. If capacity is measured as the number of trains per hour that the section can handle, then it (capacity) is just the maximum possible traffic intensity on the section, regardless of the heterogeneity or punctuality of these trains.

2.3.1.1 Scott's formula

This formula is used to calculate the capacity of a homogeneous railway section where demand is same throughout the section i.e. trains are going through from the start to end of the section:
capacity (number of trains in an hour)= 60/t
where t is the maximum amount of time in minutes it takes the train to traverse through any 2 consecutive signals on the section.(i.e. the maximum of traversal time through any block or loop) This is akin to finding the bottleneck which limits the output in a production planning scenario. More discussion on Scott's formula is included in 2.5.3 which explains this idea and its limitations in further detail.

2.3.2 Heterogeneity

A timetable is heterogeneous (or not homogeneous) when a train catches up another train i.e. there are speed differentials among trains on the same section. The result of a heterogeneous timetable is that it is not possible to run as many trains as if the timetable was homogeneous – all trains running at the same speed and having the same stopping pattern, which can be understood by this figure (the horizontal axis is the location. Figure a is mixed traffic which takes a higher time to run the same number of trains than b, which has homogeneous traffic.):

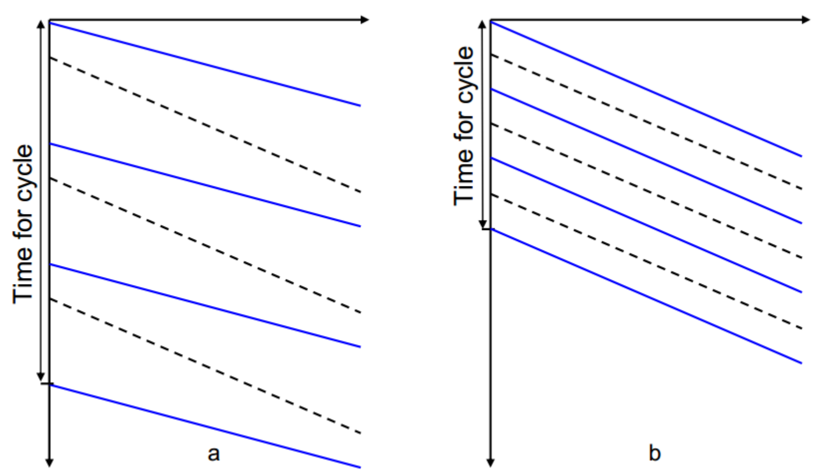

Figure 2.3: Effects of heterogenity on capacity

Image source: [1]

2.3.3 Average Speed

A train consumes different amounts of a resource at different speeds. When a train stands still, the train consumes all the resource since it occupies the block section for an infinite amount of time and hence the throughput/capacity in this case is 0. When the train speed increases from 0, it occupies the block section for shorter time so more trains can pass the same block section in a given time. Thus, capacity increases from an argument using Scott's formula. Increasing the speed beyond a certain limit will decrease the throughput though, as braking distance is increases with speed. This means that the headway distance – and headway time – is increased too. It can also be shown that the 't' in Scott's formula is nothing but the minimum headway time. Thus with Scott's formula, there is a trade-off between average speed and capacity. A study based on this was conducted for the Western suburban section, which will be discussed in chapter 4. Detailed theory of this idea has been developed in [5]

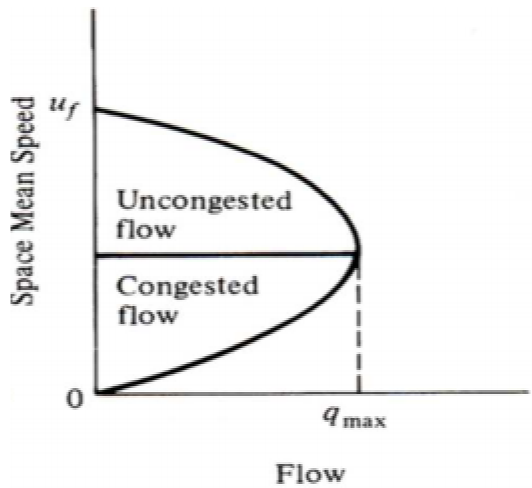

Figure 2.4: Capacity vs speed

Image source: [1]

2.3.4 Stability

Stability means punctuality, or trains reaching their designated locations on time. From principles of queueing theory, higher the capacity utilization of a resource (which will happen with a combination of high arrival rate (amount of traffic) and low service rate (high headway)), higher is the average waiting time or loss of punctuality. This holds over here too and the following generic graph is observed:

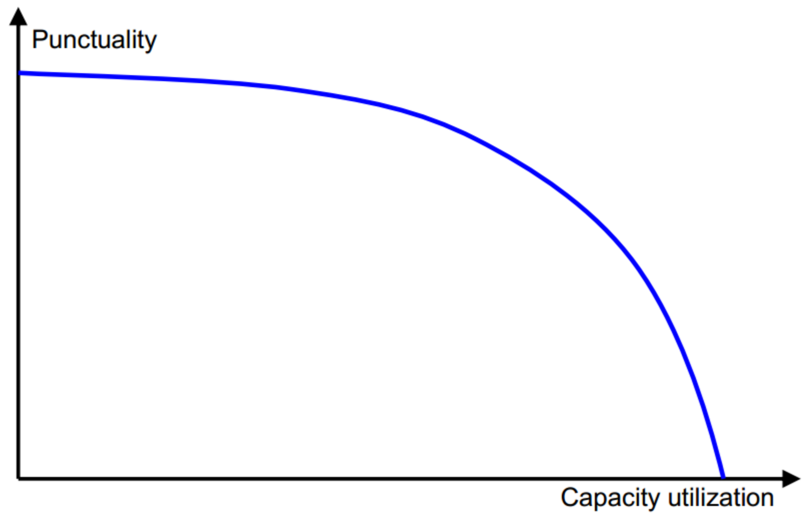

Figure 2.5: Punctuality vs capacity utilization

Image source: [4]

In chapter 3, one of the studies done on the Secunderabad Wadi division employs a definition of punctuality which incorporates stability: The maximum number of freight trains that can be operated on the section with an upper cap on the average traversal time.
The key takeaway is that capacity is not just the number of trains that can be operated on the section. The constraints of the system and the user requirements are also an important part while defining capacity. Different definitions of capacity have been used in the studies done, from simple ones based on Scott's formula using just number of trains to fairly complex ones involving mixed traffic and punctuality criteria.

2.4 Optimal automatic signal spacing algorithm

While designing a railway section layout, it is desirable to put signals is such a way that low latency and low headway are ensured, but as seen in 2.3.3, there is a tradeoff between average speed and capacity. In [7], a signaling system has been designed which enables to run train in the following conditions: (a) the minimum possible latency under normal operating conditions, and (b) the minimum headway possible for that minimum latency. i.e. It designs a layout which minimizes headway given that trains travel as fast as possible under normal conditions and aren't affected by network effects (slowing down or halting due to getting close to other trains) by design.

Note that this method employs the definition of capacity which ensures high speeds (in fact highest permissible operating speed or free flow of traffic) and homogeneity (trains of only 1 type). It is assumed that the signaling system is 4 aspect automatic signaling like the western suburban corridor. For free flow of train, it is assumed that trains should move only see a YY or green at every signal. In case of Y, although the train can move, driver's show extreme caution (they decrease the speed) as they might see a red at the next signal and would have to stop.

Thus, the train should observe a YY (at least) at all signals which are not home signals and at least a Y at home signals. This relaxation is done for the home signal (home signal means the signal just before the station) as the train is supposed to halt at the station by design. Thus a Y headway will ensure the free flow at a station.

A YY(Y) headway of a signal means the time it takes the signal to turn from YY(Y), when a train passes to YY(Y) again (when the train ha cleared the next 2(1) block from the considered signal). Thus, for signal i, we want H_i, the headway at signal i, to be the YY headway for all signals not home signal and Y headway for the home signal as discussed.

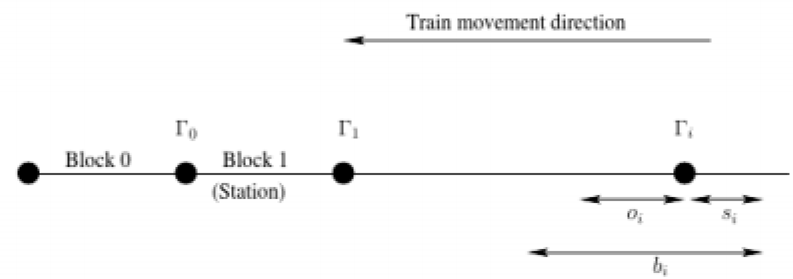

Figure 2.6: Blocks, signals, sighting, overlap, and braking

Image source: [7]

The nodes are the signal locations in the above figure. Block 0 is the block after the station. o_i,s_i, and b_i are the overlap, sighting and braking distance respectively at block i. Let the block numbering be 1,2 and so on starting from home signal and then pre-home signal and so on.L_i is the length of block i. Then H_1 and H_2 surely involve a halt time: home signal (signal 1) will turn Y again only after the train has cleared the loop and pre-home signal (signal 2) will turn YY again only after the train has cleared the home block and the loop block, which involves the halt time. For H_3, the headway calculation may or may not involve W (halt time): If L_1(loop length) is small, then due to the fact that trains have to clear an overlap distance beyond a block also for the previous signal to turn, the waiting time will be involved for signal 3 to turn YY again. It can be shown that L_1 has to be larger than $o+l_t$ for waiting time at station to not affect H_3,where l_t is the length of the train. If it is less than the said quantity, then W will be accounted in H_3.

It's easy to show that H_1 is always less than H_2, as the length traveled in converting signal 1 from Y to Y again is also traversed in converting signal 2 from YY to YY again. Now as H_i (i=1,2,3) potentially involve waiting time and decelerations, it is expected by and large that the max headway of all sections, which is the system headway used to determine capacity, will be either H_2 or H_3.

Thus, the algorithm developed finds the values of H_2 and $H_3$for different values of L_1. These can be used to determine the value of L_1which minimizes $H = \max\{H_2, H_3\}$. Details of the same are in [7].

2.5 On bottleneck and its relevance to railway systems

From an engineering point of view, a bottleneck refers to a phenomenon where the performance or capacity of an entire system is limited by a single or small number of components or resources. In production, a bottleneck is one process in a chain of processes, such that its limited capacity reduces the capacity of the whole chain. The concept of bottleneck using production on assembly line shall be explored (processes happening sequentially and in a specific order to obtain a final product) after which parallels between this with railway sections shall be drawn. Applications of these concepts to determine the bottleneck on the Allahabad-Mughalsarai division of North Central Railways shall be seen.

2.5.1 Production systems

In this section, we define a simple production system and intuitively identify the bottleneck. Then we discuss multiple definitions of bottleneck and how to detect them.

2.5.1.1 Bottleneck based on processing times

Consider a production system as shown below.

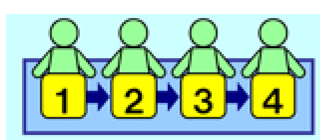

Figure 2.7: Linear production system with 4 steps

The raw materials enter at 1, they get processed sequentially from 1 to 4, and then the final product comes out at 4. Note that the steps over here are discrete. The transformation of the product is not continuous in the sense that once one machine starts working on a given part, that machine will finish processing that part when done. Then it goes to the next step. Thus, the evolution of the product is discrete and the state of the product changes state only upon completion of processing by a resource. This is similar to our railway system where the section is divided into blocks and loops by signals, thus dividing the 'flow' of the train in finite, discrete parts similar to the given production line.

Now consider that the service time for each of these processes has mean time p_{i} i=1,2,3,4 in minutes. Also assume p_2 is the largest of these. Thus, the system can have an average maximum throughput of $60/p_2$ only as although other parts maybe processed quickly, the process at 2 will take the maximum time and will thus determine the throughput. If the occupancy of resource 1 is kept higher in such a way that it processes more parts than what 2 can do in the same time on an average, then these parts will form a pile of inventory at 2 waiting to be processed at a rate of p_2. Also, given that 2 has the highest processing time (and thus the lowest processing rate), the subsequent stages can have a processing rate no more than that of 2, as they are being fed parts at that rate. Even if they have extra capacity left (i.e. they are free to process more parts), they won't be able to do so as the feed rate is slower and governed by the output of 2. Thus, the system throughput is governed by the processing rate at one of the 'nodes' of production. This node is precisely the bottleneck as it is a process which limits throughput of the system. Note that this is the principle on which Scott's formula discussed in 2.3.1.1 is based.

2.5.1.2 Different definitions of bottleneck

As seen previously, some of the resources affect the system performance like throughput more than the others. Usually, the limitation of the system can be traced to the limitation of 1 or 2 resources, which are the bottlenecks. One way to determine the bottleneck was seen in the last section, namely the processing time method. In more complex systems which involve branching and more 'graphical' (as opposed to linear) flow of parts, the definition of bottleneck needs to be changed as the demand at each node need not be the same. (The previous section assumed the same demand for all nodes and determined the bottleneck using just processing times.) Thus, expanding the definition of bottleneck in a general system is required, as processing times aren't sufficient: ex. consider a simple way to see is inventory piling up at a resource which is in heavy demand. Although the processing time at such a machine might not be the largest in the system, it will limit throughput by the very observation that inventory is piling up at that resource due to the high demand of the said resource. A few relevant definitions as discussed in [8]:

1. Congestion points occurring in the product flow.

2. The resource whose capacity is less than the demands placed upon it.

3. Any process that limits throughput.

A ‘common sense’ definition of bottleneck is anything that limits the production rate, but the bottlenecks need not be same under different definitions or detection methods as discussed next.

2.5.1.3 How to detect bottlenecks

After defining bottleneck, implementable methods of discovering the bottlenecks keeping the definitions in perspective are required. The 2 most discussed methods are the following:

1. Measuring the average waiting time:

In this method, the machine with the longest average waiting time is considered to be the bottleneck of the system.

$$B=\{i|\ W_i = \max(W_1,W_2,..., W_n)\}$$

In the above equation, B (the bottleneck machine index) is the the machine with the largest average waiting time, W_i. This method is suitable when the the intermediate nodes don’t have a preemption buffer i.e. a theoretically unlimited waiting time is allowed. As this is true in case of railways (the trains can theoretically wait for as long a time as it takes to proceed on its journey. It doesn’t have anywhere else to go!), this method is suitable for detecting bottlenecks in a railway section.

2. Measuring the average Utilization:

In this method, the machine with the largest busy time/total time ratio is considered the bottleneck section.

$$B=\{i|\ \rho_i = \max(\rho_1,\rho_2,...,\rho_n)\}$$

In the above equation, ρ_i is the utilization of the machine i. $\rho_i=\lambda_i/\mu_i$, where $\lambda_i$and μ_i are the arrival rate and the service rate of the i^{th} machine respectively. As seen here, the bottleneck doesn’t just depend on the service rates, but also on the arrival rate which is essentially the demand of the given machine. If the arrival rates are the same for all the machines, the bottleneck can be determined using just the processing times which are the inverse of service rates, thus recovering the result of 2.5.1.1.

Note that the 2 methods may not give the same bottleneck as it is not necessary that the machine with the largest waiting time observed has the highest utilization too. Next we understand how these concepts can be used to obtain throughput of linear railway sections.

2.5.2 Parallels between Railway sections and production systems

Consider the old version of the Allahabad-Mughalsarai section before the implementation of automatic signaling. The section is divided into loops and blocks by placement of signals:

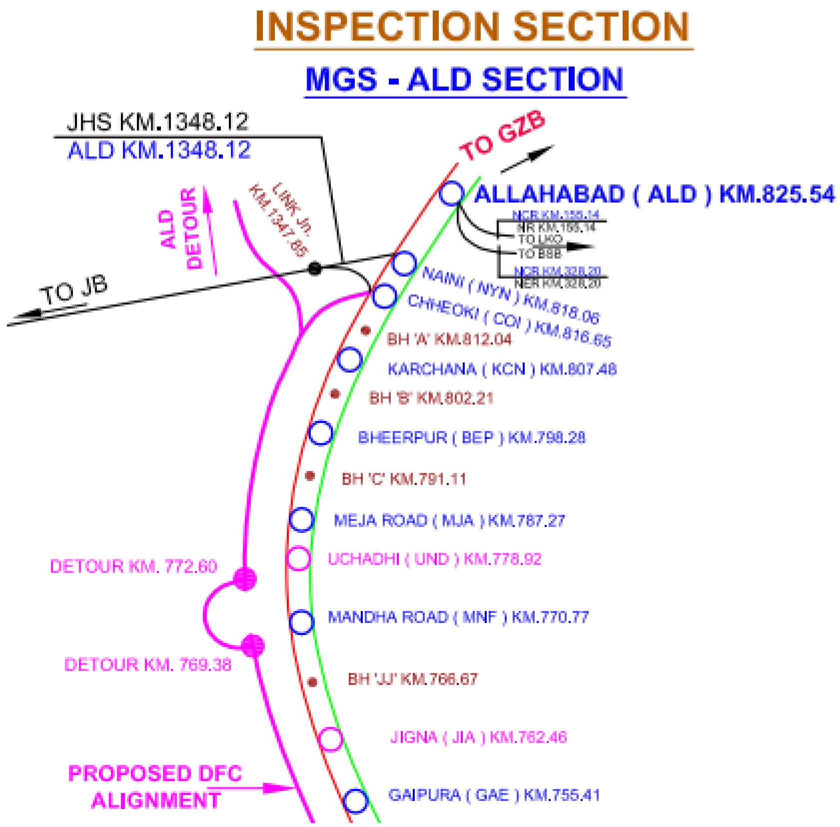

Figure 2.8: Part of the ALD-MGS IB signals

Only one train can be present on a given block or loop, thus each of the blocks or loops act as different machines of a production system. In a production system, the processing of a part at a subsequent machine starts only when the processing at a previous machine is done. This is not strictly true over here as the train can be present in more than one block/loop at the same time, but by and large it spends a major time in one block/loop only and the approximation is valid. The processing times are nothing but the traversal times through the blocks and the halt time + traversal time through the loops. Also, the pile of inventory (or the queue size) at any node is restricted by the number of blocks/loops just before the considered node. This is because the train can only wait in a block or a loop, and given that only train can be present on a given block/loop, the inventory size waiting to be 'processed' by a node is limited to the number of blocks/loops at that node. Thus, a linear railway line section can be reduced to a production graph as shown:

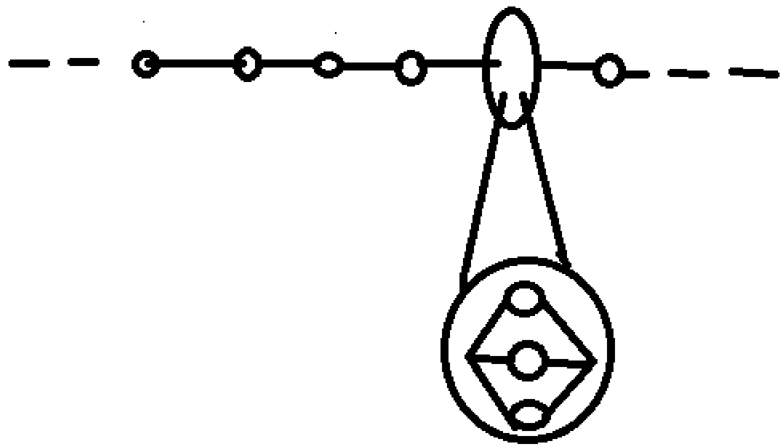

Figure 2.9: Graphical Interpretation of a railway section

The small circles represent the blocks (there may be more than 1 blocks between 2 consecutive stations) while the large circle represents a node at the station. Zooming in on the large node shows that it is comprised of smaller nodes itself, which are the loops at the stations. These nodes can be treated as multiple servers to do the same process. Their effective total rate is the sum of service rate of each of them. Only double line tracks are considered over here with no common loops at stations (as common loops are shared by trains going in both directions. A 50-50 split for each of the directions and an effective rate of half the original loop's can be assumed in order to include common loops.)

Thus, railway track can be abstracted to a production line with processing times as mentioned. Note that the demands may not be the same at all nodes as trains may enter a given node from some different part of railways and originate/terminate at a given node. Note that these can happen at the station nodes only. The nodes corresponding to the blocks between consecutive stations will have the same demand.

Given that throughput, or the number of trains the section can effectively handle in a day is limited by the bottleneck(s) of the line, we should try to figure out the same using the methods suggested in 2.5.1.3:

2.5.3 Limitations of Scott's formula

Scott's formula says that the capacity of a railway line is determined by the longest time for trains to move between any 2 consecutive signals, which will happen for the resource with highest processing times, which is not necessarily the bottleneck. For a block, this time includes traversal time and block working time, while for the loops, the time includes scheduled halt time along with the traversal time. Scott's formula also does take into account heterogeneity to some extent (by a factor, which is arbitrary, but people use with some experience) and also a block working time, to take into account the signaling technology in force. Let this combined longest time be T minutes. Then the capacity of the section according to Scott's formula is 60/T. A factor of efficiency is also included on this final number based on experience. It calculates this T using the path of a single train going through the section and identifies a node as a bottleneck based on this.

Limitations: Clearly, only 'processing' times or time to traverse a block/loop is included in this method. The demand at the each of the nodes which may or may not be the same, is not considered. This can result into choosing the wrong node as a bottleneck, which will happen in cases like the ALD-MGS section where different demands are placed on different parts of the track. Thus, using a single train's path to estimate capacity of the whole section gives us limited insight, which can in fact be flawed. Result of this flawed argument will be seen in chapter 6.

To overcome the limitations, the indicators as discussed earlier to determine the bottleneck in a railway sections should be used: Expected waiting times will increase while moving close to the bottleneck node (waiting time analysis method) and the loops and blocks near the bottleneck would be more or less permanently under use (utilization method).

2.5.4 ALD-MGS section: brief description and bottleneck detection

The ALD-MGS section has 21 stations including the terminals. There are trains which originate/terminate in the middle of the section and also trains entering from different sections into the considered section, especially at ALD, where 5 different lines converge, making it an extremely busy junction (high demand). Using the free running path of a single train, the bottleneck was thought to be the block between Dagmagpur to Pahara, which is an absolute block of 8 km. Based on the following facts, the real bottleneck was discovered to be the station terminal at ALD itself and not the considered block:

1. Upon asking experienced railway personnel at ALD section railway offices, they all told us that they believed the bottleneck to be the ALD station itself. This method of asking experienced people directly has been shown to hold merit by Cox and Spencer (Cox and Spencer 1997), but can't be used directly for quantitative research and design.

2. We got on a train from Chunar to Allahabad, 12321-Howrah-Mumbai CST Mail, which is a fairly high priority which was already delayed at Chunar by a couple of hours (details in chapter 6). Post Naini station, as the train moved towards Allahabad, the train experienced significant delays and waiting times, taking around 1.5 hours to traverse the final 5 km. It was met by a red at almost all signals, which indicated high

utilization of the nodes also. Thus, this actual travel also suggested that the bottleneck was close to the ALD station as suggested by the methods in 2.5.1.3.

3. Data analysis was also done to get an average picture: [9] was used to obtain average waiting time statistics for all the trains in the given section and observed that the delays indeed increased on an average as trains moved close to the Allahabad station, which confirms that it is the bottleneck. (details in chapter 6).

Key Takeaways of this abstraction of a railway line with a production system:

- The railway system can be approximated by the production line, with the nodes being the blocks and loops and the processing time being the traversal times.
- Using only processing times for calculating the bottleneck is not the right approach in general, thus Scott's formula is rudimentary.
- Indicators like average waiting times and utilization of resources/nodes should be used to determine the bottleneck or a railway track.

Chapter 3

Signaling technologies and impact on capacity

The first exercise conducted to measure the impact of signaling technologies on capacity was done on a standard 10 km section with 2 stations at the end. The tool used for the same was the IIT Bombay simulator tool. The tool is capable of handling sections with line side signaling only. Thus, automatic, intermediate and absolute block signaling can easily be modeled within the framework of the simulator, but cab signaling, in which there are no fixed blocks or line side signals, needed to be reconciled with the block signaling system. Thus, this reconciliation is explained first followed by the results of the exercise.

3.1 Modeling cab signaling in the simulator

As discussed before, the simulator implements block signaling for generating valid train schedules. The simulator assumes that the track is divided into blocks and compulsorily requires that the whole track is divided into blocks and loops. Note that for absolute and intermediate block, it assumes that the response time for setup of next signal is zero even though there may be some delay due to manual skills employed. Now:

Problem: Cab signaling system does not have fixed blocks per se. There are no line side signals involved with signaling which divide the track into smaller parts/blocks. Thus, the question arises as to how to model this signaling method with the simulator's logic.

Claim: Discretization: Divide the track into a large number of arbitrarily small blocks. Also ensure: Number of colors/aspects of simulation (col) = number of blocks (n) + 1.

The number of signal aspects, in automatic signaling with 4 aspects are: Green, double yellow, yellow and red, tell us how many blocks ahead of a given train is the train before it, as was seen in 2.1.3. One can theoretically have as many colors as one desires to get more information about the train in front of it in a similar way in the simulator's logic. Next we prove our claim.

Proof: The basic reason for adopting cab signaling is that it instantaneously clears the region which was previously occupied by the rear of any train for the train following it. In block system, the whole length of the blocks needs to be cleared before the next train can enter that patch of the track.

If block lengths are kept infinitesimally small, then the rear of the train will clear consecutive blocks very quickly, thus making it available for the following train in almost the same time as they are cleared, as is desired in cab signaling. Thus, it seems that a large number of small blocks should closely approximate cab signaling.

Coming to number of aspects: What cab signaling really means is that the train has real time data about the location of the train in front of it. Let's assume that there are very small blocks but only 2 signal aspects i.e. red and green. This does not capture the cab signaling behavior we want to model as although the trains clear consecutive blocks very quickly, a train can only know whether the one ahead of it has cleared at least the block right in front of it or not. It cannot determine the location of the train.

To mitigate this, as many colors as required should be used to accurately track the train ahead. One can easily see that we don't need more than n+1 colors as even if we did, the additional colors would never be used, as there are not enough blocks to see the remaining colors (red (1), Y (2), YY (3) up to n+1 colors may be observed, but none beyond that, as there are no blocks). Thus, we set col=n+1.

Essentially, with this setup, the next train can accurately locate the previous train's rear position within an error distance of L (+/- L/2 on either side of the middle of the block), where L is the size of the block the previous train's tail is in. As we use smaller and smaller blocks, this error tends to 0 and we closely approximate a feasible cab signaling solution. □

3.2 Standard section exercise

This exercise was conducted to observe the effects of capacity on a gradient free, speed restriction free track, where the trains are uniform. The capacity definition used was to calculate the maximum number of trains that can be operated in a day, regardless of punctuality, average speed or heterogeneity or other factors, while assuming that new trains are always available for firing when there is a free train path. The train and section settings, different simulation settings, results and conclusions of this exercise are described next.

3.2.1 Train and track settings

- Maximum velocity of train =25 m/s =90 km/h
- Acceleration = 0.2 m/s^2
- Deceleration = 0.2 m/s^2
- Size of train = 0.5 km (Note: The train can only use the specified acceleration and deceleration in the simulator and these are not indicative of a range. Also, the velocity profiles generated ensure that the train maintains the maximum velocity possible for as long as it can, accelerate as fast and early as possible and decelerate as late and quickly as possible.)
- Maximum velocity on all blocks and loops, station entry velocity = 27.78 m/s = 100 km/h
- Distance between the 2 stations = 10 km (This is the block length in absolute block signaling)
- Loops size (same as station size in the simulator) = 1 km (changed for cab signaling)

3.2.2 Settings for different simulations

1. Absolute block: Just one block between station A and B. Block working time of 1 min considered.
2. Intermediate block signaling: 2 blocks each of size 5 km. Block working time of 1 min considered.
3. Automatic block signaling: 10 blocks each of size 1 km with 4 color aspects.
4. Cab to cab signaling: Blocks and loops of sizes 100 m each, thus total of 120 blocks and loops. Numbers of colors/aspects were kept at 121 using the criteria of c=b+1 as discussed previously. The safety distance for cab signaling is taken as 100 m in this analysis. Safety distance is the minimum distance to be maintained between a train's rear and the next train's front. This distance was implemented by increasing the length of the train to 600 m.
5. Train settings: 5 identical trains as defined by the parameters in the previous subsection scheduled in such a way that the next enters the loop at A and is ready to get on the track as soon as possible. This is achieved by keeping their scheduled departures from the station within a minute.

6. Simulation time = 1 day.

3.2.3 Results of exercise

Typical nature of velocity profiles of trains for different signaling systems experiment:

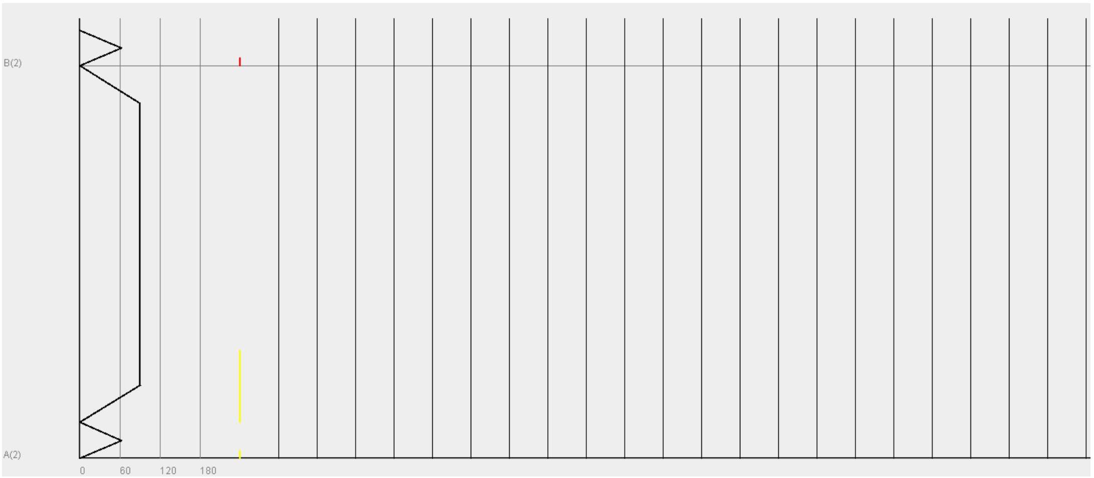

Figure 3.1: Absolute block signaling - standard section

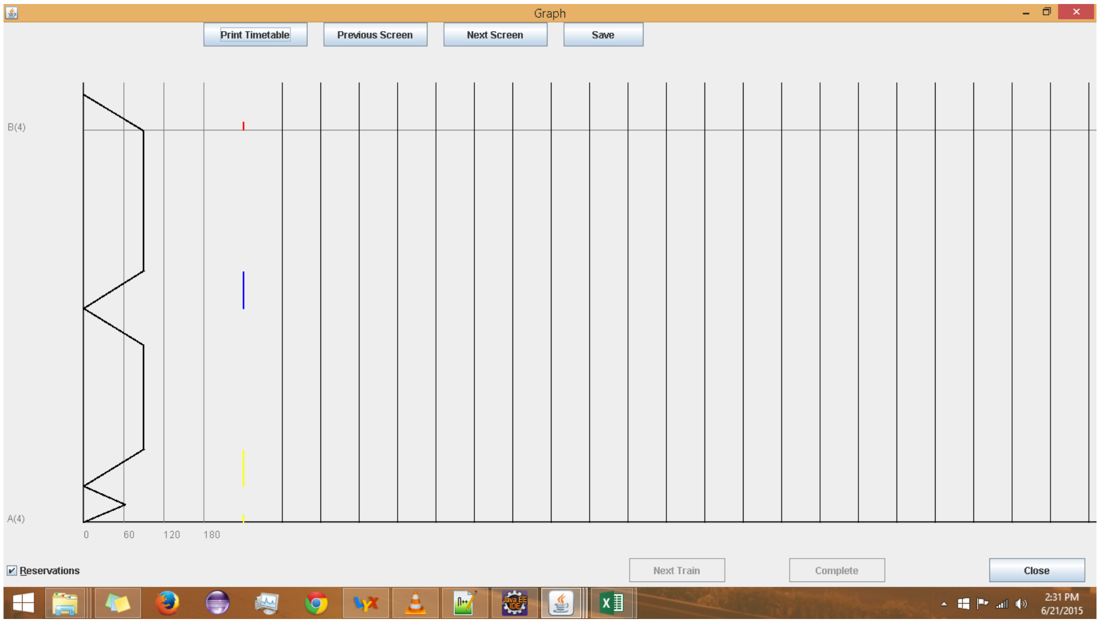

Figure 3.2: Intermediate block signaling - standard section

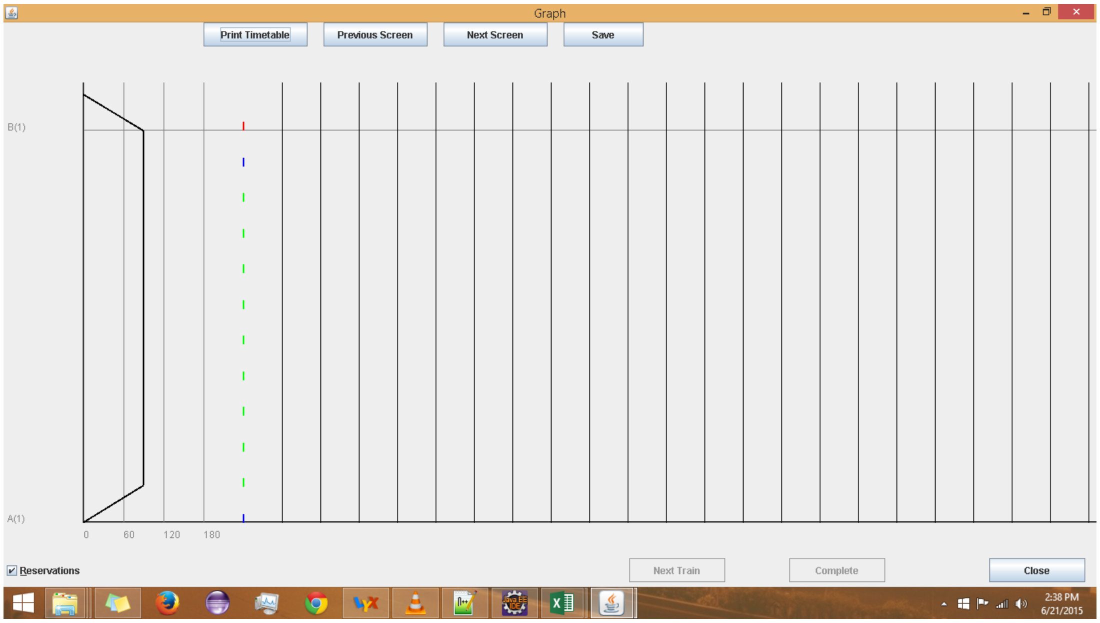

Figure 3.3: Automatic signaling - standard section

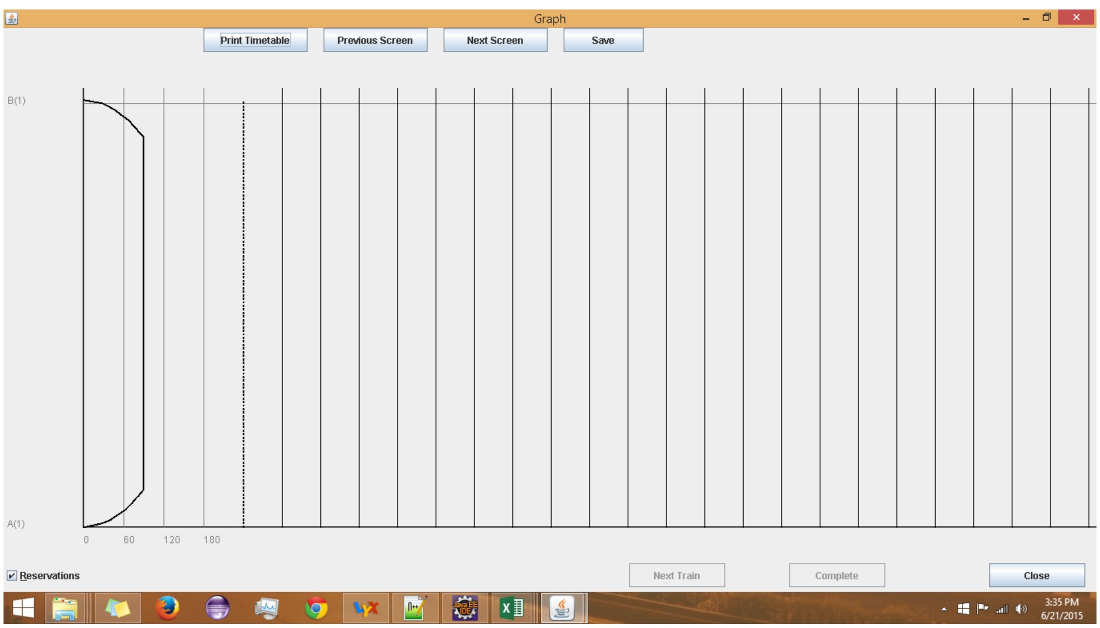

Figure 3.4: Cab signaling - standard section

The following table summarizes the results obtained from the analysis of the schedules generated by the simulator:

Signaling System	Latency (min) (departure at B - departure at A)**	Headway (min) (time difference to leave/arrive between 2 consecutive trains	Line capacity (trains/day using Scott's formula) (60% efficiency)*	Increase in capacity wrt absolute block	Increase wrt IBS	increase wrt automatic block
Absolute Block Signaling	11.1	10.05	85	NA	NA	NA
Intermediate block signal	11.5	6.75	128	50%	NA	NA
Automatic Block Signaling	8.4	3.2	270	217%	111%	NA
Cab signaling	9.5	2.5	346	307%	170%	28%

Table 3.1: Standard section results

*This is an empirical factor used in Indian Railway literature like [1] **Wait times at station B are 0: trains just pass through. We include the departure time within the latency.

3.2.4 Standard section conclusions

1. Waiting time, speed restrictions, loop entry velocity etc. aren't considered in this study. Also, it is assumed that trains are available at all times for leaving the station and they are scheduled the moment the block ahead becomes free. Thus, the capacity observed is much higher that what would be observed in practice for all forms of signaling.

2. The simulator schedules trains according to a FIFO basis and schedules trains immediately. This explains the output obtained in intermediate block signaling: trains will wait at the end of the first block. If they had been fired a certain time later, they would have found the second block also free. Thus, the capacity obtained is under the strategy described and don't correspond to the normal operating conditions of trains.

3. The capacity of cab to cab signaling is over 3 times that of absolute block signaling. This is a huge improvement and shows the potential it has if implemented on real sections currently using absolute blocks. Cab signaling gives a similar order of capacity as auto blocks of 1 km. A similar result is obtained in chapter 4 also. Thus, a cost benefit analysis should be done for the same (auto to cab) before implementing it on a real section.

3.2.5 Cab signaling and completely coupled trains

In the last section, the simulator was used to analyze capacity with a FIFO and leave as soon as possible rule. The question arises as to how should cab signaling or any other signaling should be used if it has been implemented on a section. There are a lot of feasible strategies which can be employed. Maintaining minimum latency can be one criteria. Another possible policy/way for using cab signaling is to keep the trains as close to each other as possible (braking distance) to extract more capacity. This basically means that trains are separated by the braking distance determined by the maximum velocity of the trains and a safety distance. Although this is practically never done by design on any railway track, we explore the idea here as a thought experiment. Railway operations managers prefer trains to run freely and try to obtain maximum capacity while trying to ensure that sections aren't getting too 'choked'. Despite of this desire, completely coupled behavior as described is naturally seen around choked bottlenecks on IR from time to time, and this analysis can be useful for understanding such situations.

Steady state system is assumed for this analysis of obtaining latency and capacity in such situations. Details and procedure used:

1. Two consecutive trains will be separated by the braking distance as determined by the maximum velocity plus an additional safety/clearance distance. This distance will also be kept minimal. This means that if the train ahead decelerates, the one behind will also necessarily have to decelerate.

2. Determine the maximum number of trains that can be running on the track given that point 1 is met. Let's say this number is N(b), where b is the braking distance. In general, we can assume that of these N(b) trains, one train can be entering station B while one is leaving A. The other N(b)-2 trains are completely on the track.

3. Under conditions 1 and 2, we have that the train coming next on the track from A will have to accelerate from 0 to max velocity (call it v(b)) and decelerate from v(b) to a halt (N(b)-1) times as all trains on the track are 'coupled' given 1. The rest of the distance (while not accelerating or decelerating) will be covered by this train at the maximum velocity v(b). Thus, we can find the time required for this train to move from A to B and also calculate the capacity as we shall see next. The analysis is based on back of the hand basic approximations and may not hold exactly in all situations.

3.2.5.1 Symbols for the exercise

Basic variables:

b = braking distance

c = clearance distance (trains can never be closer than this distance for safety)

t = track length

l = train length

D = acceleration/deceleration of the train. We assume them to be same for simplicity.

Derived parameters:

v(b) = maximum velocity achievable corresponding to the braking distance b.

N(b) = number of trains on the track as a function of b.

T(b) = latency or the time taken by train to travel from station A to B

T_1(b) = time spent by the train either accelerating or decelerating.

T_2(b) = time spent by the train traveling at v(b).

T_a = time to accelerate

T_b= time to decelerate

C(b) = capacity of section as a function of the braking distance.

Note: We must have T(b)= T_1(b) + T_2(b), as there are only those two cases over here, the train is either accelerating, decelerating (corresponding to T_1(b)) or moving at v(b) (corresponding to T_2(b)) ... **(1)**

3.2.5.2 Derivation

1. To derive v(b): (using elements of [5])

In the extreme case that the train ahead comes to an instantaneous or stone-wall stop, the train following it should be able to brake from v(b) to 0 within a distance b. Using elementary constant acceleration/deceleration kinematics from classical physics, we must have that: $v(b)^2 - 0^2 = 2*D*b$.

Thus, we have v(b)=√(2Db). ... **(2)**

2. To derive N(b):

Assume that in the general case one train is just leaving the track while the other one is entering it. Thus of these N(b), at least (N(b)-2) trains will be completely on the track. For each of these trains, there is a distance (b+c) behind its length l. The train at the front will also have a distance of (b+c) on the track which separates the first of the (N(b)-2) trains. To derive N(b), we note that our policy assumes that the trains are packed as closely as each other and that even a single more train can't be allowed on the track Thus, we must have:

$(b+c) + (l+c+b)*(N(b)-2) \leq t < (b+c) + (l+c+b)*(N(b)-1)$

The inequalities corresponds to the fact that at most N(b) -2 complete trains are on the track on the track.

Using the fact that N(b) is an integer, we get that N(b) = 2 + [(t-b-c)/(l+c+b)] ... **(3)**

Where [x] is the floor function of x, or the largest integer smaller than or equal to x.

3. To derive T(b):

a) $T_1(b)$: Train accelerates and decelerates a total of 2*(N(b)-1) on its journey from A to B. Again from constant acceleration/deceleration kinematics, we have that Final velocity – initial velocity = Acceleration*(time taken to go from initial to final velocity) Using this to find T_a and T_d

$v(b) - 0 = A*T_a$ and $0 - v(b) = -A*T_d$

Thus we get Ta = Td = v(b)/A ... **(4)**

Hence, $T_1(b)$ =2(N(b)-1)v(b)/A ... **(5)**

b) $T_2(b)$: time spent by train moving at constant speed v(b):

It's evident that the train covers the same time to accelerate or decelerate from 0 to v(b) or v(b) to 0 respectively as acceleration and decelerations are assumed to be same. And this distance is precisely the braking distance. Thus, the train travels a total of 2(N(b)-1)b distance while accelerating or decelerating. Thus, it travels (t + l - 2(N(b) - 1)b) distance at constant speed v(b). Note that we include t+l in the last expression and just t as the train's head has to clear an additional l distance before it can said to have cleared the section. The train leaves only when its complete length arrives.

Hence, $T_2(b)$ = ((t + l - 2(N(b) - 1)b))/v(b) ... **(6)**

Note: We require that $t + l \geq 2(N(b)-1)b$. Otherwise, T2(b)=0.

(1), **(5)** and **(6)** $\implies$

T(b)= 2(N(b)-1)v(b)/A + ((t + l - 2(N(b)-1)b))/v(b) ... **(7)**

4. To derive C(b):

In time T(b), N(b)-1 complete trains enter station B. Also, the steady state assumption means that this is repeated continuously and this can be effectively used to compute the capacity.

C(b)=((N(b)-1)/T(b))*1day

Note: This capacity problem becomes an optimization problem wrt the variable b. Assuming the other variables are all provided to us and can't be modified, we can vary the braking distance/maximum velocity in order to get the maximum capacity.

Using the same parameters as used in the simulator exercise in 3.1, we get the following results:

Latency = T(b) = 20.83 min

N(b)= 6

Capacity = C(b) = 345 trains a day.

3.2.5.3 Comparing the 2 cab signaling solutions

Important parameter values	Using simulator's output (decoupled free flow)	Using completely coupled flow done in the calculations
Latency (minutes)	9.03	20.83
Capacity (trains per day)	346	345

Table 3.2: comparing cab signaling policies

This is similar to a traffic conditions: Vehicles crawling behind one another. The latency is much higher in this case than the output of the simulator, but capacities are similar. This is because the parameter values are such that extra vehicles on the track compensate the higher latencies and give similar throughput. Despite of this, it can be shown that such a running strategy can give better throughput than the simulator's strategy if the acceleration and deceleration parameters are extremely high i.e. braking distances are small.

As mentioned before, no operation manager wants trains to run in the coupled manner described above and desire free flow of trains with small latency. In congested sections like ALD-MGS, there are 4 automatic signals between the penultimate station Naini to Allahabad, thus the station pair is divided into 5 blocks and the distance between them is 7.44 km. While moving towards Allahabad from Naini, it was observed that trains halt at each and every block of this section, which indicates that each and every block contains a train. Thus trains are moving in an extremely coupled fashion, which although is undesired, is seen in extreme congestion cases. This is akin to traffic on a congested road.

3.3 Secunderabad Wadi section exercise

SC-W is a section in South Central Railway (SCR) zone of IR. This section has 21 stations with absolute blocks. Few station pairs have intermediate blocks also. Station jurisdictions, loop configurations, gradients and signal information were provided within SC division rolling diagrams, route wise indexes.and the working time table[11]. We were provided with the gradients on the track, the loops at the stations. Details on other parameters were assumed as follows:

Length of train: 500 m

Acceleration: 0.5 km/min^2

Deceleration: 1 km/min^2

Maximum velocity: 90 km/h
Loop velocity: 20 km/h
Block velocity: 100 km/h
Gradient effects: As mentioned in table 8.1
Capacity of section was defined again in a manner similar to the last section: What is the maximum number of trains of the type mentioned that can be pushed in a day without halting at intermediate stations and just going through from start to end of the section. Waiting time at stations wasn't considered here either. The current signaling system is intermediate block signaling and we want to measure the impact of cab signaling on the section. Now, the SC-W division is around 194 km long. To simulate cab signaling on this track would require roughly 1700 small blocks each of size 100m and correspondingly 1501 colors (assuming loops of 1 km each). As the version of simulator on which this exercise was conducted had manual data entry directly into text files involved with the back end of the code, this was infeasible. To resolve it: As seen on the standard section, even 1 km blocks with 4 signal aspects gave the same output as cab signaling. We used this result over here and approximated cab signaling with 1 km blocks and 4 color aspects on the section. 21 trains ready to leave as soon as a path is available were used again in this exercise. Typical velocity profile generated on the track and the results of the exercise are:

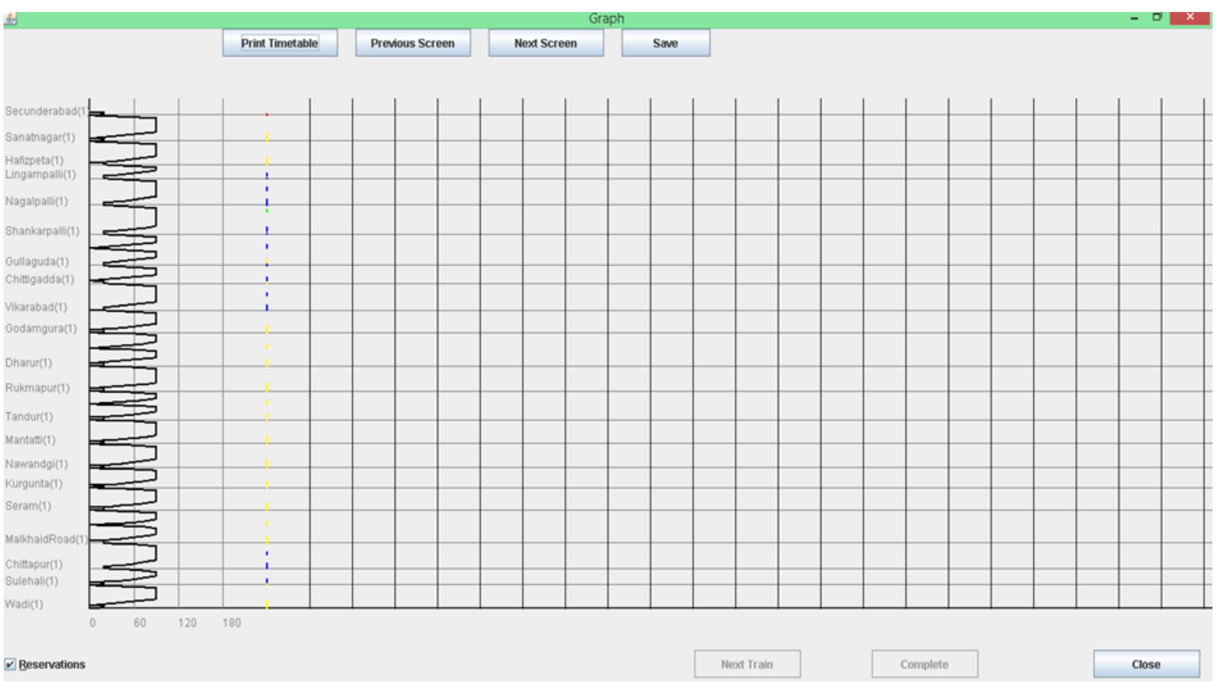

Figure 3.5: Velocity profile on SC - W division

Signaling Method	Average Latency (hours) of 21 trains	Capacity (trains/day)
Existing Intermediate Block	3.6	140
Cab Signaling/Automatic Block	3.2	267

Table 3.3: SC - W intermediate and auto/cab signaling

Thus, with mentioned assumptions and definition of capacity, there is a 90% increase in capacity on a real section of IR by adopting automatic/cab signaling over intermediate blocks.

3.4 Capacity analysis with mixed traffic and punctuality constraints

The capacity definitions used so far are simple and don't consider complications mentioned in 2.3. The next exercise conducted was to come up with a definition of capacity which can be used in a practical scenario with mixed traffic and punctuality requirements along with waiting time at stations. The definition adopted was the following: Capacity is the maximum number of freight trains that can be run on a section and a scheduled trains time-table given that the average running time of these freight trains is less than twice the free running time of a freight train. [12]

The SC-W section was considered again for this exercise. The time table was obtained from [11]. A single freight train is run initially for calculating the free running time through the section. After this the time-tabled trains were added to the model. Following this an initial set of 6 freight trains (3 in both directions) spread uniformly across a day's operating time were added. The arrival times of freight trains at origin station were prohibited to be from12:00 to 4:00 for maintenance purposes. We compute the average traversal time for these trains and if it comes out to be less than twice the free running time we perform next iteration. For every iteration we increase the number of freight trains that run on the section till we do not cross the traversal time limit. After doing that we then converge to a number which is our capacity (according to the modified definition) of the section with the particular signaling technology.

Automatic signaling was used with blocks of 2 km length while cab signaling was considered with blocks of 1 km length. Passenger trains with a maximum speed of 75 km/h and freight trains with max speed of 60 km/h were considered for this exercise.

3.4.1 Results of the exercise

The results obtained for different signaling technologies on the Secunderabad Wadi section (iterative technique mentioned above in 3.4 applied to determine the section capacity) are tabulated below:

Number of trains	Average Latency (min)	Upper bound* (min)
1 (free train)	197	394
6	274	394
12	312	394
24	334	394
80	353	394
90	382	394
96	412	394

Table 3.4: SC - W intermediate signaling iteration results
*(determined by travel time of 1 freight train without any passenger trains i.e. free running time of freight train.)

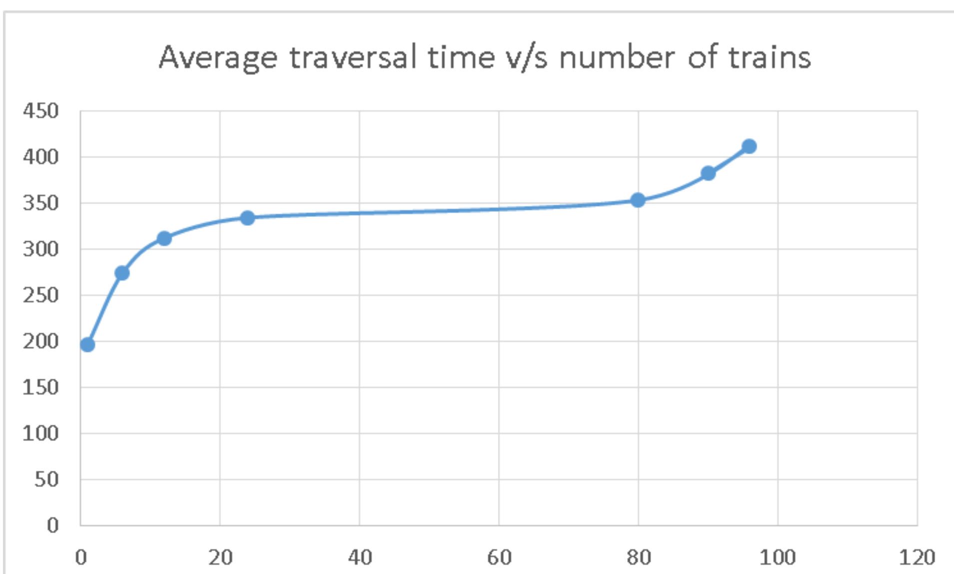

Figure 3.6: IB: Traversal time v/s number of trains

Thus the capacity is ∼90 freight trains a day (combined for both the directions) for the existing intermediate block signaling system according to the simulator with uniformly spaced freight trains.

Number of trains	Average Latency (min)	Upper bound (min)
90	234.2	394
96	234.6	394
160	258	394
480	824	394
240	335	394
300	456	394

Table 3.5: SC - W automatic signaling iterations results

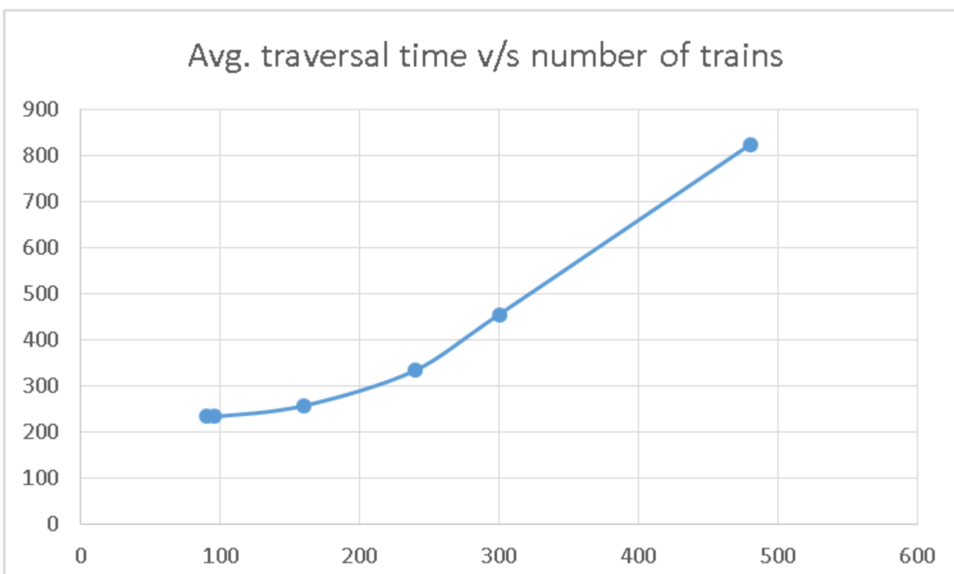

Figure 3.7: Auto block: Traversal time v's number of trains

So the capacity is ∼269 freight trains (using a linear approximation between 240 and 300 train results) in a day for the proposed automatic block signaling on the Secunderabad Wadi section.

Number of trains	Average Latency (min)	Upper bound (min)
300	290	394
480	Simulation didn't terminate	394
400	Simulation didn't terminate	394

Table 3.6: SC - W cab signaling iteration results

Signaling Method	Line capacity (freight trains/day)	Increase wrt IBS
Existing Intermediate Block	90	NA
Automatic Block	269	199%

Table 3.7: Comparing automatic and intermediate signaling

Thus, the study shows that capacity can increase by almost 200% by adopting automatic blocks instead of intermediate block signaling (in both directions). Limitations of this procedure: It doesn't take into account the stochastic nature of freight train arrivals. In principle, Monte Carlo methods should be used with the simulator to obtain the average waiting time at a given arrival rate and distribution of arrivals. The simulator doesn't have a function of multiple calls and random number generators within it and the only resort currently is to do this manually, which is time consuming. Thus, only 1 simulation was done with every arrival rate of trains in this exercise. Also, the section will always be choked with trains in this method as freight trains are departed at every possible time slot, which is not the case in reality. The modeling of yards/complex stations, which themselves could be bottlenecks and limit throughput, isn't included in this exercise.

Chapter 4

Automatic signal spacing algorithm and cab signaling

The basic idea behind the automatic signal spacing exercise was described in 2.4. This chapter shall describe the idea for the same algorithm, results on western corridor using it, how to model cab signaling using the same pseudo-code, results obtained for cab signaling and then finally comparing the 2 technologies. The definition of capacity used in this study was using Scott's formula with homogeneous and periodic traffic. We would like to thank Professor Abhiram Ranade of Computer Science and Engineering Department, IIT Bombay for sharing with us the code of the algorithm mentioned next.

4.1 Signal spacing algorithm

This algorithm was developed by the authors of [13]. It includes details including acceleration, deceleration, halt time, block overlap, sighting distance of signal and length of train. Speed restrictions within blocks were not considered as part of this exercise. The basic outline of the algorithm is as follows:

1. For loop: (vary station block length L_1 from 100 to $o+l_t$)
   Station length should be at least 100 m as smaller station sizes are generally not observed and people have to be given sufficient place to stand. Station block larger than $o+l_t$ doesn't need to be considered as increasing beyond this will increase H_2 and keep H_3 the same, thus increasing H (=max$\{H_2, H_3\}$), which is undesirable as the idea is to minimize the headway.

2. Create 2 new blocks: The pre station and pre-pre station block.
   This uses the safety measure of $s_i + L_i \geqslant$ braking distance. As minimum headway are wanted at these locations, strict equality is set (smaller the block size, smaller the time

to traverse it). A function called brakefrom(x) is used which calculates the position (and hence the braking distance) from which the train needs to break so that it will reach at x. The block locations are found using this function.

3. Find YY headways H_2 , H_3.
   This uses a function called T(x,y), which calculates the time it takes from x to y in the path of the train under normal operating conditions of a train (when the train is moving freely according to its acceleration, deceleration and top speed criteria and unaffected by network effects. This is precisely the minimum latency requirements). x here will be the location of the signal and y will be the location the train has to reach for signal at x to turn YY again.

4. Print block lengths and the headways obtained.

5. Repeat for all station lengths in the for loop.

6. Using this, H, the minimum headway across the section can be calculated as:
   H = min (max$\{H_2, H_3\}$ obtained for the different station lengths station length).

4.2 Results for Western Railway (suburban corridor)

4.2.1 Parameter values for the exercise

The following train and section characteristics were used for this study. These are in line with the operations in Western corridor suburban traffic in Mumbai as mentioned in [13]:
Booked speed (v) = 18 m/s. This is the max operational speed on the section currently.
acceleration (a) = 0.5 m/s^2
deceleration (d) = 0.3 m/s^2
block overlap (o) = 120 m
sighting distance (s) = 200 m
Waiting time at station (W) = 20 s
Length of train (l_t) = 258 m for 12 rake cars

4.2.2 Results for automatic block signaling

The inter station length in the western corridor varies from around 800 m to 1800 m ([7]). The headway between 2 consecutive stations as a function of inter station length at fixed booked speed of 18 m/s was analyzed first:

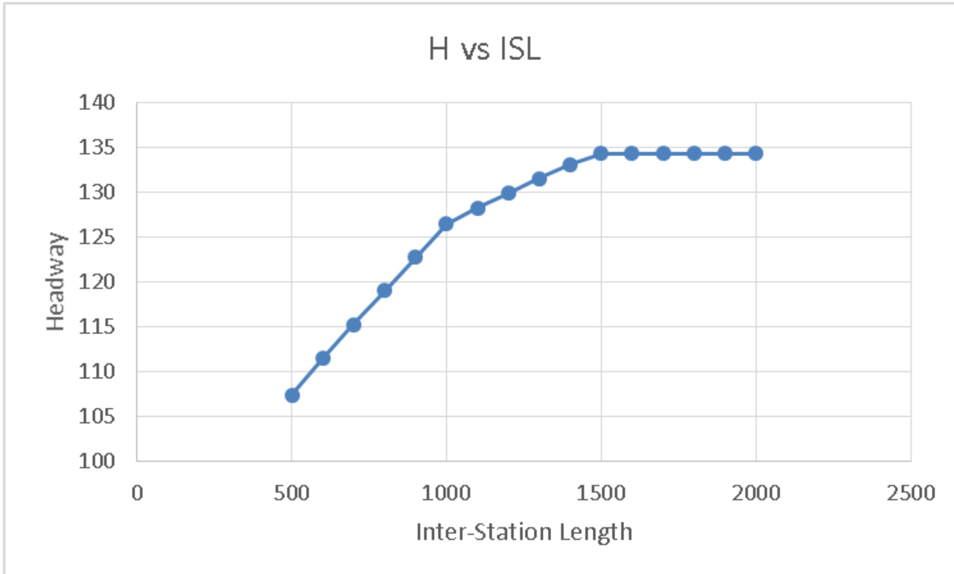

Figure 4.1: Headway v/s inter-station distance in auto block signaling

As seen in the figure, the headway doesn't change beyond a certain Inter station length. This is expected as the factors dominating the headway are seen at the end of the station. Pushing the starting point beyond a certain point will not change what happens towards the end. This is because larger lengths between stations just translate to a larger time spent at constant speed, and by designing the intermediate auto blocks small enough, we can almost always ensure that the bottlenecks are always at the end and not in the middle. Thus increasing inter-station distance beyond a point will just mean larger numbers of blocks but not a larger headway. The end looks the same beyond a certain ISL. In case of small inter station distances, the train may not accelerate completely before it starts decelerating and have effects on the end. The block sizes also stabilize after a certain inter station length as expected:

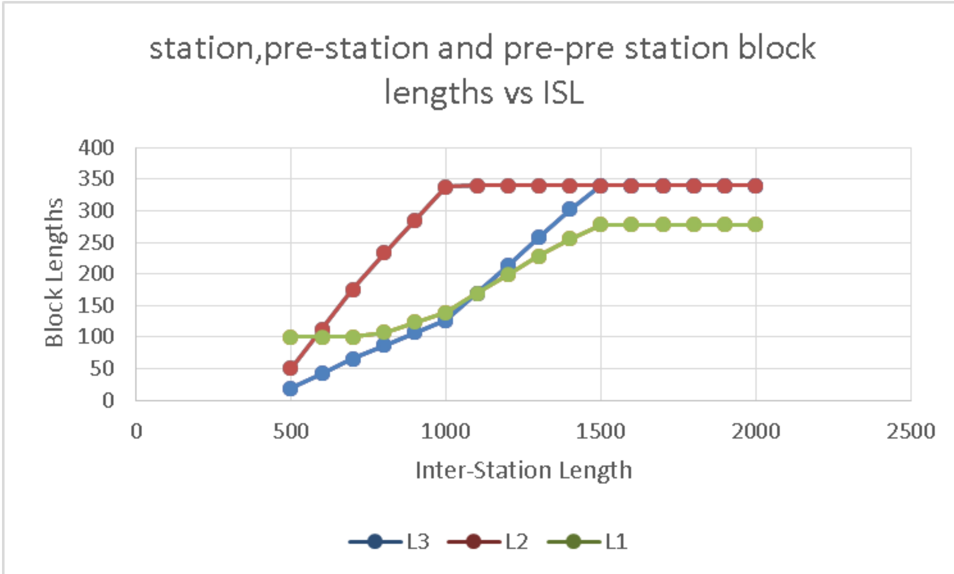

Figure 4.2: Optimal Block lengths sizes vs inter station distance

The prime takeaway from these results is that the best possible headway that can be achieved with automatic signaling and running trains with minimum latency on a YY headway on the western suburban corridor is 134.33 seconds as that is the maximum headway across all possible inter station lengths. We'll see in the next section that headway with the same strategy in cab signaling comes down to 117 seconds.

The variation of headway with booked speed (at fixed inter station distance of 2000 m) was also observed and as expected (details in 2.3.3), headway increases as max/booked speed increases:

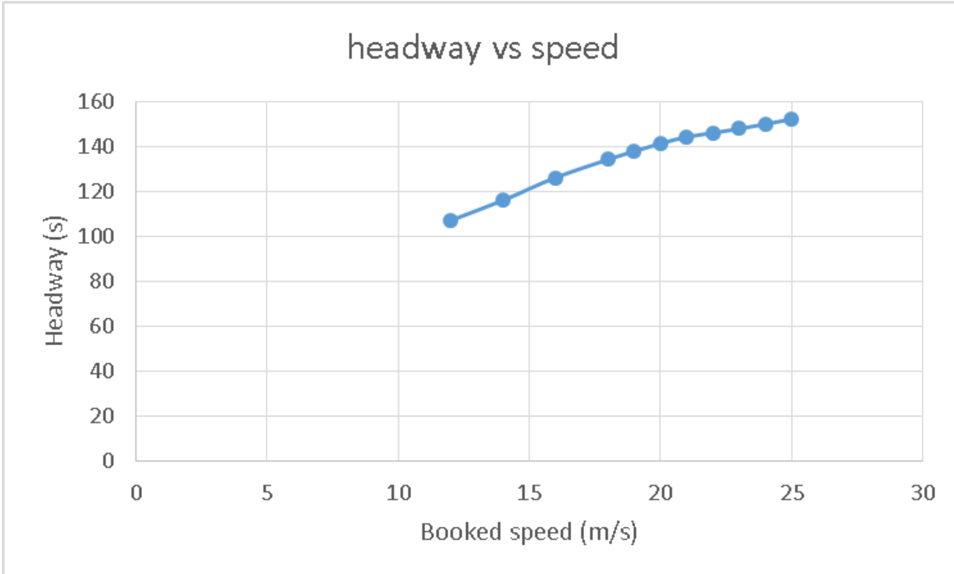

Figure 4.3: headway vs booked speed in auto block signaling

This algorithm was developed next to handle cab signaling: The idea was to calculate the minimum headway for normal operating mode of trains with cab signaling and quantify the improvements in throughput with respect to the best that can be done with auto block signaling. The logic is explained in the next section.

4.3 Cab signaling algorithm using brakefrom and T functions

The code for obtaining cab signaling results using the 4 aspect signaling code described 4.1 was implemented. The main idea was to use the brakefrom and T functions of the earlier algorithm and get the headway at each location of the track, and thus choose the maximum of all of these headways as the system headway.

1. for(int x=(l_t+1); x<(sl+l_t+1); x++)
   Here, x denotes location of the head of the train and sl is the inter-station length. x starts at 0 and increases as the train moves forwards. The idea is to obtain the headway at each and every location along the track by first calculating the size of the 'shifting block' at every location using the brakefrom function and then calculating the headway at that location by using the T(x,y) function described in 4.1. This is precisely what is done in step 2. As the position of the rear of the train may vary from 1 to sl (we're discretizing locations over here for simplicity without loosing the basic idea, in principle the position varies as a real number, not integer.), the position of the head (x in this case) varies from

2. h[(x-ltint-1)] = T(brakefrom(x-lt),x)
   Array h stores the minimum headway at different locations along the track under the normal operating conditions strategy which was used for auto signaling. Explaining the idea again: brakefrom finds the closest location to break from in a way that collision doesn't happen between the next train's head and the previous one's tail (which is at x-lt as the head is at x). T finds the usual time it takes the time to go from the brakefrom distance to x in normal conditions. If the train reaches at the brakefrom distance sooner than this time (as in sooner than this usual time), the normal operating conditions may get violated. This is the idea behind shifting blocks: Find the minimum block length ending at each possible location of the end of the train and assign the headway at that location as the time it takes in normal operating conditions to cover this block shifting dynamically with the position of the train. This is the same idea as used in 4.1 when the block's were fixed and headways at only these locations were found. In order to understand the idea better, please refer to the starting sections of [5].

3. print << (x-lt) << ' ' <<h[(x-ltint-1)]<< ' ' <<brakefrom(x-lt)
   Obtain all the locations and the headways and the shifting block sizes at those locations.

4. max over x {h[(x-ltint-1)]} is the headway for the given inter station distance. Maximum is used to detect the bottleneck location using the theory described in 2.5.1. This is the same as choosing the max{H_2, H_3} done in step 6 of 4.1, but there are a lot of blocks now, namely the shifting blocks at every location.

This is how the headway is calculated for cab signaling given train characteristics and inter-station distances. Next, the results for cab signaling were obtained and were compared with those obtained with automatic signaling.

4.4 Results of cab signaling and comparing with auto blocks

The same parameters values for a,d and W were used in this exercise. In cab signaling, a safety distance is also desired between 2 consecutive that should never be breached. This distance was kept at 100 m in this exercise (the current safety distance on Western corridor with auto signaling is kept at 110 m). This distance was incorporated by increasing the length of the train by a 100 m, as practically the length of the train and this additional distance serves as the length for calculation purposes. Thus, l_t= 258+100 = 358 m for this exercise. Headway versus inter station distance (at booked speed of 18 m /s) and headway versus booked speed (at a fixed inter station distance of 2 km) were compared in both the cases.

The following results were obtained:

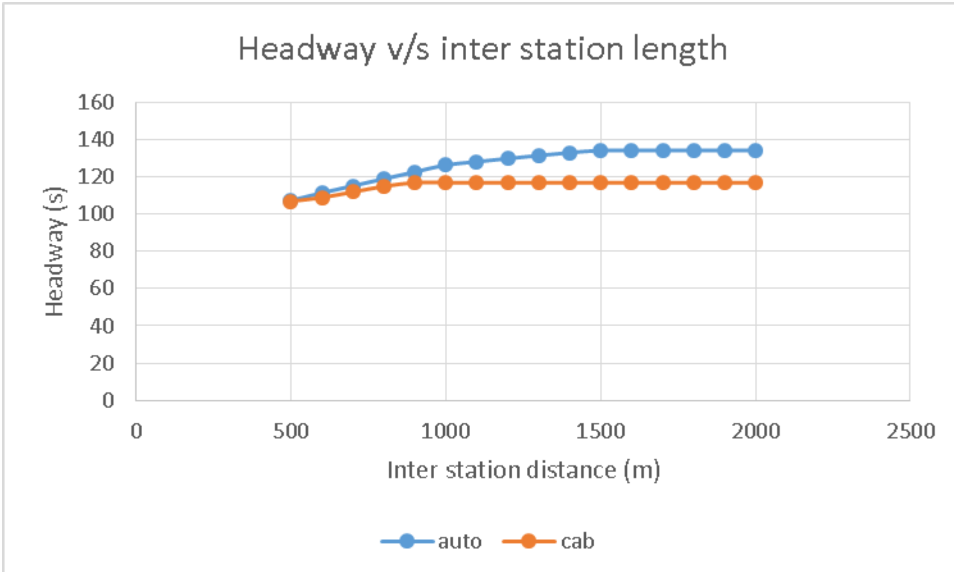

Figure 4.4: Headway v/s inter station length comparison for auto block and cab signaling

Inter station length (m)	Auto signaling headway (s)	Cab signaling headway (s)
500	107.419	107
800	119.036	115
1100	128.273	117
1400	133.141	117
1700	134.333	117
2000	134.333	117

Table 4.1: Headway v/s inter station length comparison for auto block and cab signaling

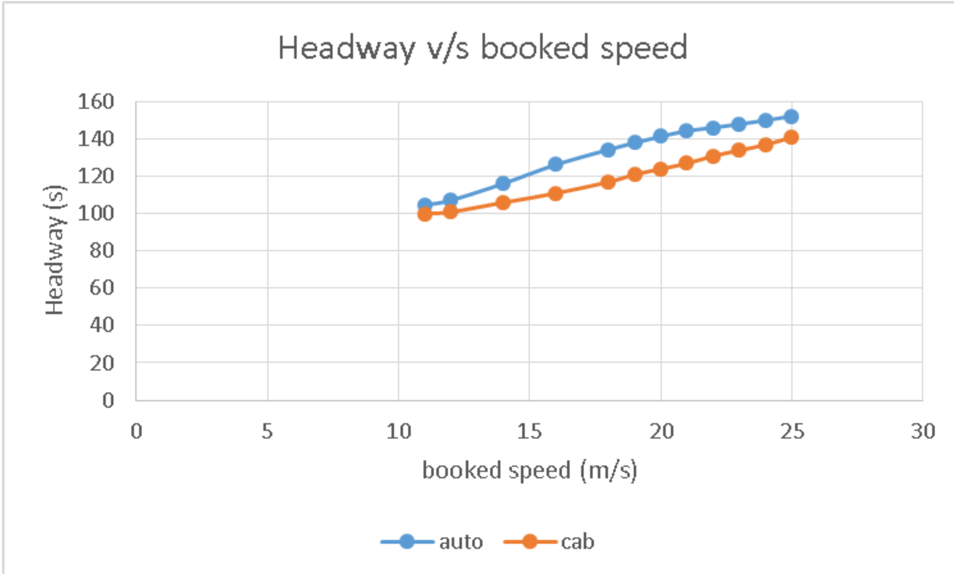

Figure 4.5: Headway v/s booked speed comparison for auto block and cab signaling

booked speed (m/s)	Auto signaling headway (s)	Cab signaling headway (s)
11	104.62	100
12	107.218	101
14	116.204	106
16	126.301	111
18	134.333	117
19	137.987	121
20	141.451	124
21	144.408	127
22	146.161	131
23	148.037	134
24	150.093	137
25	152.298	141

Table 4.2: Headway v/s booked speed comparison for auto block and cab signaling

4.5 Takeaways and limitations

The key takeaway from this analysis is that the best headway on Western corridor possible after repositioning signals at the locations found by our algorithm under normal operating characteristics is 134.333 s, while the headway with cab signaling is 117 s. This results in a 14.8% increase in throughput in terms of number of trains using Scott's formula.

Limitations of this analysis: The train path under normal operating conditions is assumed to be that of a train accelerating to max speed, moving at the max speed for as long a time as possible and then decelerating as late as possible just to reach the station at the other end. There may be speed restrictions on the track in reality which are being ignored. Including them may change the headway calculations. It is actually likely that despite of restrictions, unless these speed restrictions are close to the station, the analysis result will hold the same as what we have done over here: If there is a speed restriction somewhere in the middle of the section, the headway for blocks containing this restriction (in case of auto signaling) or shifting blocks containing parts of this restriction area is bound to increase. Now the overall headway of the system is determined by the max of all possible headways. (H_2 and H_3 in auto case and max over all shifting block headways in cab case) If the restrictions are not near the station, it is likely that the the headway of blocks (fixed or shifting) near the station will be largest of all of them as they involve waiting times, thus limiting the system headway. The bottleneck area will remain the same. If the speed restrictions are close to the stations, the existing headway may increase, thus further exacerbating the bottleneck. Despite of this argument, if the speed restrictions are significant compared to the booked speed, we may get a case of bottleneck shifting from the station to the restriction area.

Chapter 5

Constraints on train path parameters and implementation via IIT Bombay rail simulator

A train is capable of some acceleration rate, deceleration rate and a certain maximum velocity, but velocity profiles of trains are not just based on these three parameters. If a train is moving on a curve or going on a bridge, there may be speed restrictions placed due to track requirements on the train although it is capable of more. Similarly, when the train is pulling into the station on a side loop (not the main line), it negotiates a curve and pulls into the station slower than when it would pull into the main line which would be straight. These and other situations are considered and their implementation via the simulator is also explained.

The railway simulator has developed at IIT Bombay over the last decade and was one of the main tools used for experimenting in this project. Its main function as described before is to generate a feasible schedule based on section, scheduled and unscheduled trains input as described before. The basic principle used is priority based FIFO scheduling techniques, along with feasibility in terms of at most one train between 2 consecutive signals. Path is scheduled as soon as a block/loop is available.

5.1 Block working time

When a train clears a block, there may be a lag involved in clearing the signal for the other end of the block where another train may or may not be waiting to get a path. This lag is known as the block working time. This parameter is included in the param.dat file of the simulator. 2 unscheduled trains were scheduled on a section with 3 stations and 2 aspect signaling. In one scenario the block working time was kept 0 min and in the other it was kept at 10 min. The results obtained were the following:

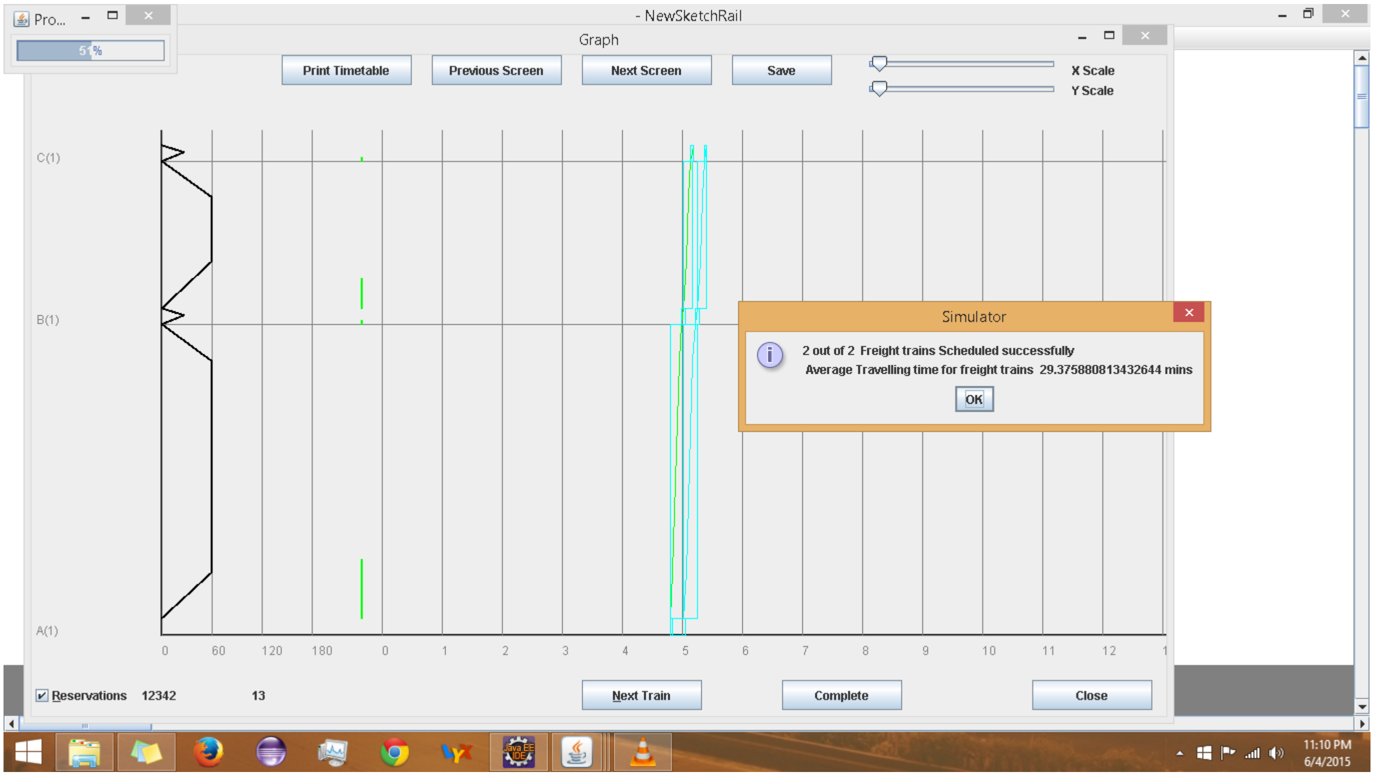

Figure 5.1: Block working time 0 mins

Block working time	T.No.	station Name	Loop	A.Time	D.Time								
0	12341	A	11	4:48:00	4:48:00	B	12	4:59:32	5:01:47	C	13	5:08:49	5:11:05
0	12342	A	11	4:50:02	5:01:34	B	12	5:13:07	5:15:22	C	13	5:22:24	5:24:39
10	12341	A	11	4:48:00	4:48:00	B	12	4:59:32	5:01:47	C	13	5:08:49	5:11:05
10	12342	A	11	5:00:02	5:11:34	B	12	5:23:07	5:25:22	C	13	5:32:24	5:34:39

Figure 5.2: Block working time working

Thus, we see that the second train, which is available as soon as the first has cleared the block, is delayed by 10 min due to the block working time.

5.2 Link velocity

Links are parts of track which join blocks/loops with other blocks/loops. They can have lengths which extend into the resources it joins. For example, a link of 100 m joining a block and a loop can extend to 50 m into the block and 50 m into the loop, or any other combination adding to the link length. Links can have speed restrictions which the whole length of the train has to respect. To validate this working, a section with 3 stations and a non zero length link between loop of second station and the block joining station b and c as seen in the image was considered. The working of the same is seen in these figures:

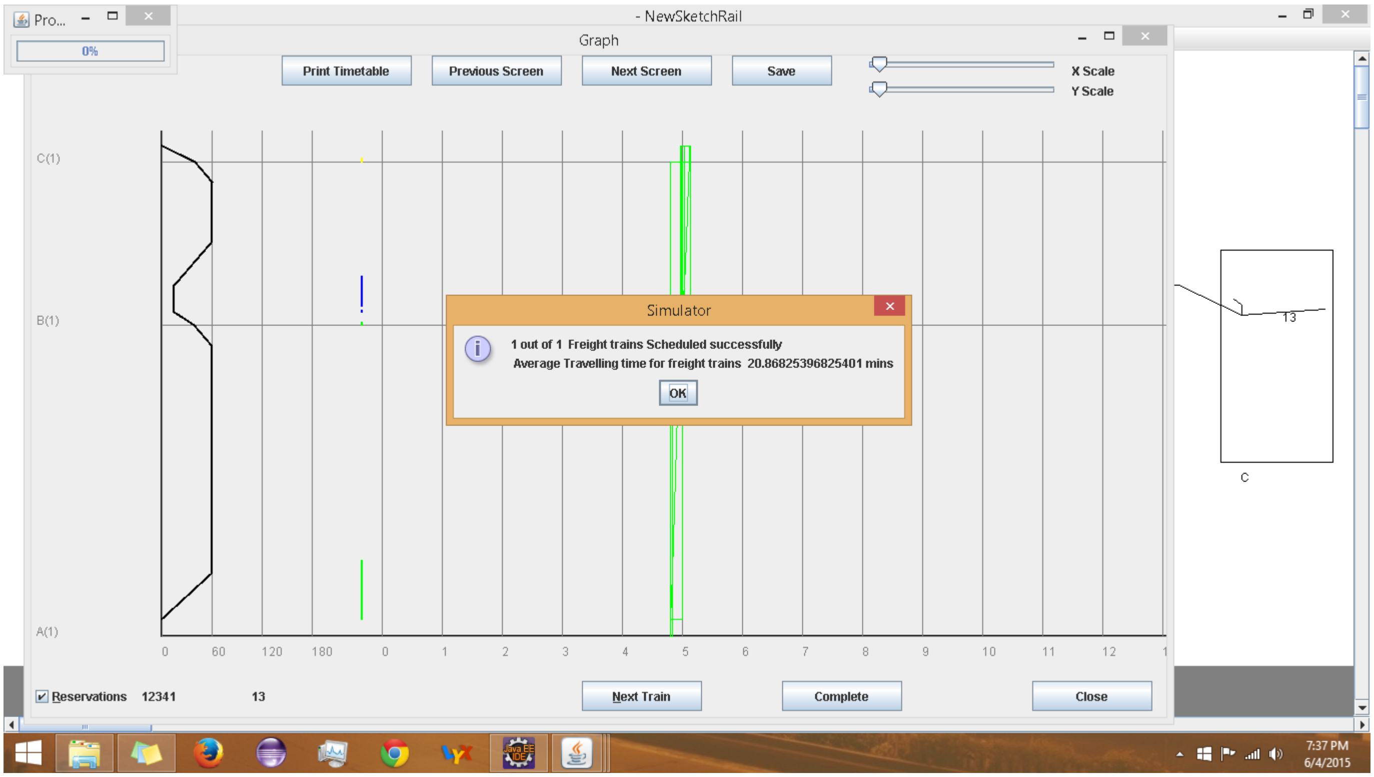

Figure 5.3: Link velocity working with link length of 200 m

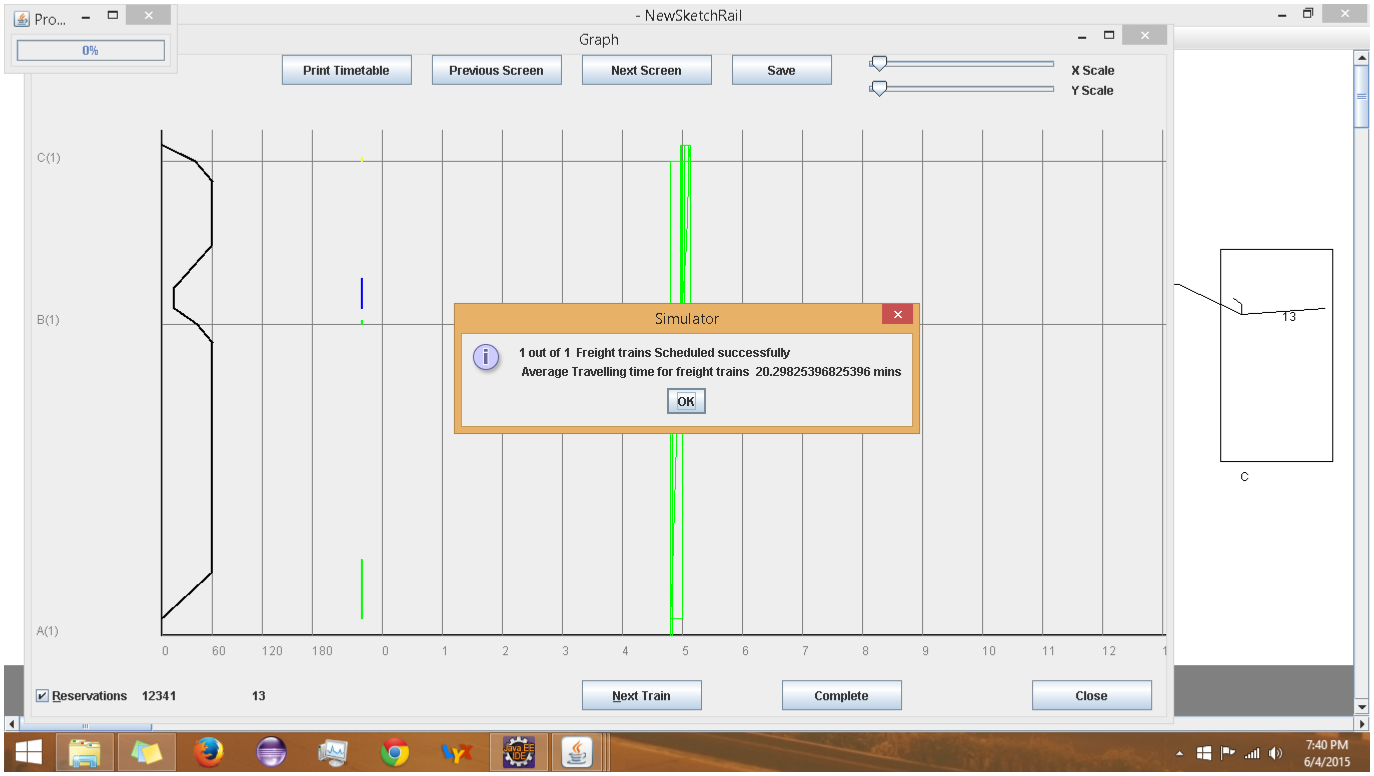

Figure 5.4: Link velocity working with link length of 10 m

In the last case, although the link length is just 10 m, the train's whole length has to manoeuvre it, thus the lesser velocity is held through a longer distance. Note that if the link length is brought down to 0 m, it simply allows the train to pass through as seen in the next figure. Thus to enter a zero link length or point velocity in the section, an arbitrarily small but positive link length will have to be used.

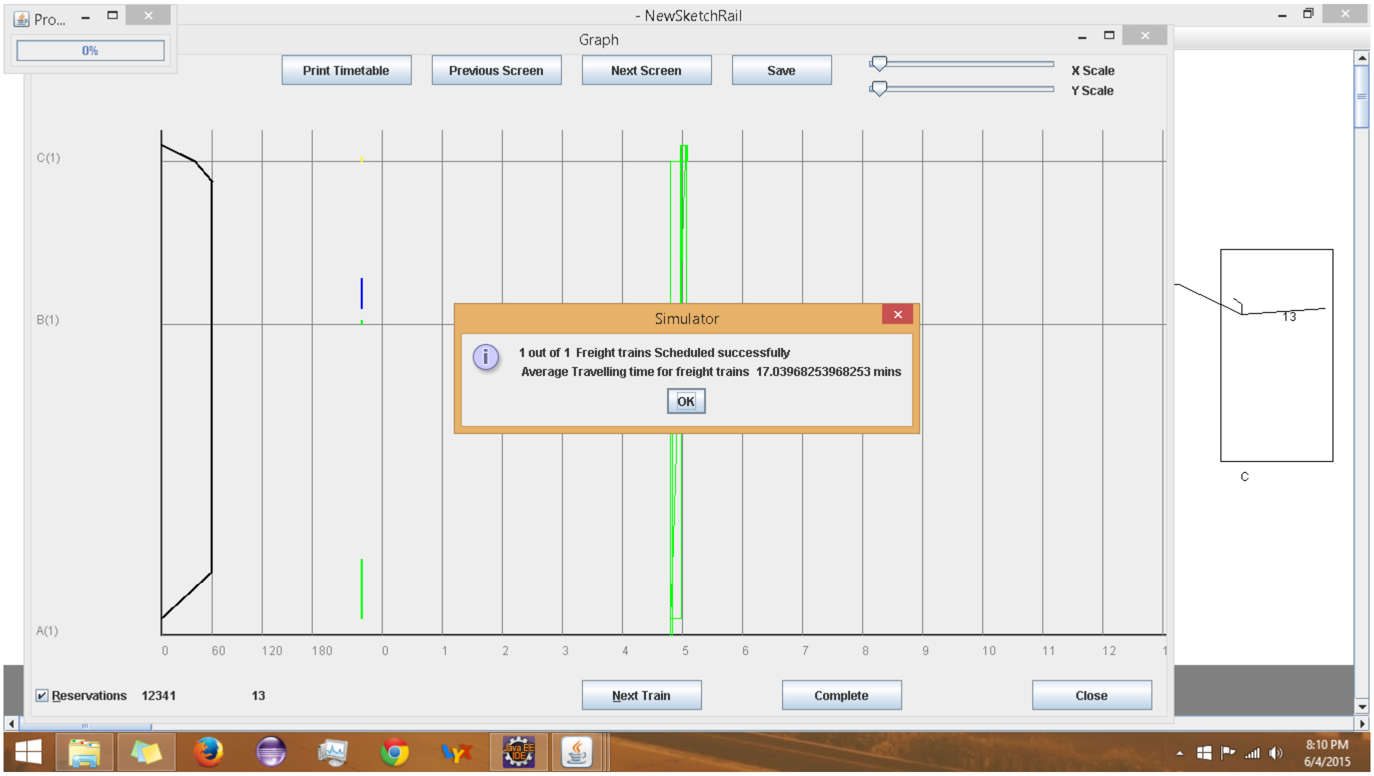

Figure 5.5: Link length working with link length of 0 m

5.3 Gradients

To understand gradient effects on the section, we modeled a 2 station section with a single train on it. 2 cases were tried to understand gradients: No gradients and a uniform up gradient throughout the section (train chugging up throughout the journey situation). The following are the results obtained:

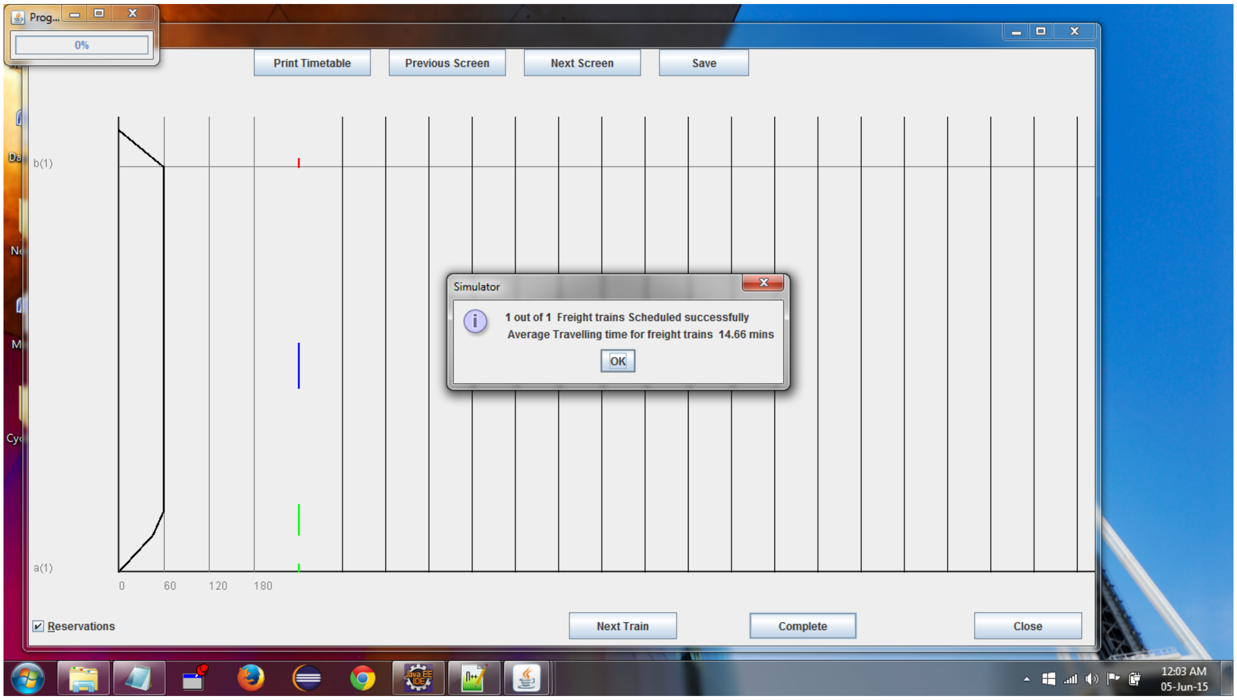

Figure 5.6: no gradient effect

A gradient of slope 1/100 which reduces acceleration by 0.15 m/s^{2}and increases deceleration by 0.15 m/s^{2}were added throughout the length of the section.

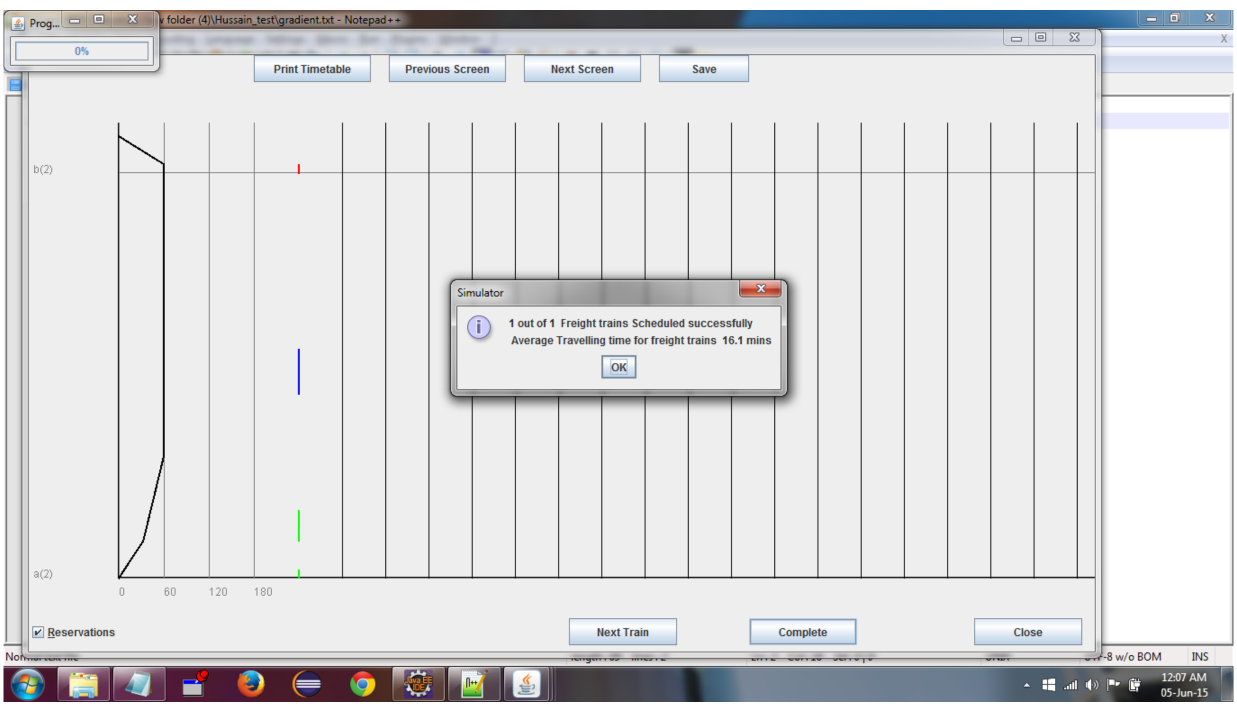

Figure 5.7: Uniform up gradient result

Interpretation: Acceleration takes a larger time than in the previous case as expected. The deceleration at the next station also starts much later as the slope helps the train to decelerate at a faster pace, thus the deceleration can be done in a smaller distance which is seen by comparing the 2 figures. Also, once the train achieves the maximum velocity, the gradients don't affect its working. The simulator's model is that gradients only matter when the train is trying to accelerate or decelerate, not when it is cruising at max speed. Note that ideally stations are supposed to have (and always have to the best of our knowledge) no gradients as trains are supposed to halt there. A free roll forwards or backwards is not desired.

5.4 Loop/block velocity

There are velocity restrictions imposed on the whole loop or block based on the track technology being used. These are effectively modeled. To test it, the loop at station B in the 3 station case was allowed a max speed of 50 km/h. The freight train which was moving at a speed of 60 km/h had to stay down while passing through the loop as seen in the figure. The same idea works for a block too. Note that different loops in the same station can have different speed restrictions in the simulator's logic.

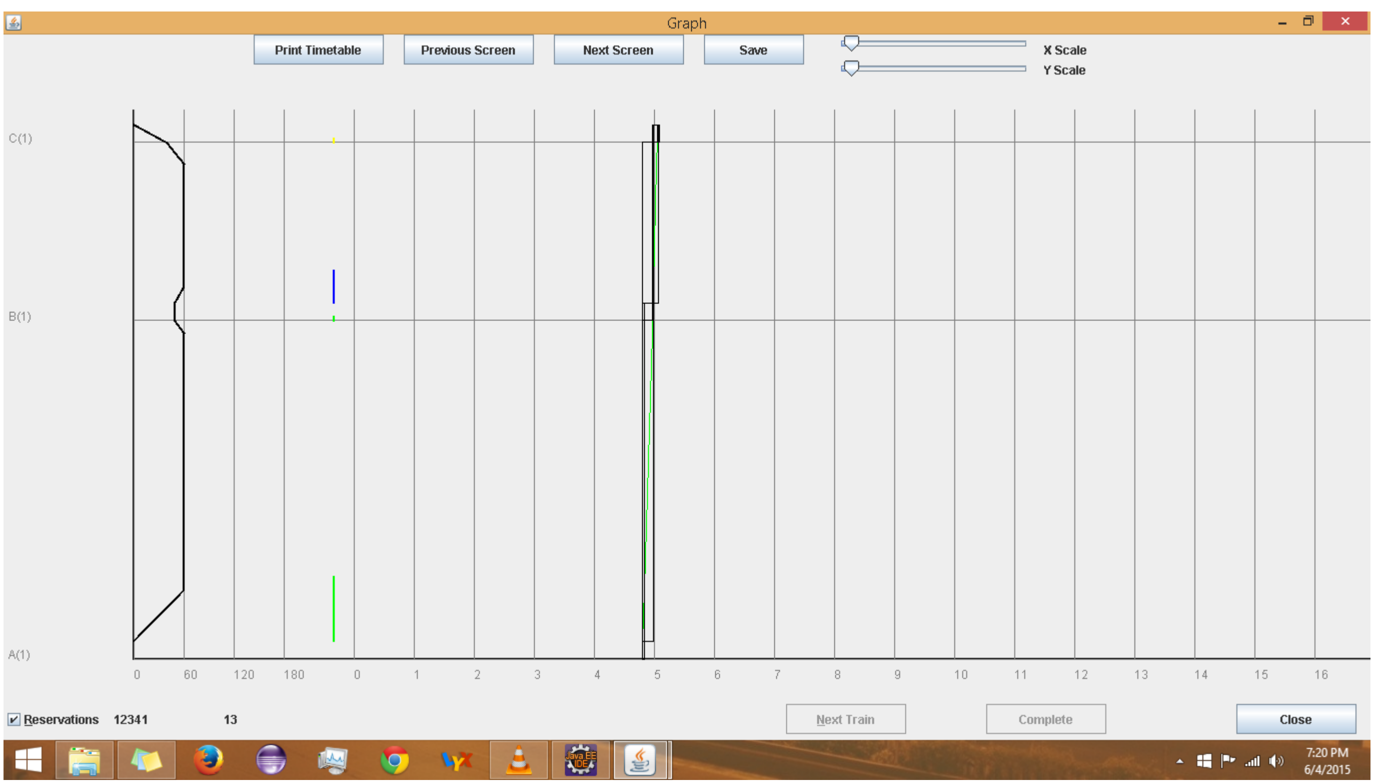

Figure 5.8: Loop velocity restriction

5.5 Speed restrictions within block

Tracks may have permanent or temporary speed restrictions due to curves of different types, bridges, passing through populated areas (example: caution shown in level crossings), and track maintenance or other activities happening near a part of the track. These are restrictions not limited to a whole block: It may span multiple blocks or may be present in just a small part of a block. Ex. When the train is making a sharp curve, the speed needs to be reduced for a short distance. This distance is most likely part of a single block. Thus, there is a need for a speed restriction of this kind also.

In the form based version of GUI (see Appendix), the speed restrictions on blocks have inputs in the form directly, while in NSR, they need to be entered within the block fields at the end. Only one speed restriction per block can be entered in the simulator. This is restrictive and needs to be relaxed.

Demo example: A few of the permanent speed restrictions on the SC-W section (obtained from [11]), were used to show the effect of speed restrictions within blocks with working of a single train:

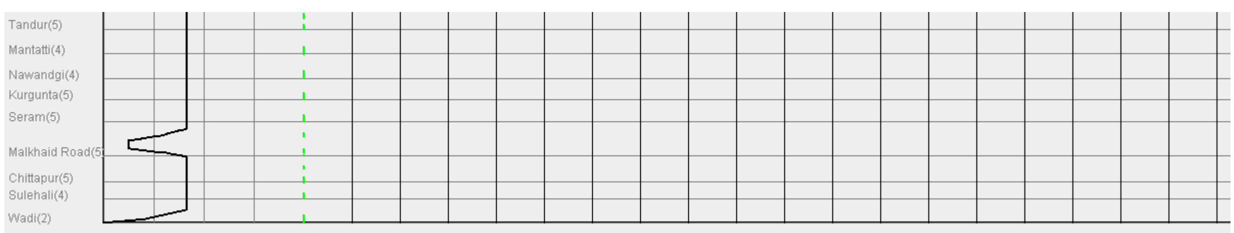

Figure 5.9: Speed restriction in blocks

5.6 Overtaking of freight trains

Freight trains have a lower priority as compared to scheduled trains. If there are 2 trains available and wanting to use a resource, the passenger/scheduled will be given the chance over freight trains. Not only that, if the passenger train is coming from behind and if there is a possibility that the movement of a freight train might hamper the movement of scheduled trains, former will wait in a loop until the latter overtakes it. To see this working, 3 trains were simulated on the ALD-MGS section: A slow moving freight train which started before the 2 scheduled trains, which have a higher priority over freight trains (priorities are to me mentioned for each train in the simulator. Up to 10 levels of priorities are handled within the logic of the simulator.) The slow freight train is overtaken by the faster train while it waits in the loop line at Uchadhi. The green line is the freight train.

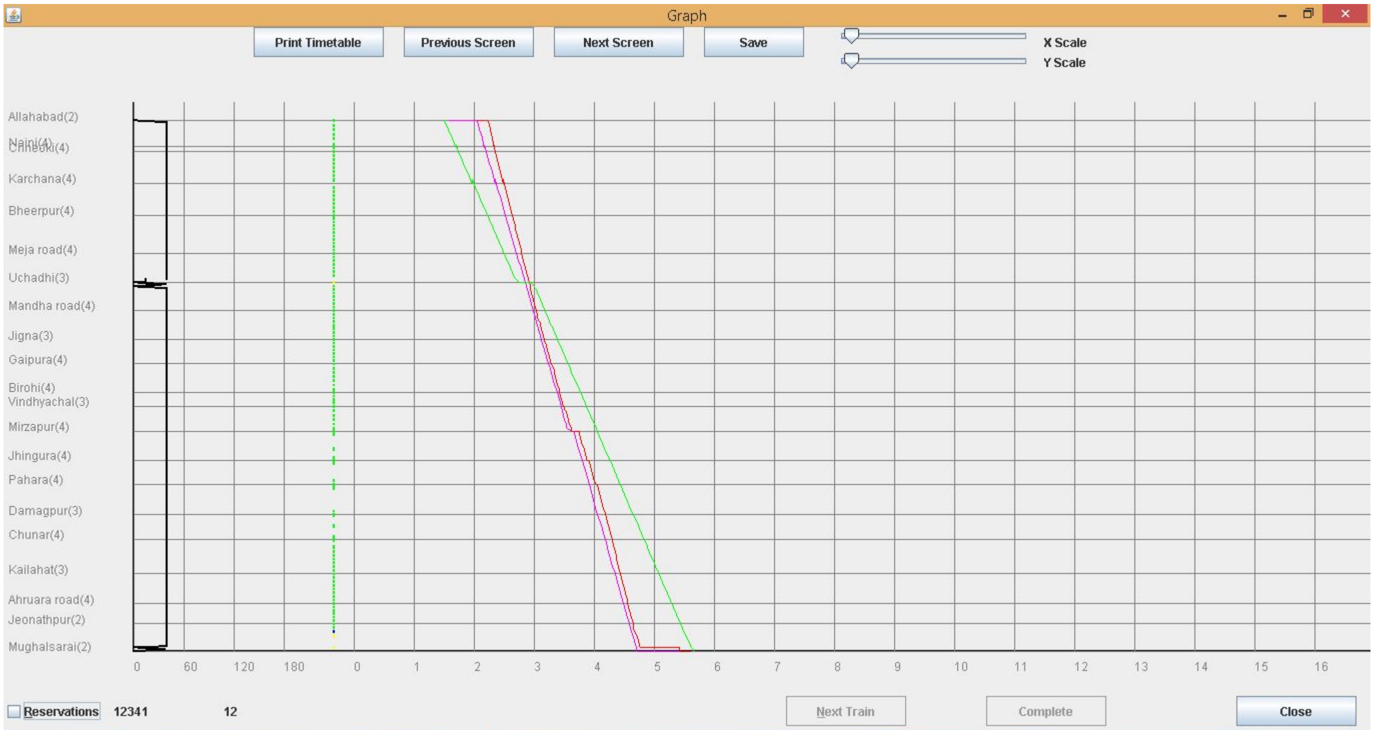

Figure 5.10: Overtaking of freight trains by passenger trains

5.7 Overtaking of scheduled trains

Just as there are priorities between freight trains passenger trains, there are different levels of priorities among passenger trains too. For example: A Rajdhani train will have a higher priority over mail express which will have a higher priority over intercity local trains. These trains don't overtake wherever they get a chance though like freight trains as seen in the last section. Scheduled trains need to have planned overtakes as mentioned in working time tables. Thus, they need to be mentioned within the scheduled.txt file in terms of delays so that the desired trains coming from the back may overtake. Not only that, the loops on which the trains will stop also need to be mentioned for the scheduled trains.

Example: A higher priority, higher speed train following a lower priority, low speed train (both of which are scheduled to go through from start to end without a halt) doesn't overtake it as seen in this figure. The velocity profile is that of the higher priority train which has a lot of start-stop motion due to the slower train in front of it:

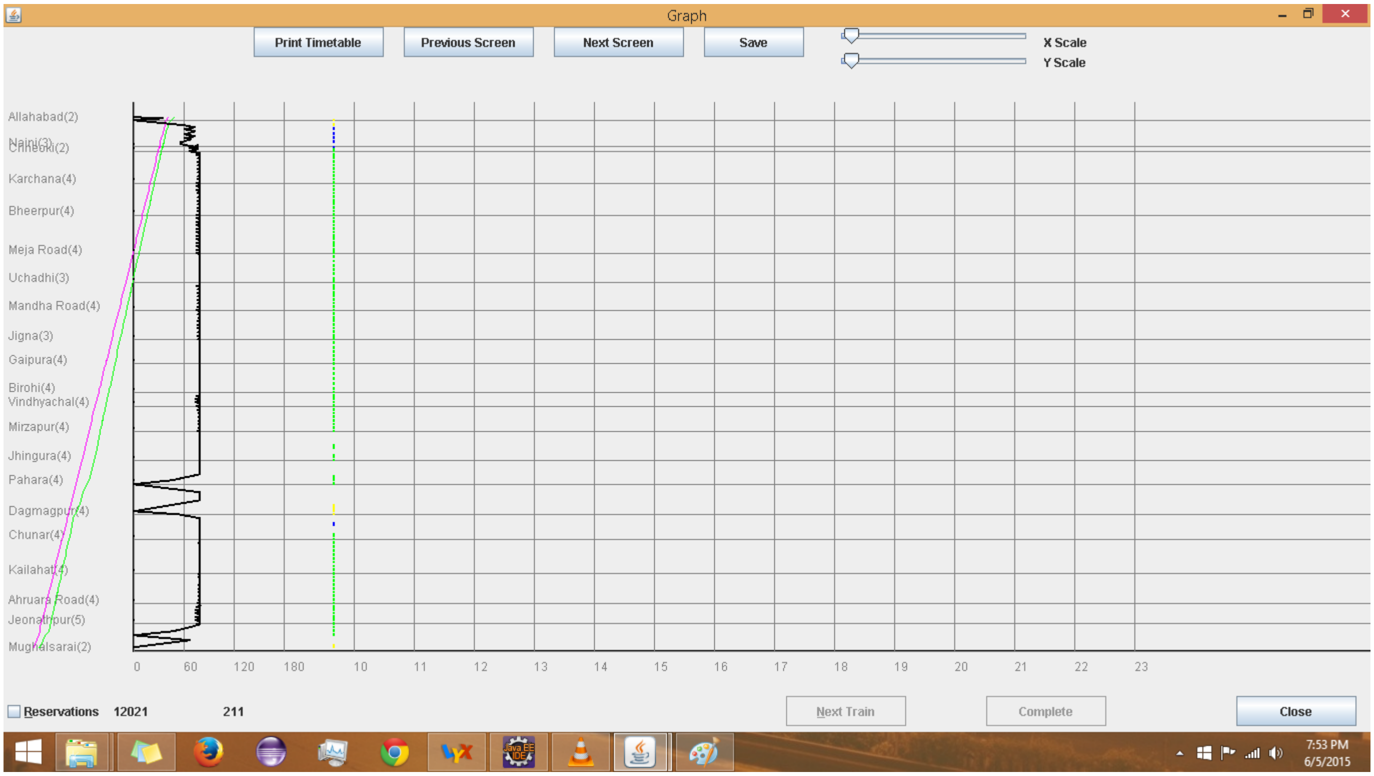

Figure 5.11: Overtaking of scheduled trains

5.8 Crossover movement

Consider the following test section to understand crossover movement. The numbers on the blocks and loops refer to their ID's used within the simulator logic:

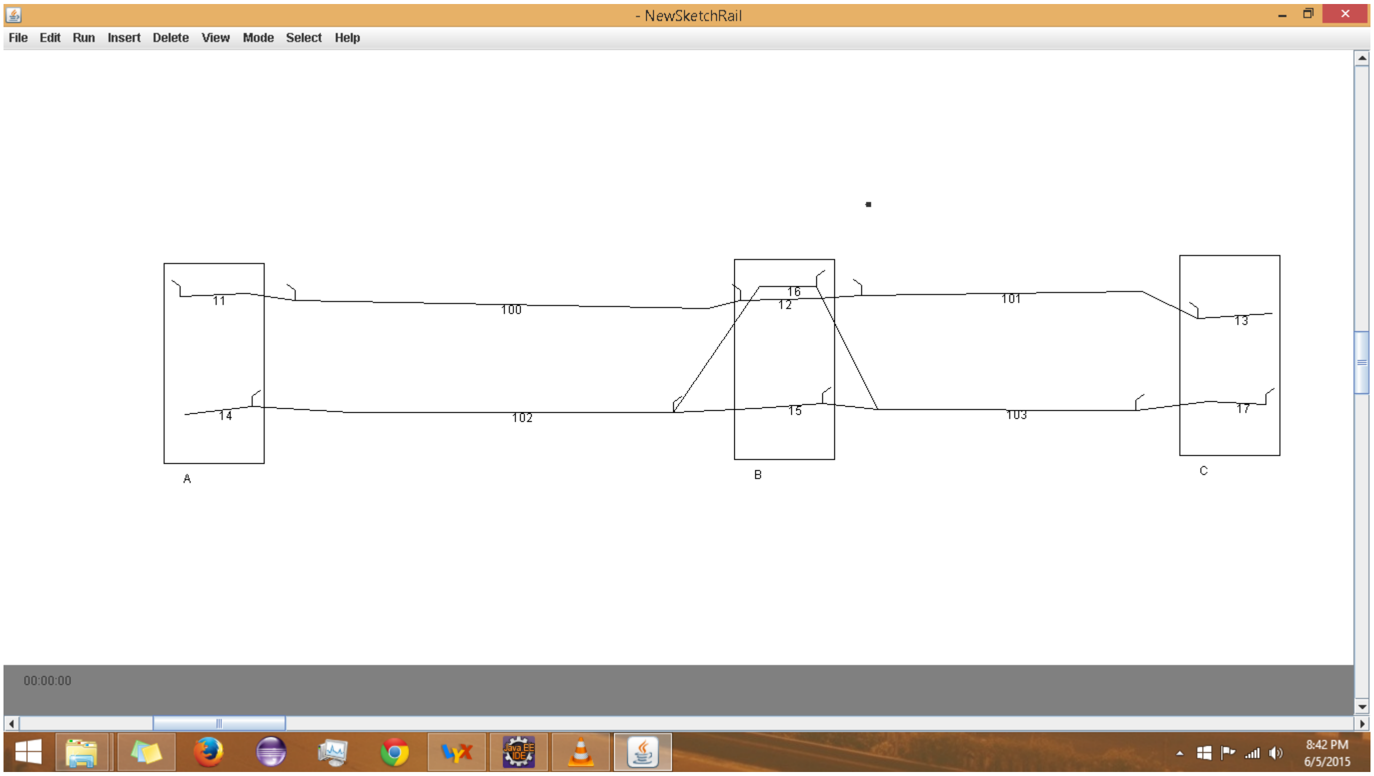

Figure 5.12: Crossover test section

When a train goes from 103 to 16, it 'crosses' the loop number 12 i.e. a train can't be present in 12 while this movement is happening. If the train length is larger than the loop length, it will in fact occupy a part of the block 101 and no train can travel through that block then. Similarly, when train goes from 16 to 102, it uses part of the same track as 100. Thus no train can occupy that block when this train is going down. This is precisely the crossover movement: Trains using a part of another resource for a while to get to a different resource. NSR figures out the crossovers itself using the diagram just showed, while there is crossover detection in form based GUI too based on the loop configuration mentioned. Only in the old version of the simulator is the manual entry of the crossover required. Working Example:

Consider the following part of the ALD-MGS section:

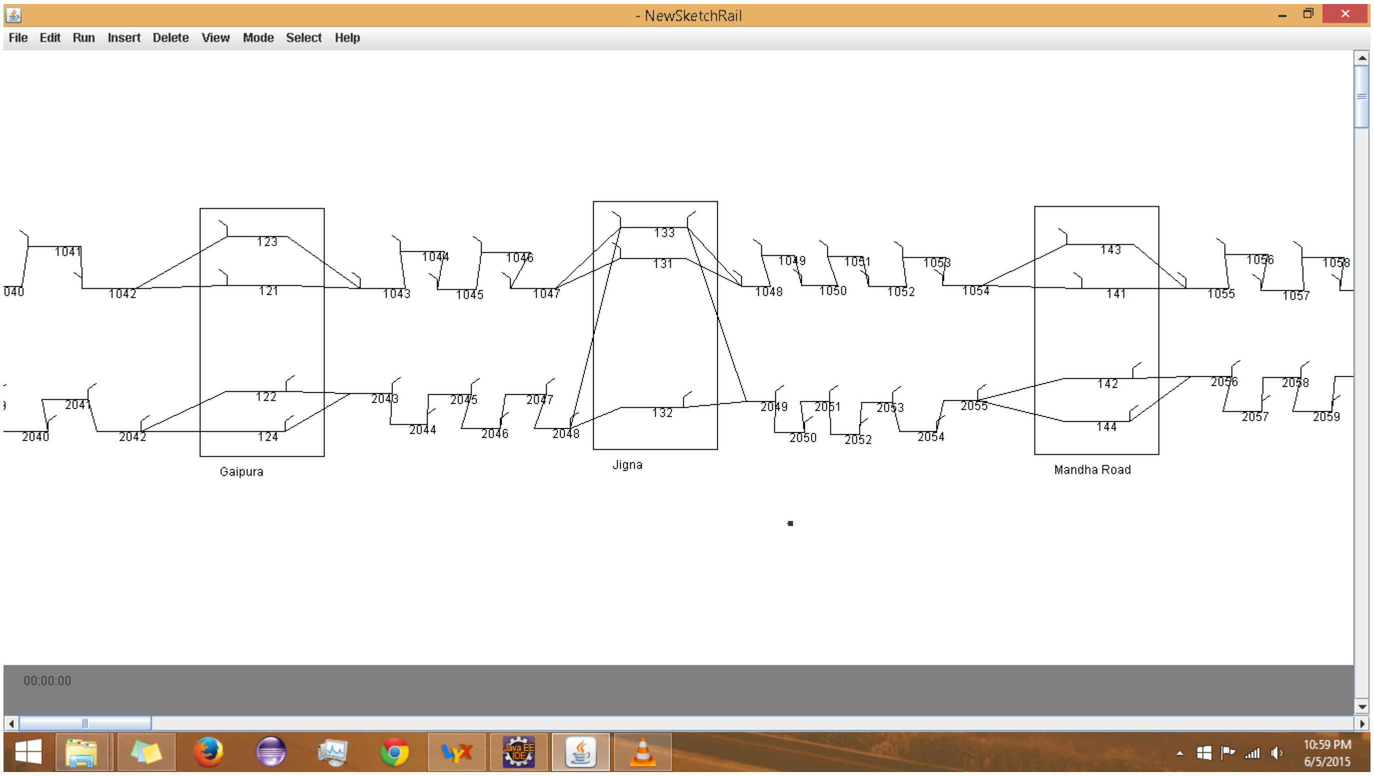

Figure 5.13: crossover at Jigna

When the train is going from 2049 to 133, it uses 131 while crossing, and similarly when it goes from 133 to 2048, it uses 1047. To test this, 2 scheduled trains in opposite directions were considered: 1 from 121 to 141 while passing through 131, another going from 142 to 122 passing through 133. The times were set in such a way that they would want to use the common resource 1047 at the same time, thus creating a conflict which would result in the one who managed to start using it first the winner of the outcome, which turned out to be the train going from Mandha Road to Gaipura in this case, while the train going in the opposite direction waiting in 1046 while the track was being occupied by the former train. The results are as follows:

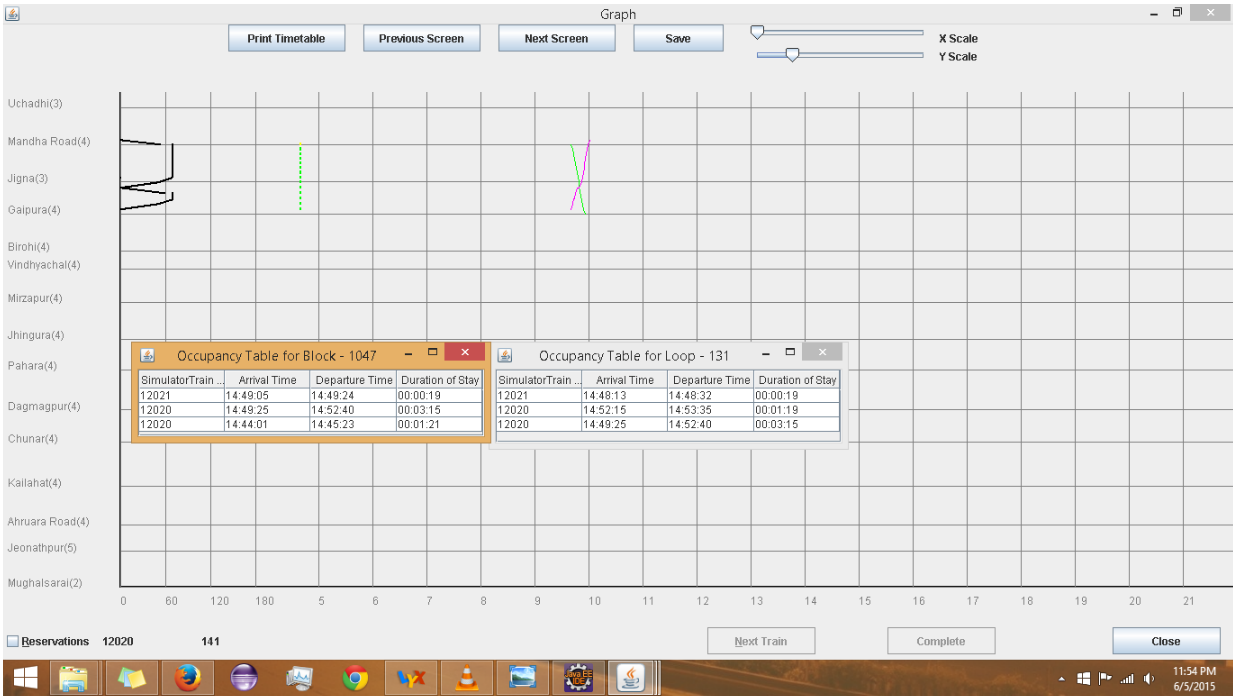

Figure 5.14: Result of crossover

This clearly shows that the train going towards Mandha Road waited for a while before moving on as is seen both in the time-space and velocity-space graph. If there was no crossover movement involved with either 131 or 1047, this would have been the result:

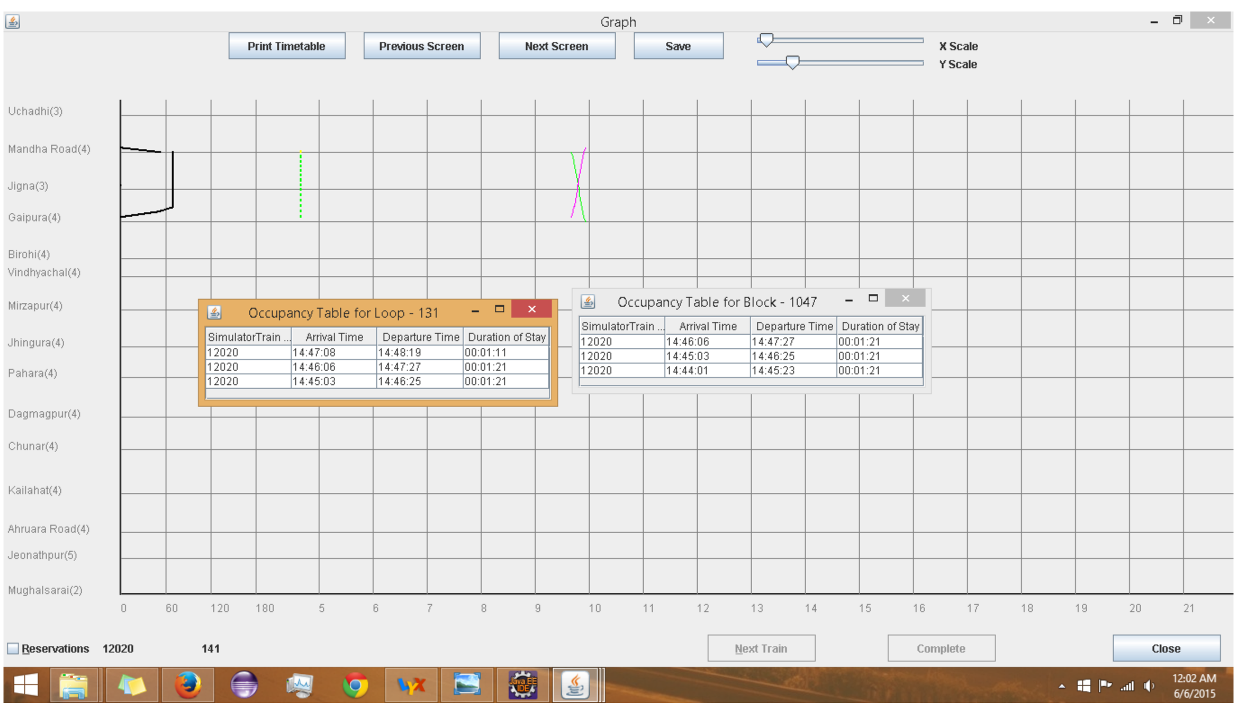

Figure 5.15: No crossover result

As seen, only train number 12020 would have used these resources, and would have gone through without any stops. The time-tables of both these scenarios can also be seen over here:

	T.No.	station Na	Loop	A.Time	D.Time								
no crossover	12021	Mandha R	142	14:40:00	14:40:00	Jigna	133	14:48:13	14:49:05	Gaipura	122	14:54:16	14:56:30
	12020	Gaipura	121	14:40:00	14:40:00	Jigna	131	14:47:08	14:48:00	Mandha R	141	14:54:16	14:56:30
with crossover	12021	Mandha R	142	14:40:00	14:40:00	Jigna	133	14:48:13	14:49:05	Gaipura	122	14:54:16	14:56:30
	12020	Gaipura	121	14:40:00	14:40:00	Jigna	131	14:52:15	14:53:16	Mandha R	141	14:59:32	15:01:46

Figure 5.16: Time-table comparison for crossover movements

This shows how crossovers are implemented in the simulator.

5.9 Parameters not implemented

There are a few parameters which are not yet implemented in the simulator although some of them are regularly encountered in IR.

1. Multiple signaling regimes: Not all signals in a section need be of the same aspect numbers. For example, on the suburban section, there are 3 as well as 4 aspect signals observed. The simulator input considers the number of aspects of each signal as a global input, i.e. it is the same for all. The simulator logic is capable of handling multiple aspects and the same should be changed.

2. Level crossings: Roads are blocked when trains are passing through them. These crossings are called level crossings. Trains have the right of way in most cases of level crossings and that's what the simulator assumes, thus neglecting their implementation, but there are cases in which these is not true and the train comes to a halt at level crossings. Thus their modeling needs to be done in the simulator. A possible abstract way to do this within the current framework is running a point sized, high priority train between two pseudo stations on either side of the level crossings. The run time of this point train should be similar to the level crossing time that we need to show.

3. Maintenance blocks: A few hours every day the track near a station is kept under maintenance. As the maintenance car might be using the blocks near the platform, trains can't operate during this time. These times aren't timetabled and are decided by station operation personnel in a way to minimize the effect on passenger traffic. These are not modeled either and can be modified in a way similar to that described in point 2 above.

4. Station entry velocity: Provision for this parameter has been provided in the station.txt file, but not yet implemented. The purpose of the same is unclear as loop velocity and link velocity for blocks and loops are implemented and these should suffice to describe station entry velocity too.

5. In the param.dat file, the fields corresponding to 'delayFactor', 'redFailWaitTime', 'redFailVelocity' and 'warnerDistance' are not functional as of now. If automatic blocks fail, the signal shows red all the time.In such a situation the following protocol is followed by trains: It'll wait for a fixed time (redFailWaitTime), and if the block seems clear ahead and no train can be observed on the track, it moves ahead with a certain restricted velocity (redFailVelocity). warnerDistance: The status of the signal is warned to the driver some distance before the signal distance by another signal. This is the warner signal and the distance is the warner distance.

Chapter 6

Allahabad Mughalsarai section: Bottleneck alleviation

This chapter discusses the overview of the section, details on the visit to Allahabad and on the section observed during the visit, how the bottleneck was detected, and then strategies to alleviate bottlenecks and improve operations.

6.1 Section details and model

The section of tracks from Allahabad to Mughalsarai is part of the Allahabad division of North Central Railway zone. The part of the track on the right side from Allahabad onwards is the section considered for this exercise:

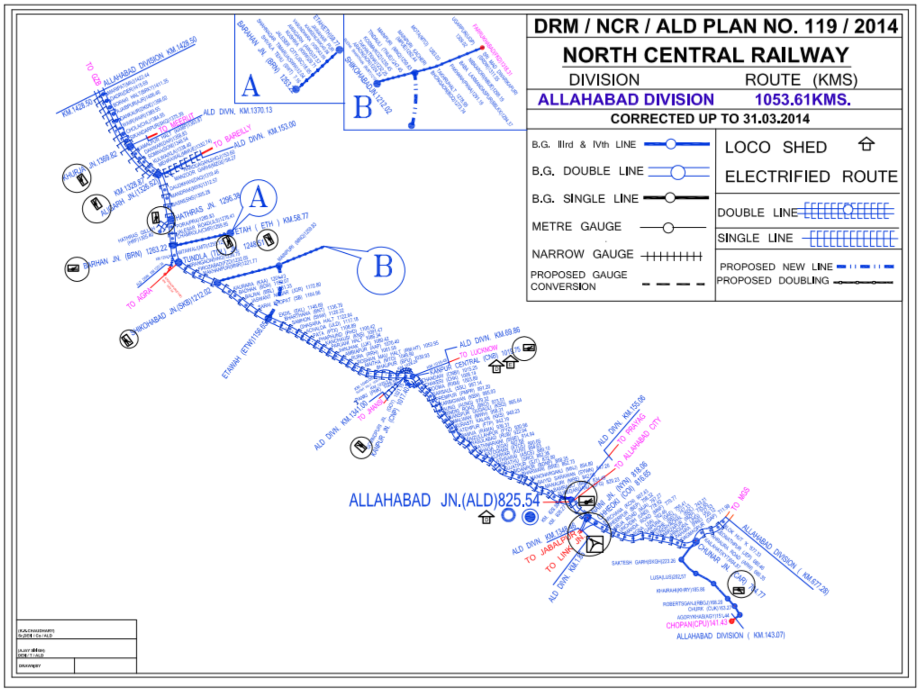

Figure 6.1: Allahabad division

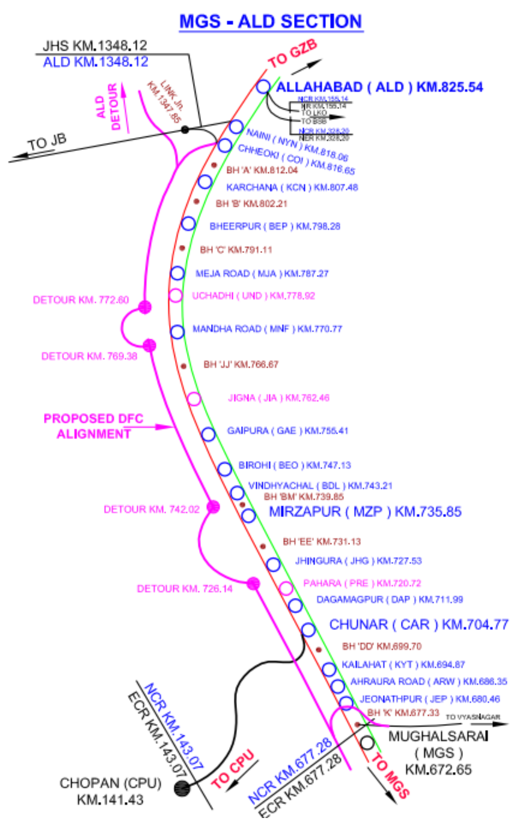

Figure 6.2: ALD-MGS section

This division is on an important location as it connects the west India (New Delhi, Punjab etc.) to the east of India (Kolkata etc.) and has heavy passenger and freight traffic. Most of the double line track in this section has automatic block signaling (4 aspect), apart from a part in the middle which still has absolute blocks, as seen in this table:

Automatic Section	Station At Kilometre	Name of Station	Distance between station	No of Semi Auto + Auto Sig. UP/DN
27.44 Km	677.33	K	3.13	2/2
	680.46	JEP	5.89	4/4
	686.35	ARW	8.52	6/6
	694.87	KYT	9.90	7/7
31.08 Km.	704.77	CAR	4.08	Absolute Block System
	708.70	DC	3.14	
	711.99	DAP	8.13	
	720.12	PRE	7.41	
	727.53	JHG	3.60	
	731.13	EE	4.72	
89.69 Km.	735.00	MZP	8.21	5/4
	743.21	BDL	3.92	2/2
	747.13	BEO	8.28	4/5
	755.41	GAE	7.05	4/5
	762.46	JIA	8.31	6/6
	770.77	MNF	8.15	6/6
	778.92	UND	8.35	5/6
	787.27	MJA	11.05	9/9
	798.32	BEP	9.11	7/6
	807.43	KCN	9.22	7/6
	816.65	COI	1.45	-
	818.10	NYN	7.44	4/4
	825.54	ALD		

Figure 6.3: ALD-MGS block info

Explanation: K, DC and EE are block huts which serve as intermediate signals between the stations they are located. Thus, there is an intermediate block between CAR (Chunar) - DAP (Dagmagpur) and JHG Jhingura and MZP (Mirzapur). There may not be the same number of automatic signals between 2 consecutive stations. For example, there are 4 up direction signals while 5 down direction signals BEO (Birohe) and GAE (Gaipura). Note that the direction towards ALD is the Up direction in this division.

Stations may have loops apart from the up and down main lines which are there at every station. The next image gives the number, types and lengths of loops at intermediate stations on the ALD-MGS section. UP loops are on the side of Up main line only and don't affect the movement in down direction. Similarly for Down direction. Common loops may be on Up side (Common loop - Up ML - other loops/Down ML configuration), down side, or sandwiched between the 2 main lines:

Sr. No.	Name of station	UP Loop (length in meter)	DN Loop (length in meter)	Common Loop (length in meter)
1.	JEP	680.20	(i) 680.14 (ii) 729.63	-
2.	ARW	703	685	-
3.	KYT	-	776	727 (UP)
4.	DAP	715	-	715(DN)
5.	PRE	712	768	-
6.	JHG	688.84	692	-
7.	BDL	695	705	-
8.	BEO	695	693.4	-
9.	GAE	747	692	-
10.	JIA	-	-	695(UP)
11.	MNF	733.75	837	-
12.	UND	-	-	685(Sandwich)
13.	MJA	742	745	-
14.	BEP	688.15	753	-
15.	KCN	699	735	-

Figure 6.4: ALD-MGS Loop info

The section was modeled in the IIT Bombay Rail simulator for experimentation and it looked like this:

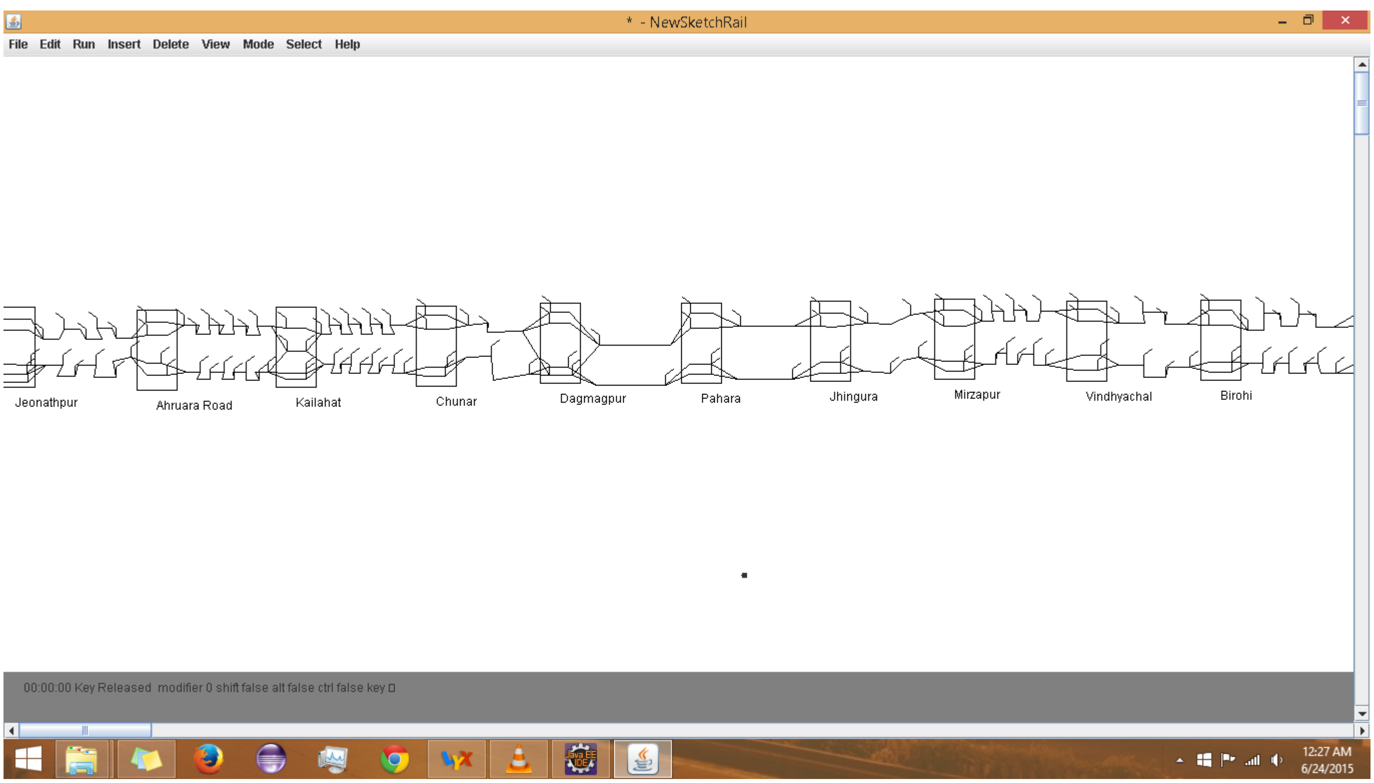

Figure 6.5: ALD-MGS Model

6.2 Visit to Allahabad and observations on the section

As mentioned in the literature review section, a trip was conducted to ALD at the end of April 2015. While traveling in passenger train number 12321, the following observations were made on the section and the train running: (Note: The same hold true for a lot of sections on Indian Railways.)

1. Mast numbers: There are long masts alongside the track pocated periodically. They're of the type 'x/k', where x is the current km from some reference station and k is the mast number for that km. the left side of the track had odd k's while the other side had even k's.

2. Top speed of the train was 110kmph.

3. Driver behavior/protocol on yellow: The speed was decreased to around 30kmph on every block which was entered on a yellow.

4. Double distant and distant signal were present on the section before the home signal. Starter and warner signals were present before the actual block signal.

5. At Dagmagpur: A freight was pushed over a priority train (12321). 12321 had to wait at the loop till the absolute block between Dagmagpur and Pahara.

6. The train was already late at Chunar by around 5 hours and it was traversing mostly through double yellows and yellows given by freight trains (no other passenger trains at that time). This was the first indication of congestion observed.

7. Speed restriction of 'x'kmph is marked by sign of x in yellow triangle board. End of the restriction is indicated by a through or 'T' signal (again in yellow)

8. There was a long halt at Mirzapur. Yellow signal was showing for home signal but train didn't leave. The train did leave on yellow only eventually. This is a possible indication of driver behavior again or a guard getting down.

9. Home signals (signals which guard the loop) have a stem with lights on it. The direction of the stem indicates the direction of the loop at the station. If the lights are on, it indicates that the stop is at the loop, otherwise it is through the mainline which is straight up.

10. Driver behavior on speed restrictions of 30: Slows down to 30 much before the 30 mark starts. Suggests extremely cautious behavior of the driver. The same was true on all kinds of restrictions.

11. Train took 1.5 hours to travel the last 3-4 km before Allahabad. This was a clear indication of the strong node effects at ALD and that the bottleneck is also probably at ALD.

6.3 Detecting bottleneck

The ALD-MGS section handles a lot of traffic (roughly 104 trains with 52 passenger train and rest being freight and other train paths like loco movements) in one direction (similar numbers hold for the opposite direction, MGS to ALD). This average has roughly doubled over the traffic of the mid-90s (Statistics obtained from calculations done by the ALD division). Due to the heavy traffic, there have been a lot of delays in the movements of trains, thus finding the bottleneck and alleviating it is important. Using a criteria similar to the waiting time method described in 2.5.1.2, we consider average delay at subsequent locations to figure out the bottleneck. As trains reach close to the bottleneck, the delays compared to the normal timetable should increase. This idea was used along with the delay information provided in [9]. Trains going from CAR to ALD and in the return from ALD to CAR were considered to obtain average delay statistics. The following results were obtained:

Try these Via Stations: Varanasi

No.	Name	Type	Zone	From	PF	Dep	Avg	To	PF	Arr	Avg	Duration	Halts	Dep Days	Classes	Qt	Distance	Speed
84369	Triveni Link express	Exp	NR	CAR	4	00:10	+0	ALD	9	02:35	+0	2h 25m	4	S M T W T F S	II SL	GNWL	120 km	49 km/hr
14369	Triveni Express	Exp	NR	CAR	4	00:10	+1:49	ALD	9	02:35	+1:49	2h 25m	4	S W F	II SL 2A	GNWL	120 km	49 km/hr
24369	Triveni Express	Exp	NR	CAR	4	00:10	+2:05	ALD	9	02:30	+1:48	2h 20m	4	M T T S	II SL 2A	GNWL	120 km	51 km/hr
12873	Jharkhand Swarna Jay...	SF	SER	CAR	2	02:45	+2:07	ALD	3	04:35	+2:34	1h 50m	0	T W F	II SL 3A 2A	RLWL	121 km	65 km/hr
13007	Udyan Abha Toofan Ex...	Exp	ER	CAR	2	03:16	-	ALD	2	05:05	-	1h 49m	2	S M T W T F S	II SL 3A	RLWL	121 km	66 km/hr
13201	Rajendra Nagar Patna...	Exp	ECR	CAR	2	05:12	+46	COI	??	08:20	+37	3h 8m	4	S M T W T F S	II SL 3A 2A	PQWL	112 km	35 km/hr
12175	Chambal Express	SF	NCR	CAR	2	05:42	+55	COI	??	07:45	+1:42	2h 3m	2	M W T	II SL 3A 2A	RLWL	112 km	54 km/hr
12177	Howrah-Mathura Chamb...	SF	NCR	CAR	2	05:42	+1:21	COI	??	07:45	+2:54	2h 3m	2	S	II SL 3A 2A	RLWL	112 km	54 km/hr
22912	Shipra Express (PT)	SF	WR	CAR	2	05:42	+1:05	COI	??	07:50	+1:53	2h 8m	2	S T F	II SL 3A 2A	RLWL	112 km	52 km/hr
54103	Chunar - Allahabad P...	Pass	NCR	CAR*	1	06:10	+0	ALD*	2	09:15	+45	3h 5m	14	S M T W T F S	Unreserved		120 km	39 km/hr
18101	Tatanagar-Jammu Tawi...	Exp	SER	CAR	??	06:40	+53	ALD	3	08:40	+2:22	2h 0m	3	S M T W T F S	II SL 3A	RLWL	121 km	60 km/hr
18109	Rourkela - Jammu Taw...	Exp	SER	CAR	4	06:40	-	ALD	6	08:40	-	2h 0m	3	S M T W T F S	II SL 3A 2A	GNWL	121 km	60 km/hr
53345	Chopan Allahabad Pas...	Pass	ECR	CAR	4	08:05	+40	ALD*	2	12:00	+1:05	3h 55m	14	S M T W T F S	Unreserved		120 km	30 km/hr
12321	Howrah - Mumbai CST ...	SF	ER	CAR	2	09:02	+2:12	ALD	6	11:07	+3:00	2h 5m	2	S M T W T F S	II SL 3A 2A 1A	RLWL	121 km	57 km/hr
11094	Mahanagari Express	Exp	CR	CAR	2	12:30	+49	COI	??	15:10	+45	2h 40m	3	S M T W T F S	II SL 3A 2A	GNWL	112 km	41 km/hr
13131	Kolkata - Anand Viha...	Exp	ER	CAR	??	15:05	+2:28	ALD	??	17:55	+3:48	2h 50m	9	S M T W T F S	II SL 3A	RLWL	120 km	42 km/hr
54105	Mughalsarai Allahaba...	Pass	NCR	CAR	2	17:00	+1:07	ALD*	2	20:55	+2:07	3h 55m	14	S M T W T F S	Unreserved		120 km	30 km/hr
19046	Chhapra-Surat Tapti ...	Exp	WR	CAR	2	18:50	+1:05	COI	3	20:48	+1:31	1h 58m	2	S T W F S	II SL 3A	RLWL	112 km	56 km/hr
19048	Bhagalpur-Surat Tapt...	Exp	WR	CAR	2	18:50	+1:25	COI	3	20:45	+1:57	1h 55m	2	M T	II SL 3A	RLWL	112 km	58 km/hr
18202	Nautanwa - Durg Exp...	Exp	SECR	CAR	2	20:02	+2:00	ALD	8	22:30	+2:58	2h 28m	2	S F	II SL 3A 2A	RLWL	121 km	48 km/hr

Figure 6.6: Train going from CAR to ALD

Try these Via Stations: Varanasi Muri Mughal Sarai

No.	Name	Type	Zone	From	PF	Dep	Avg	To	PF	Arr	Avg	Duration	Halts	Dep Days	Classes	Qt	Distance	Speed
11093	Mahanagari Express (...	Exp	CR	COI	??	00:32	+47	CAR	3	02:58	+51	2h 26m	2	S M T W T F S	II SL 3A 2A	RLWL	112 km	45 km/hr
12874	Jharkhand Swarna Jay...	SF	SER	ALD	6	04:50	+5:29	CAR	5	06:35	+7:09	1h 45m	0	W T S	II SL 3A 2A	RLWL	121 km	69 km/hr
19047	Surat-Bhagalpur Tapt...	Exp	WR	COI	2	06:40	+30	CAR	3	08:23	+58	1h 43m	2	S W	II SL 3A	RLWL	112 km	65 km/hr
19045	Surat-Chhapra Tapti ...	Exp	WR	COI	2	06:40	+42	CAR	3	08:23	+1:11	1h 43m	2	M T T F S	II SL 3A	RLWL	112 km	65 km/hr
54106	Allahabad Mughal Sar...	Pass	NCR	ALD*	7	07:00	+11	CAR	3	10:17	+1:30	3h 17m	14	S M T W T F S	Unreserved		120 km	36 km/hr
13132	Anand Vihar T. - Kol...	Exp	ER	ALD	4	08:00	+7:38	CAR	??	10:55	+7:31	2h 55m	9	S M T W T F S	II SL 3A	RLWL	120 km	41 km/hr
18201	Durg - Nautanwa Expr...	Exp	SECR	ALD	7	10:45	+1:34	CAR	1	12:28	+2:54	1h 43m	2	T S	II SL 3A 2A	RLWL	121 km	70 km/hr
53346	Allahabad Chopan Pas...	Pass	ECR	ALD*	3	14:50	+11	CAR	5	18:55	+57	4h 5m	14	S M T W T F S	Unreserved		120 km	29 km/hr
18110	Jammu Tawi - Rourkel...	Exp	SER	ALD	6	16:02	-	CAR	5	18:10	-	2h 8m	3	S M T W T F S	II SL 3A 2A	RLWL	121 km	56 km/hr
18110-Slip	Jammu Tawi - Tatanag...	Exp	SER	ALD	6	16:02	-	CAR	??	18:10	-	2h 8m	3	S M T W T F S	II SL 3A	RLWL	121 km	56 km/hr
12178	Mathura-Howrah Chamb...	SF	NCR	COI	??	17:15	+1:09	CAR	3	18:53	+2:40	1h 38m	2	M	II SL 3A 2A	RLWL	112 km	68 km/hr
22911	Shipra Express	SF	WR	COI	??	17:15	+40	CAR	3	18:53	+1:22	1h 38m	2	S W F	II SL 3A 2A	RLWL	112 km	68 km/hr
12176	Chambal Express	SF	NCR	COI	??	17:15	+1:41	CAR	3	18:53	+2:53	1h 38m	2	T T S	II SL 3A 2A	RLWL	112 km	68 km/hr
54104	Allahabad - Chunar P...	Pass	NCR	ALD*	2	18:30	+5	CAR*	1	21:40	+1:48	3h 10m	14	S M T W T F S	Unreserved		120 km	38 km/hr
12322	Mumbai CST - Howrah ...	SF	ER	ALD	6	20:55	+2:36	CAR	3	22:46	+3:59	1h 51m	2	S M T W T F S	II SL 3A 2A 1A	RLWL	121 km	65 km/hr
14370	Triveni Express	Exp	NR	ALD	10	21:20	+1:31	CAR	5	00:45	+2:06	3h 25m	5	M W F	II SL 2A	RLWL	120 km	35 km/hr
24370-Slip	Triveni Link Express	Exp	NR	ALD	10	21:20	-	CAR	5	00:45	-	3h 25m	5	S T T S	II SL	RLWL	120 km	35 km/hr
24370	Triveni Express	Exp	NR	ALD	10	21:25	+1:36	CAR	5	00:45	+2:05	3h 20m	5	S T T S	II SL 2A	RLWL	120 km	36 km/hr
13008	Udyan Abha Toofan Ex...	Exp	ER	ALD	6	22:15	-	CAR	3	23:54	-	1h 39m	2	S M T W T F S	II SL 3A	RLWL	121 km	73 km/hr
13202	Mumbai LTT - Rajendr...	Exp	ECR	COI	??	23:30	+1:39	CAR	3	02:35	+1:13	3h 5m	3	S M T W T F S	II SL 3A 2A	RLWL	112 km	36 km/hr

Figure 6.7: Trains going from ALD to CAR

Direction of train	Delay at CAR (min)	Delay at ALD (min)
CAR to ALD	76	112
ALD to CAR	154	105

Table 6.1: SC - W intermediate and auto/cab signaling

The hypothesis was that the bottleneck is the ALD yard, so as expected, the waiting times while moving towards ALD from CAR do increase. This was in line with our observations while traveling on 12321 and also from what was heard from the people working at the ALD division themselves. If the hypothesis was true, then in the reverse direction (i.e. ALD to CAR), the delays while moving from ALD to CAR should have remained the same or decreased, not worsened further. This is not what is observed here from the statistics obtained in the previous table.

The statistics obtained from indiarailinfo correspond to the average delays wrt to the time tabled time. One reason why our expectations are not met by the statistics are that the time provided in the time-table for the trains to go from one station to the next might not correspond to the normal traversal (or processing times). Due to slacks (or the lack of them which is unlikely), the delay times don't reflect the real state of congestion.

Also, delays at a station may not be a perfect proxy for waiting times for the next resource (i.e. block) which is actually the metric considered in 2.5.1.2. This is another reason why the increase in delays while moving from ALD to CAR are not reflective of the bottleneck being the ALD yard (assuming the hypothesis is true, which most likely is).

Under these circumstances, the way to resolve this is obtain the waiting time statistics at all loops and blocks for all the trains and conside the bottleneck as the resource which has the maximum average waiting time involved with it.

6.4 Better operating strategies and investments

1. Make the complete section automatic block signaling: Running a single train gave the following running characteristic:

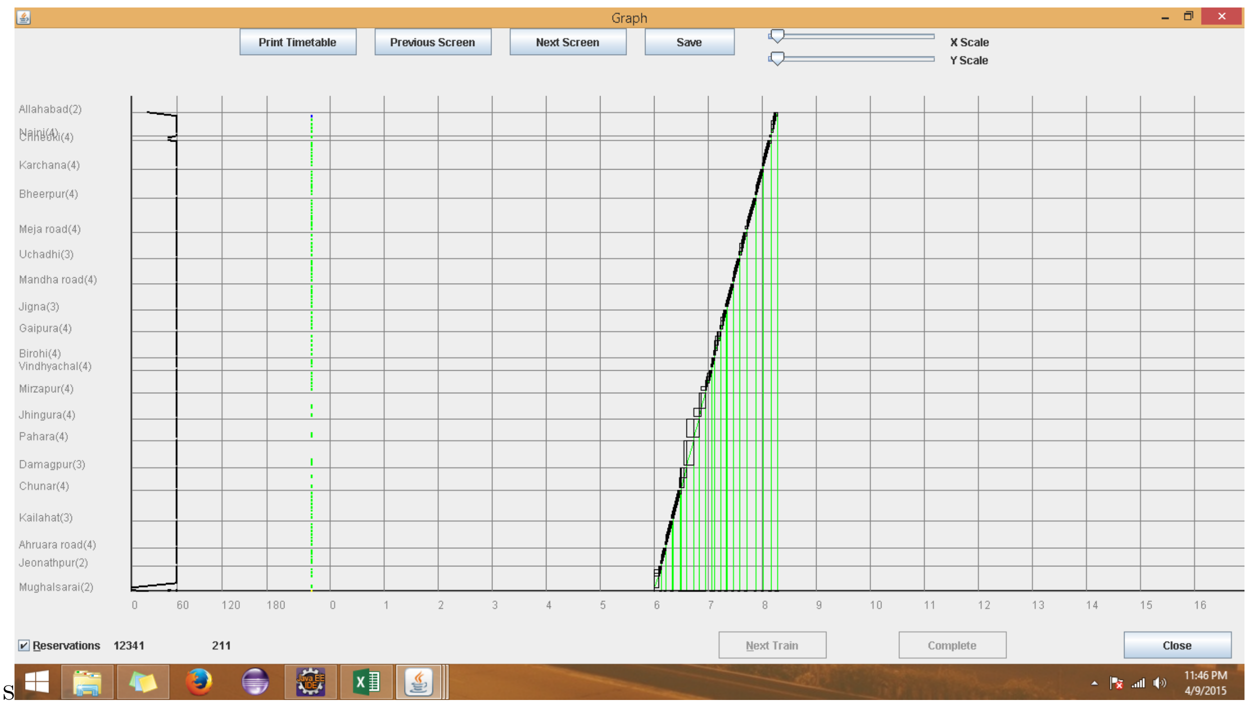

Figure 6.8: ALD-MGS single train path

As seen in this figure and as expected, the block occupany at the 2 absolute blocks, namely Dagmagpur to Pahara and Pahara to Jhingura are the highest. Ignoring the yard bottleneck at ALD, this would likely be the second biggest bottleneck. Even having one absolute block while making the rest of the section automatic doesn't improve the capacity at all as the absolute block limits the throughput. Thus, to improve capacity, the complete section should have automatic block signaling.

2. The option for twin single line signaling should be there between Naini and Allahabad. This option may be used at certain times of the day, when there is directionality of

traffic. At best, it will ease the congestion in the yard, if used appropriately. At worst, we will not lose anything by providing this flexibility. Although it seems like a win-win, the controllers will have to be trained sufficiently on how to operate trains judiciously. Clear periods of time when there is directionality of trains need to be determined. Twin single line operations in Vijayawada can be used as a model to study for implementation at ALD.

3. All terminating trains be considered for movement away from ALD, either to Subedarganj or to Allahabad city, so that platform and line occupancy at ALD is minimized and at least some shunting movements are shifted. This will lead to alleviation of the bottleneck, i.e. the ALD node.

4. The two daily trains Howrah-Mumbai mail via ALD and the Yeswantpur Patna express should bypass ALD completely, by using Link Junction – Cheoki. It is obvious that this will create additional paths on the Naini ALD section (including the slow movement on the bridge). The resultant savings can very well justify additional passenger facilities at Cheoki. In the section, it is clear that the traffic is heaviest between Naini and Allahabad from the hypothesis that ALD is the bottleneck. The suggestions regarding Howrah Mail via Allahabad and Yeswantpur Patna are relevant in this context also.

5. As seen before, the absolute blocks between Dagmagpur-Pahara and Pahara-Jhingura are the critical blocks and limit the throughput of the section. It is also known that using loops at a station (not the mainline) involves speed restrictions due to the turns taken while entering the loop. Till this signaling improvement is done, the up loop at Dagmagpur and the down loop at Jhingura should not be used at all, except in emergency situations. This will ensure that movements on these sections will happen faster and the effect of the critical blocks won't be exacerbated.

6. Re-timetabling of some passenger trains can improve the path availability for freight trains in this section. Platooning of trains is one way of re-timetabling. Platooning means grouping trains of similar speeds together. In the following hypothetical scenarios, fast trains and slow trains (all fast/slow are of one type) go one after the another alternatively in the first figure while platooning happens in second scenario. Faster train group is pushed earlier in the first scenario:

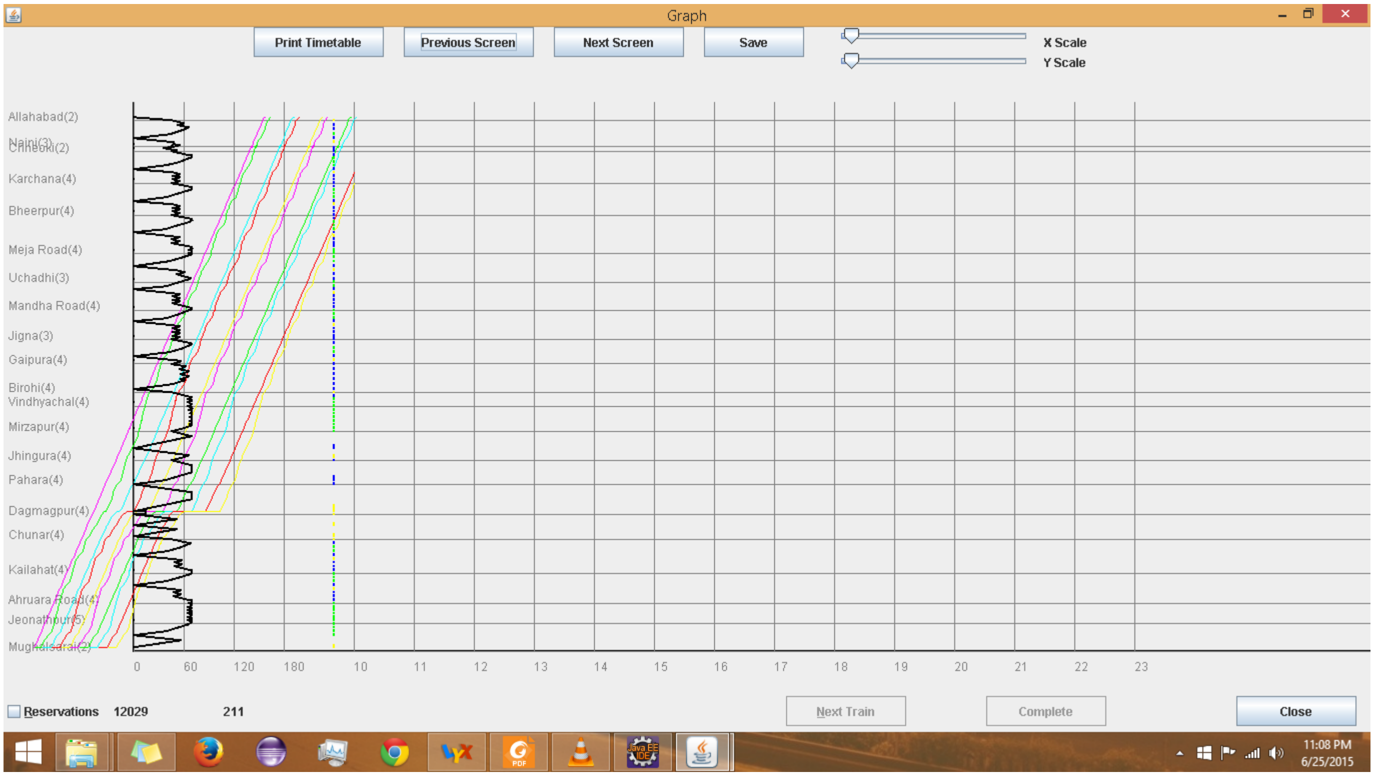

Figure 6.9: Fast and slow trains moving consecutively

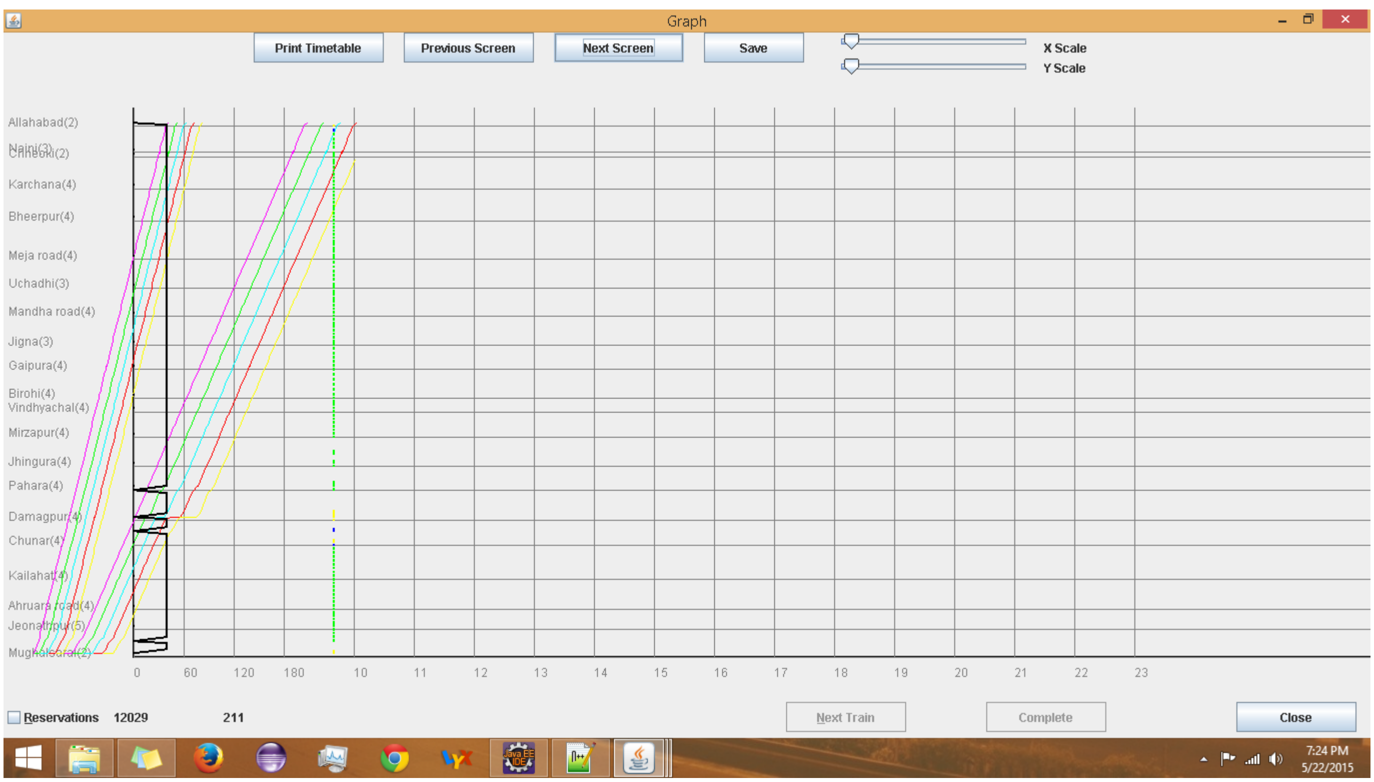

Figure 6.10: Platooning of trains

Faster trains have speed of 70 kmph while slow have speed of 40 kmph.11 minutes are saved on the ALD-MGS section by platooning trains as mentioned, which can be roughly used to create an additional train path. The larger the speed differential, the larger the extra capacity obtained via platooning.

7. As a short term measure, all scheduled overtakes by passenger trains on this section should be avoided to the extent possible. Overtaking should be allowed only if the speed differential between 2 trains is large enough, otherwise the slowing down and waiting of the train being overtaken in a loop might lead to an eventual system loss.

8. It is also known that the receiving of trains at MGS is fairly slow: passenger trains take around 10 min while freight around 15 min to be received. This is because of the complex structure of the MGS yard which involved a lot of crossover movements. This is clearly thus another bottleneck for movements in the opposite direction. Better understanding of the yard needs to be done in order to come up with strategies to alleviate it. It may even be possible that this is a bigger bottleneck than the absolute blocks on the track, thus making the tracks automatic might not help in increasing the throughput! (Assuming all trains traverse the whole section.)

All of these strategies will work in theory, but their numerical impact on capacity should be estimated, which hasn't been done as part of the current exercise.

Chapter 7

Conclusion

The work done primarily revolves around estimating capacity of a section under different scenarios. Capacity is a multifaceted object and will vary from section to section as the system changes. The 4 main factors involved in the definition of capacity are the number of trains, heterogeneity, average speed and punctuality. The definitions adopted for different sections involve a combination of these factors. UIC 406 gives a general framework for modeling capacity constraints and the possibility of defining capacity on Indian sections using this framework should be examined.

Chapter 3 and 4 have demonstrated how capacity of a section is impacted by the signaling technology being used. There is a significant improvement when going from absolute block signaling to automatic/cab signaling on long distance tracks (inter station distances are long). On Mumbai suburban corridor, there is a 15% increase in capacity over the best that can be done with 4 aspect signaling. The freight traffic can be increased to 3 times of present on the SC-W section (with punctuality constraints) by moving from the current IB system to automatic signaling.

Basic operations of train running including different types of restrictions and the interaction between two trains is discussed in chapter 5 in hopefully a fairly exhaustive manner. This also consolidates our understanding of train path generation which takes place in the IIT Bombay simulator. Important concepts like level crossings and maintenance blocks are yet to be modeled in the simulator framework and should be done to get a more accurate abstraction of the real world.

Chapter 6 shows that it is pointless to discuss 'local minimas': Only if every part of the system is bettered will the system better. It is important to identify the weak link (slowest step) of a chain of sequential events and make that step faster in order to improve the throughput of the whole system. In context of the ALD-MGS section, these weak links are the ALD yard, the two absolute blocks discussed in Chapter 6, and the receiving at MGS. Any attempt to increase the section capacity should be an attempt to alleviate these regions.

Suggestions have been made to improve the operations on this section, but they should be quantified before being implemented.

Chapter 8

Appendix

Three different versions of the IIT Bombay railway simulator exist which have been used in the course of this project. These simulators have the same logic in terms of the inputs and outputs to/of the simulator, but the method of inputs are different in each of them:
1. The basic version of the simulator was developed as part of a project with IRISET. In this version, the inputs for the user were in the form of text files. This was cumbersome for the user, as details up to loop/block IDs as well as their linkings and crossovers are expected to be entered by the user, that to directly into the text files. The next 2 versions alleviated these shortcomings of user friendliness. The usage manual for this version has been documented in [3].
2. New Sketch Rail Simulator: This version was developed as part of a student's dual degree thesis. In this version, the section details are entered by physically sketching the stations, loops and blocks as explained in the manual instead of inputting via text files.
3. Form based GUI developed as part of a project with RDSO: No ID's or linkage details need to be mentioned in this version of the simulator. They are automatically assigned, created according to the other inputs given via the interface as explained.

Operating manual for the form based GUI of the IIT Bombay simulator

The following are the steps required to enter section details via the GUI interface. Again, we assume that the tool is already loaded on the workspace of Java eclipse of the user's machine:

1. The first screen of the interface. This is where the user chooses to make a new section or run an existing section. The 'report' and 'options' tab aren't functional as of the writing dates of this report. Select 'GUI input' from here to input a new test case.

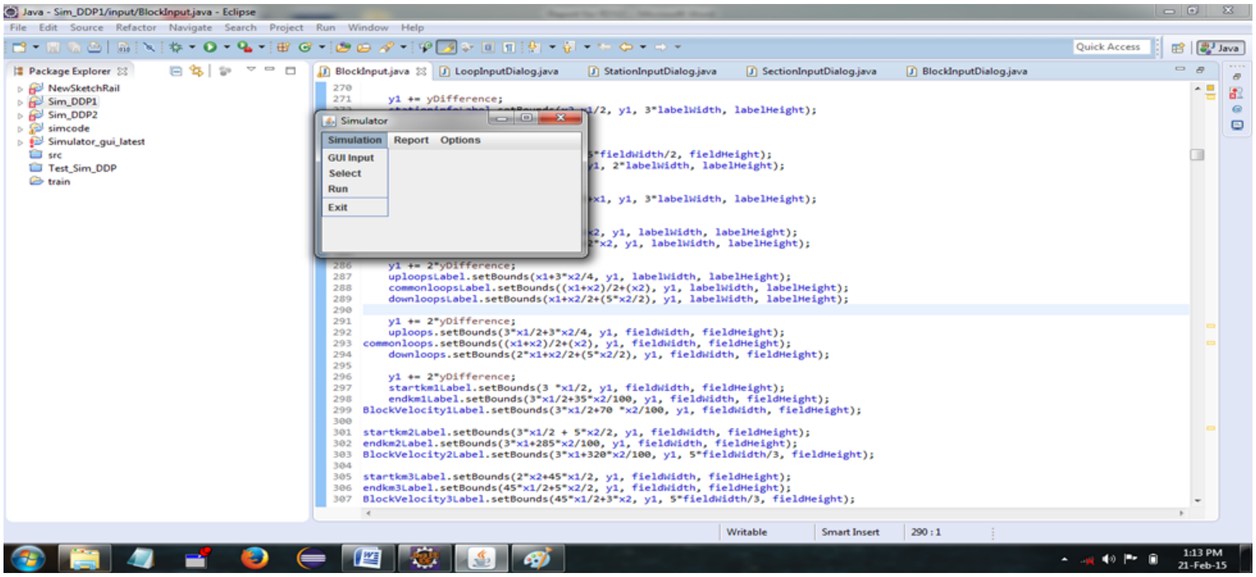

Figure 8.1: Simulator main window

2. This is the first screen when the user decides to make a new section. It requires input data about stations and their loops, the blocks which connect the sections, global parameters of the section, gradients on the section and the trains which the user wants to run on the section, both passenger and freight.

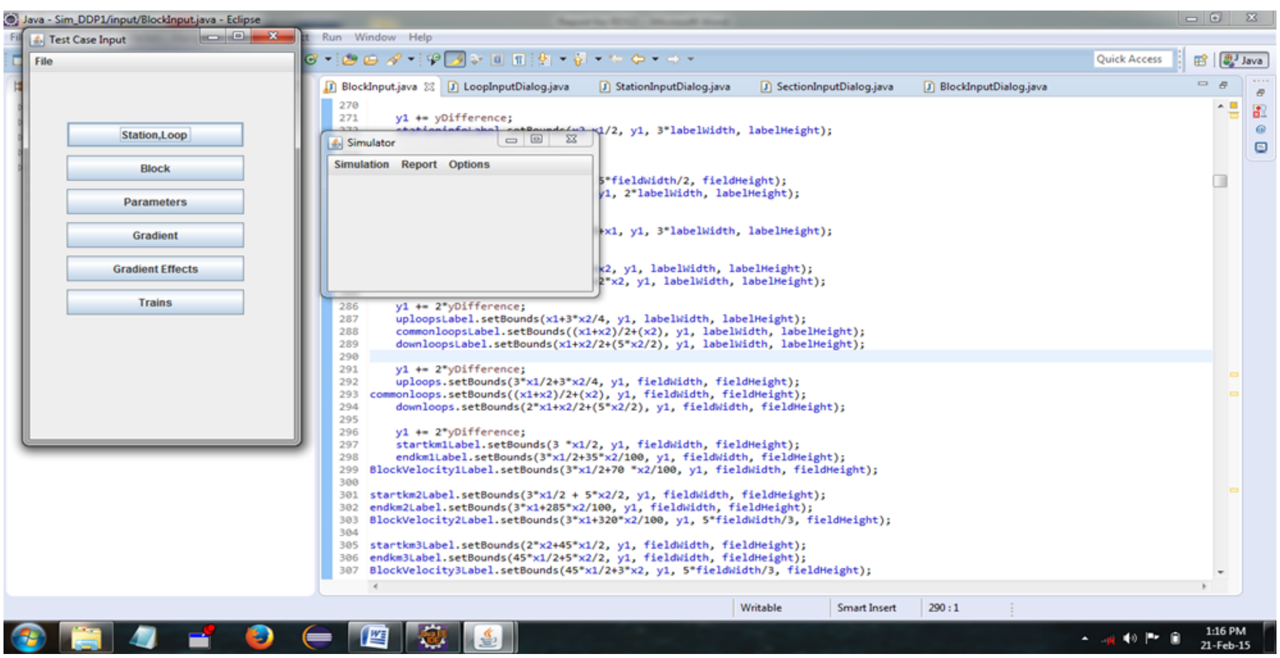

Figure 8.2: GUI input main window

3. Station input: First, the user must begin by entering the data of the station: The data required for a station are the station name, their start and end mile posts and station entry

speed restrictions. After this, the particular station can be added to the section by clicking on ‘add station’. This will lead to the next screen where the user enter the loops at the newly created station.

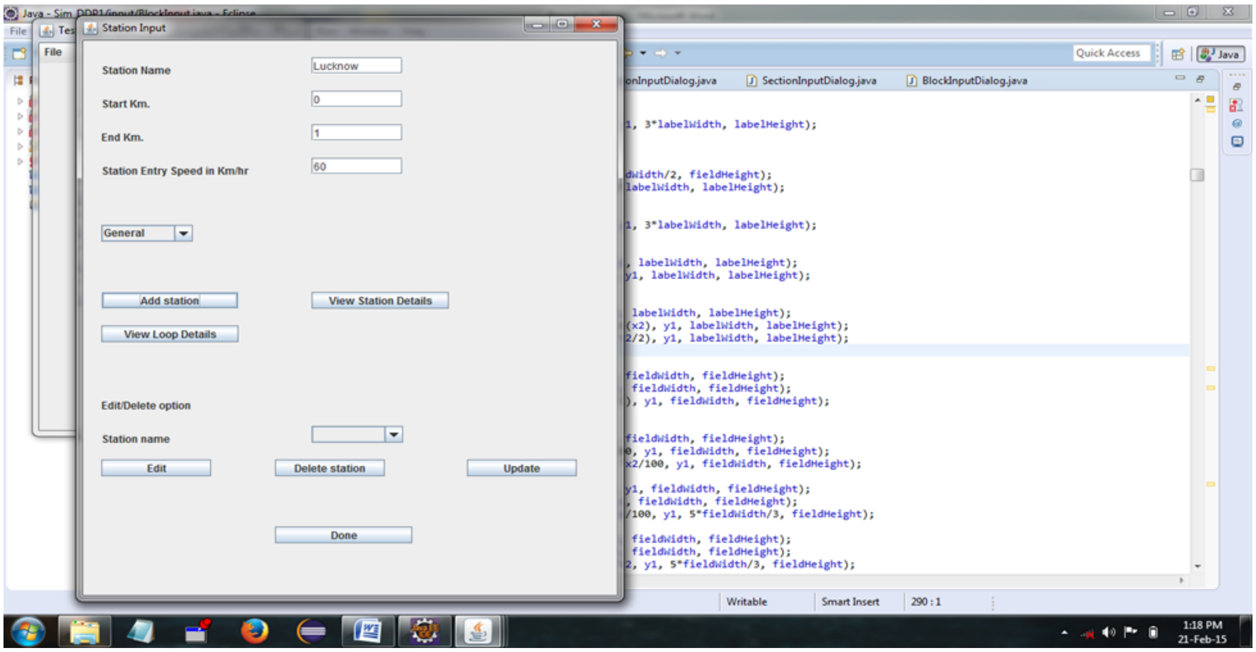

Figure 8.3: Station inputs

4. Loop input: There can be multiple loops at a station and of different types: Mainline, side loops and also in up, down or common directions as required (see image below). Click on ‘Add new loop’ when done adding all the details. The loops created at the station being considered can be viewed by clicking on ‘view all.’ Once done with adding the loops, click the ‘Done’ button on the loop window to get back to the station window. Add as many station as required in the same way. When done, click on the ‘Done’ button on the station window. Now, the user is ready to make blocks to create lines between stations.

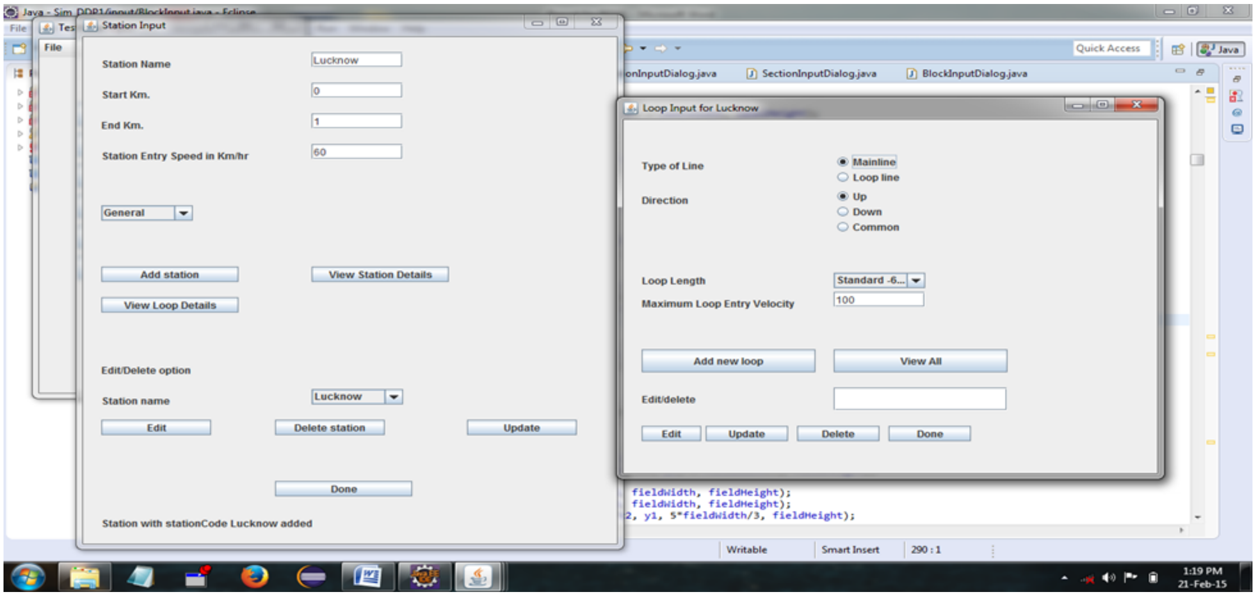

Figure 8.4: Loop inputs

The next screen shows the loops made at the station just shown to have been created: (This window will show when the user clicks on view all buttons of a particular station.)

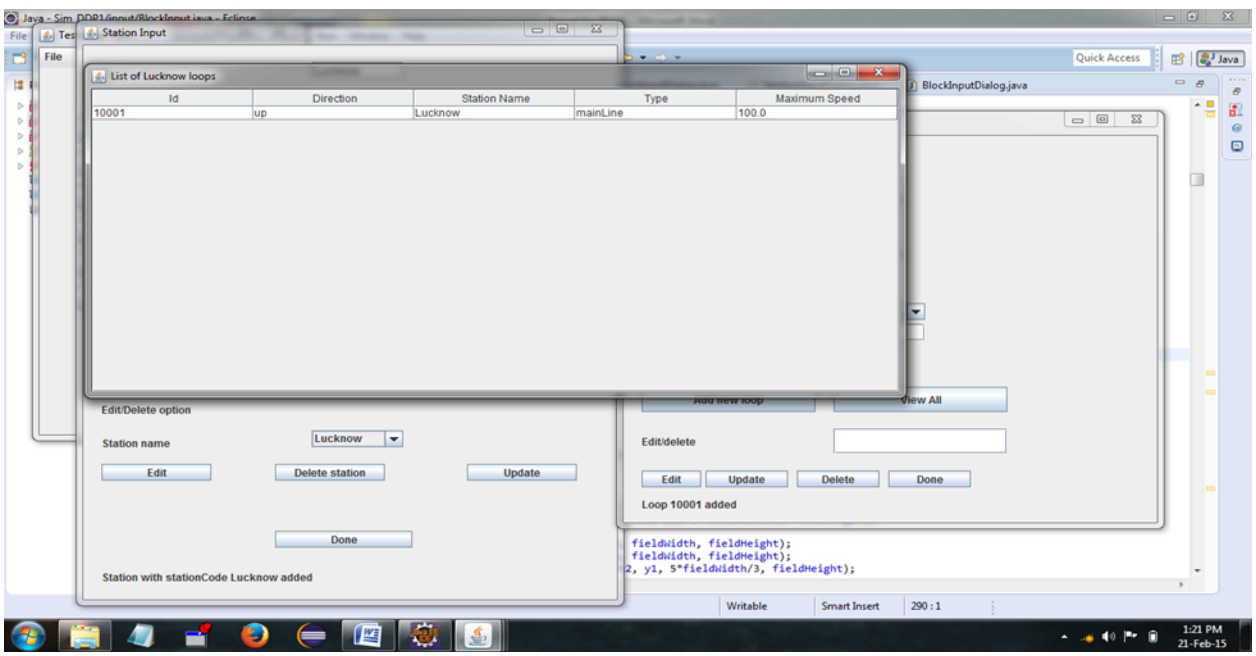

Figure 8.5: Loop list

5. Block input: Blocks need to be added between consecutive stations to connect them. In the first drop down list, the user needs to choose the station from which he needs to make the blocks (the consecutive station pops up automatically. Block types can be Bidirectional (i.e. a single line in which the trains can go in either direction using the Bidirectional block

resource, or it can be a double line (i.e. 2 lines between the consecutive stations, each dedicated for a particular direction. There are 2 cases:

Case 1: The user chooses a double line. In this case, there will be 2 lines and it's required to enter the block data for both directions. There can be up to 10 blocks in each direction. The input is their start and end mile posts, maximum speed in a block, and possible speed restriction in the given block. Note that there can be multiple set of speed restrictions within the same block, but the simulator just accepts one restriction.

Case 2: The user chooses a bidirectional line. As the blocks are common, there will only be a single set of inputs for the blocks rather than 2 different ones for up and down seen in double line.

Once done creating blocks between the considered stations pair, click on the 'submit' button. Repeat the process for all possible station pairs and hence connect all the stations in the section.

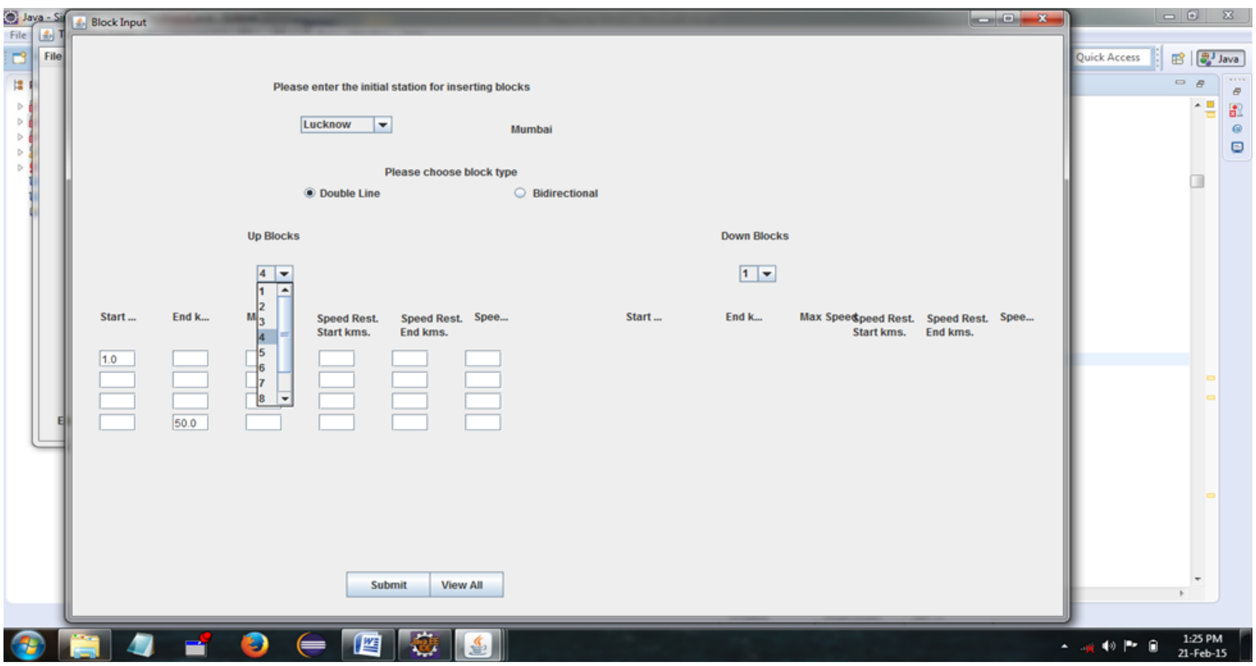

Figure 8.6: Block inputs

6. Parameter input: Enter the 'simulation time' (in days) according to the simulation length the user wants to see, the 'block working time' in min (the time after which a signal will turn from red to a go signal after the train has left a block. i.e. it's a delay time and the 'number of signal colors' (2,3,4 etc.)

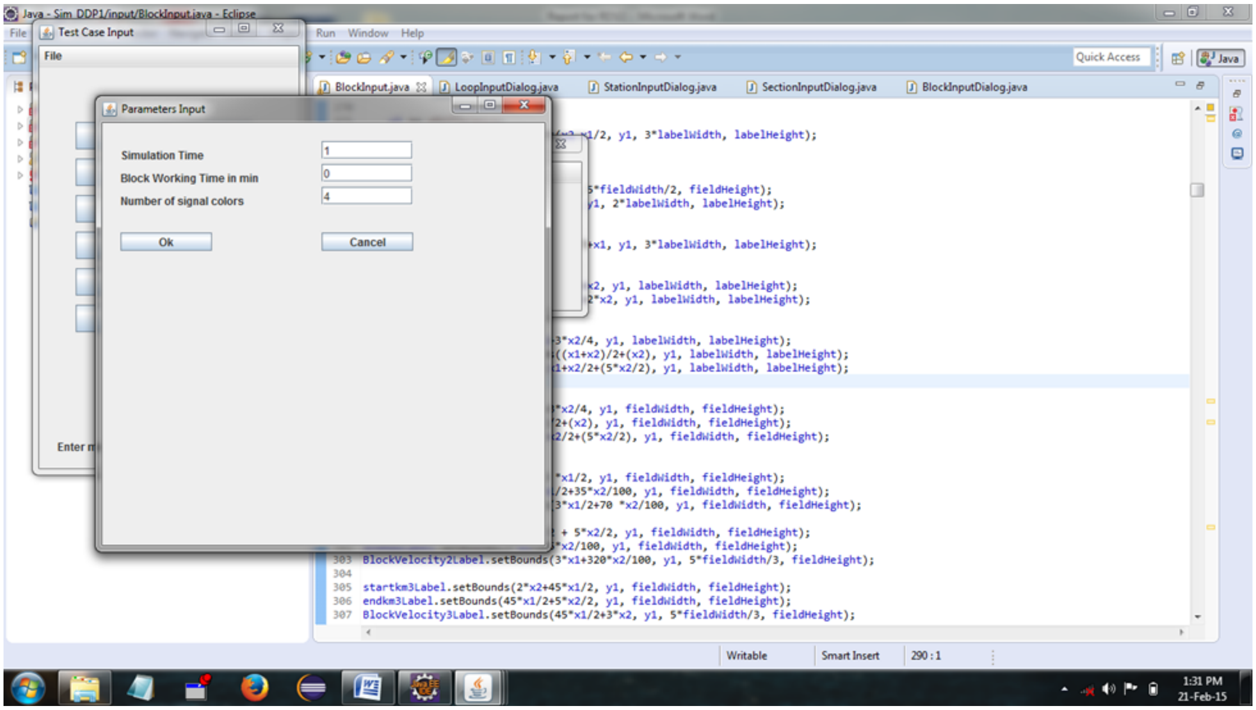

Figure 8.7: Parameter inputs

7. Gradients: The ground on which the train travels can have gradients. The user can mention those in this file. The inputs include the 'gradient value', 'start km' and 'end km' for the whole section. Note that gradients can take any real values, but a finite discrete set has been assumed for simplicity. Standard effects of the gradients on acceleration and deceleration have already been included in the simulator and don't need to be added manually. Now the user are done with creating the section. The last part is to enter the train data for the simulation.

Note: Effect of gradient on acceleration and deceleration: Up gradients will decrease acceleration and increase deceleration while down gradients will not affect acceleration (as driver will be cautious) and decrease deceleration according to the following table:

Gradient magnitude (direction considered in next columns) Unit: meter/meter	No effect on acceleration if gradient in DOWN direction Unit (m/s2)	Decrease in acceleration if gradient in UP direction Unit (m/s2)	Decrease in deceleration if gradient in DOWN direction Unit (m/s2)	Increase in deceleration if gradient in UP direction Unit (m/s2)
1/100	0	0.15	0.15	0.15
1/150	0	0.13	0.13	0.13
1/200	0	0.11	0.11	0.11
1/250	0	0.09	0.09	0.09
1/300	0	0.07	0.07	0.07
1/350	0	0.05	0.05	0.05
1/400	0	0.03	0.03	0.03
1/450	0	0.01	0.01	0.01

Table 8.1: gradient effects

Gradients smaller than 1/450 are almost close to level ground in terms of their effects on acceleration and deceleration and gradients larger than 1/100 are generally avoided by trains. Also, down gradients don't increase acceleration as drivers are cautious on down slopes and maintain low accelerations by braking often.

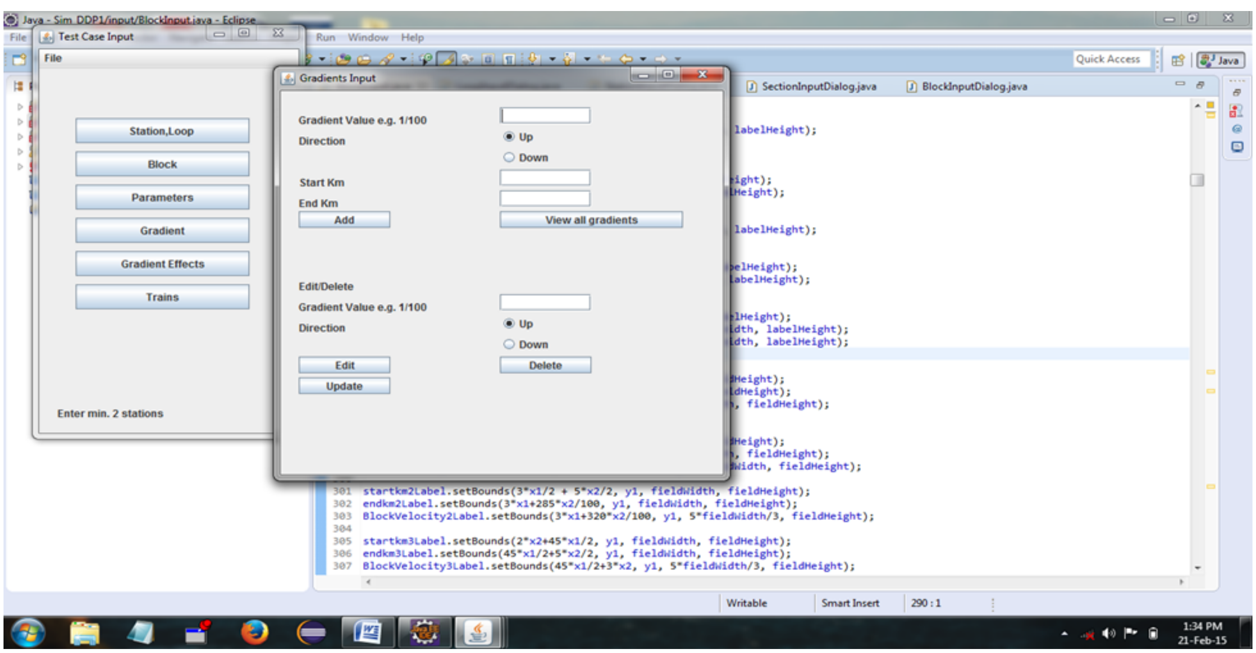

Figure 8.8: Gradient inputs

8. Train input: Trains can be passenger or freight. For each train, the user needs to enter its length, acceleration, deceleration, maximum speed, the direction in which it is going

(up or down) and a unique ID. For freight trains, only the start and end loops along with the starting time need to be mentioned, while for passenger trains, arrival and departure time for all stations it is planning to go through are required. The priority for trains is also incorporated in the simulator. Simulating with only 1 freight train is considered in this version of the GUI as the primary interest is in getting an idea of the capacity using Scott's formula.

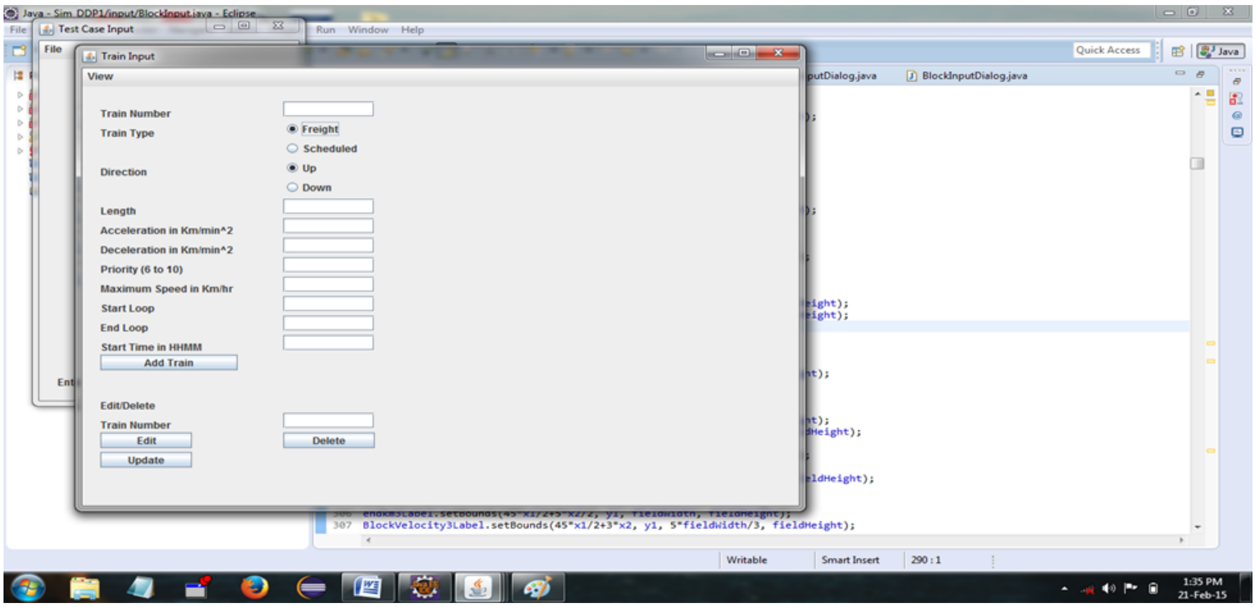

Figure 8.9: Train inputs

9. Building the simulation: The user is done with all the inputs now. From the simulation main window, choose 'select' from the simulation drop down menu. Enter the relevant file paths in the specific field as seen in the window. After that, click on 'run' from the simulation drop down to see the familiar output of the simulation next. (Note: This version of simulator also gives capacity as one output calculated using Scott's formula.)

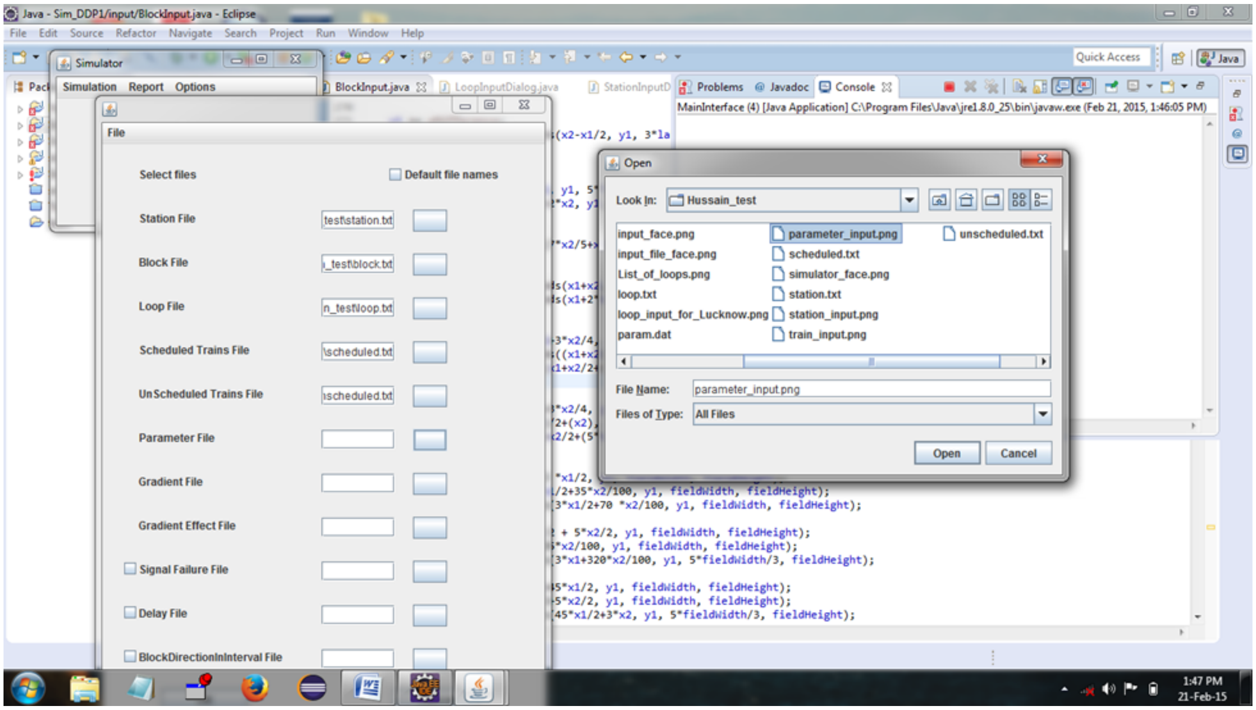

Figure 8.10: File selection window

Bibliography

[1] Research Design and Standards Organization, Line Capacity Calculation documents

[2] Online weblink: http://www.cr.indianrailways.gov.in/view_section.jsp?lang=0&id=0,6,287,394,576

[3] Keshav Singh, IITB, "Modeling and Design of algorithms for rail operations and Capacity Planning" (Ch2: Simulator by IIT Bombay), M.Tech dissertation in Industrial Engineering and Operations Research, June 2013

[4] Landex, Kaas, Schittenhelm, Schneider-Tilli, "Evaluation of railway capacity"

[5] Van Jalentin, (2011) "Traffic Flow Theory", Prague, Czech Republic: Czech Technical University in Prague

[6] Tobias Lindler, "Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity", Journal of Rail Transport Planning & Management 1 (2011) 49–57

[7] A. Ranade, D. Shelar, P. Gupta, N. Rangaraj, R. Gopalakrishnan, "System Headway on the slow corridor for suburban services (Western Railways - Mumbai)" , October 1, 2014

[8] Yong-Cai Wang, Qian-Chuan Zhao and Da-Zhong Zheng, Bottlenecks in Production Networks: an overview

[9] Online weblink: indiarailinfo.com

[10] Online weblink: http://en.wikipedia.org/wiki/Railway_signalling

[11] South Central Railways, Working time table no. 69

[12] H. Bharmal, S. Bothra, S. Dangayach, N. Rangaraj, "Section capacity analysis for different railway signaling technologies", International Simulations Conference of India, Pune, 2015

[13] A. Ranade, D. Shelar, P. Gupta, Narayan Rangaraj, R. Gopalakrishnan, "System Headway on the slow corridor for suburban services (Western Railways - Mumbai)", October 9, 2013

Acknowledgements

I've been very fortunate to be in IIT Bombay, a place where students can exercise a large amount of academic freedom. I would like to express my gratitude to my supervisor Professor N. Rangaraj for being extremely patient with me and giving me opportunities to work on interesting and practical problems. I want to thank all the professors in the IEOR Department, each of whom have taught me some course(s) and wired my way of thinking in its current state. I would like to thank my parents for supporting me in my endeavors and friends for their pleasant company. Last but not the least, I am indebted to Led Zeppelin and rock music at large, which has accompanied me through good times (and) bad times.

Hussain Bharmal
Department of Mechanical Engineering
IIT Bombay, Powai.