

A Machine Learning Approach to Eliminate Bus Bunching of Public Transportation

Harshit Kaushik, Pallavi Tomar, Shrishti Katiyar, Prachit Luthra
B.Tech. CSE, Department of Computer Science and Engineering
SRM Institute of Science & Technology,
Delhi - NCR Campus, Ghaziabad, India

Partha Sarathi Chakraborty
Assistant Professor, Dept. of CSE,
SRM Institute of Science & Technology,
Delhi - NCR Campus,
Ghaziabad, India

ABSTRACT

The framework presents the opportunity to provide a solution in dealing with the problem of Bus Bunching (BB) which adversely affects the traffic in India. Recent advancements in technology have allowed real time monitoring and interactive usage of the Public Transport (PT) System. Regression Analysis and Gradient Descent are the Machine Learning (ML) principles used to make a predictive methodology that helps to prevent BB. The methodology is then used to provide a Corrective Solution (CS) in order to eliminate the BB event. The proposed method is performed for Kaushambi - Meerut bus route using the schedule of the bus time table and various assumption based data for traffic and BB events. The system proposed could be used to provide a decision support system and improve the control room operations. This methodology considers the adverse effects of Bus Holding (BH) and provides discrete and streamlined solutions, given the complexity of BB problem.

General Terms

Intelligent Transportation System, Public Transportation

Keywords

Bus Bunching, Machine Learning, Public Transportation, Traffic Congestion, Predictive Methodology

PT	Public Transport
BB	Bus Bunching
ML	Machine Learning
GPS	Global Positioning System
PDF	Probability Distribution Functions
AVL	Automatic Vehicle Location
CS	Corrective Solution
BH	Bus Holding
SS	Stop Skipping

1. INTRODUCTION

In the recent age of technology, people have inherently started using innovations in their daily lives. The advancements in telecommunication and the wider reach of internet has enabled this technological boom. The data storage complexity has now become irrelevant

with huge amount of data storage possible in a micro-chip enabling real time sensors to collect data and embedded systems to provide a solution in a span of seconds.

These innovations are hugely being used in present day transport. The availability of current location using Global Positioning System (GPS) over the internet has enabled the development of large scale frameworks to monitor and control the public transportation. [1] Many industries are taking the advantage of using this data to provide services and monitor their operations. One such application of these technologies can be used for Public Transportation.

The PT System across the globe have acknowledged the advent of technology and have deployed Control Centers that monitor the real time information from the fleets for better operations and services. This in turn helps in maintaining the service reliability of the service providers resulting in a good service performance. The service reliability plays a key role for user satisfaction, especially in countries where private companies provide public transportation services.

The reliability of service can be understood as a combination of

- (1) The timeliness of the service - service provided in planned schedule.
- (2) The frequency of service - evenly spaced vehicles to accommodate adequate passengers.

Both these reliability metric, when in unison provide a very reliable service to the passengers and also help maintain the profitability of the company.

At present, public transportation in India uses few of these technologies to determine the current location, ticket booking, online time table and few more services. With more traffic flow and increasing vehicles on road it has become evident that buses have difficulty maintaining their schedule due to changing traffic conditions. This eventually leads to deviation from usual speeds of the bus meeting traffic and the bus next in schedule reaching closer to the bus ahead. Upon reaching a free flow traffic condition the earlier bus reaches the concurrent stop where more passengers have arrived and are waiting.

The waiting passengers board the current bus and the bus departs, while in the meantime, the next bus is also about to reach the stop but with fewer waiting passengers as the arrival time difference be-

tween the buses has now become lesser than scheduled. The event continues making a chain for the next stops on the route. This event results in the collection of buses as platoons or groups of buses that were supposed to be evenly spaced but end up reaching same place at around same time. This is referred to as BB. [8][14][15]

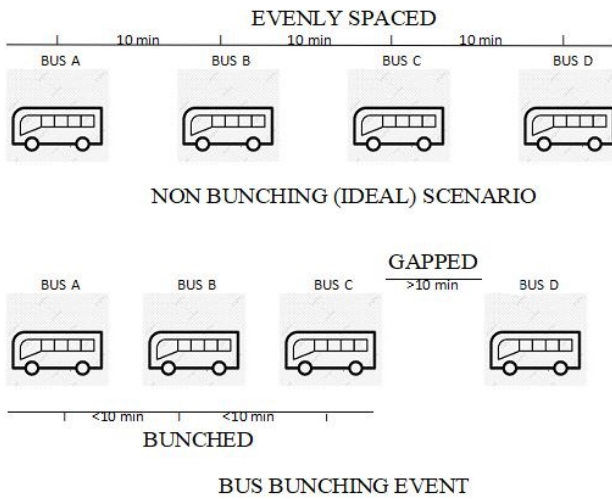


Fig. 1. BUS BUNCHING ILLUSTRATION

The occurrence of a BB event is a clear indicator of an unreliable service. The running of more than two buses together on the same route is a sight that no service provider intends to see, therefore better operation control is needed to counter this event. There arises the need to provide correction actions for the BB event keeping in mind the practical issues that may be linked with the impact. This project aims to provide the solution using ML principles of Regression Analysis and Gradient Descent using the offline and online data to make decisions.

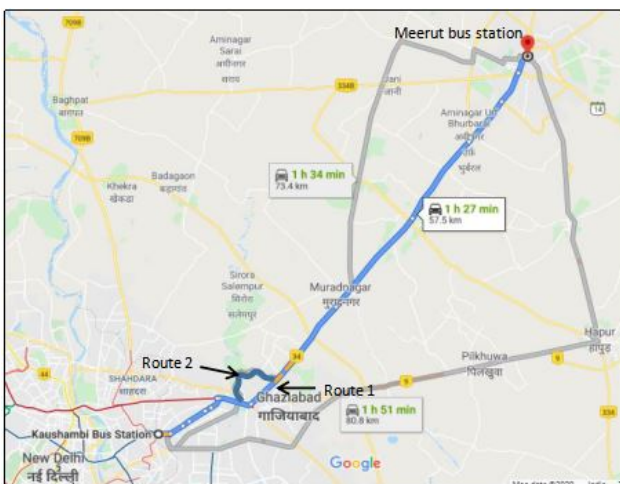


Fig. 2. BUS ROUTE MAP

2. LITERATURE REVIEW

Previous studies and works in this field have enabled the use of a wide range of analytical and operational models. These models depict the bus service operations fundamentals and providing simple control solutions along with their impacts on fleet operation. [9] The analytical studies conducted earlier examined and brought to consideration the characteristics of BB and the unreliability of service that happens due to recurrent changes in traffic conditions. Later these studies used Probability Distribution Functions (PDF) to find relationships in the data and preparing a probabilistic model that helps in identifying the event of BB.[7] In recent decades, the analytical models have advanced and have contributed in creating a control operation that enable in adjusting the bus schedule by managing the bus stop duration time and the bus running time between two successive stops based on the traffic conditions ahead.

The control strategies that have been recommended in the operational models generate solutions taking into view the best possible situation to provide a better performance and service [5]. In recent times the BH strategies are being commonly used to control the bus service and for providing a regularity in service [10]. The implementation of these strategies to counter BB is done by determining

- (1) What decision has to be taken
- (2) How to carry out the decision
- (3) Where (what location) to deploy the decision

These BH operations have been accounted for minimizing the total waiting time of passengers that have been proven in various mathematical models.

The previous studies [1] have turned out to be effective in reducing the intensities of occurrence of BB, thereby improving the reliability of service. There have been propositions of a real time solution in a few studies. However, these studies have remained unsuccessful in providing a solution for the BB event in real time so far. These studies propose a real time elimination of BB taking into use the methods of ML providing comprehensive solution to make bus service perform even better.

This study provides a methodology to provide a complete bottom-up solution to the BB event. For this methodology [2] it is proposed to use Automatic Vehicle Location (AVL) data that constitutes a self updating stream of data arriving from the fleet. Batch learning on this data is performed next in relatively small data sets to train the algorithm. Since a prediction framework for BB occurrence is prepared in this method, it generates output while collecting data. In order to limit the output generation, another algorithm using Drift concept is incorporated that handles the output by adapting to the learning model of the current system state set.

The predictive methodology this study proposes uses [4] Regression Analysis, Probabilistic Reasoning [16] and Perceptrons learning with Gradient Descent [13] for reaching a solution from a huge data set that continuously keeps adding data from fleet. The modelling framework to prevent BB from happening in real time proposes:

- (1) An algorithm using regression that can provide accurate predictions for Drift
- (2) The probability for prediction of BB occurrence further down the route
- (3) To select and implement a solution to correct and prevent BB from happening in real time

After the implementation of a solution, the impact of operations on the service reliability is evaluated. This study is based on the data provided by STCP (Sociedade de Transportes Colectivos do Porto) [3] for 18 routes that their fleet operates in. This generated a large data set of 18 routes along with records collected over a period of one year. This data set was pre-processed and then the learning could be performed. Operational Control and Operational Planning are the two concepts that the study proposed to evaluate reliability of service.

Operational Planning considers all the factors that can be planned and executed which includes bus schedule, driver behaviour (driving habits), passenger boarding and deboarding at stops, etc. Operational Control on the other hand handles more chaotic situations that include traffic congestions, mechanical failures, etc. For these concepts, the prediction of BB event is evaluated at Planning level and at Control level, the solution is provided and that solution is evaluated in terms of service reliability. A travel time prediction is made that helps in both Planning and Control Operations.

The control strategies the study proposes to mitigate BB at stop or between stops are BH and Stop Skipping (SS). BH is commonly used to regularize the bus service[11]. It is usually performed by instructing drivers to hold until the scheduled time at some pre-defined stop when bus arrives early thereby maintaining the schedule. But this strategy fails to be effective when the bus operates on a high frequency route where waiting passengers determine the service reliability. SS is a rare event where bus is allowed to skip some stop informing passengers on board to maintain the schedule / service reliability and shorten the travel time to some extent. Therefore, these actions require a detailed sought out solution to deploy and correct BB.

The study [12] revealed that the performance of the predictive framework varied from route to route, even for opposite line of the same route. However, more than 83% of predictions for BB turned out to be accurate. The framework failed to be effective on low frequency route and was mostly prevalent for high frequency routes so, the correction solutions deployed in the study regulated bus progress on these routes. 67.59% of the BB occurrences were reduced using this framework. This is a good success keeping in mind the forecast of 83% accurate predictions [1] [15]. The framework can be used in a decision support system as it is as the main prerequisites are already fulfilled in this methodology. However, it is difficult to communicate to a driver to vary bus speeds constantly. The prevailing attitude towards BB as a regular and inevitable feature of bus travel is also a mindset that needs to change considering the effective solution rate.

3. SCOPE AND OBJECTIVE

The studies conducted so far are hugely dependant over the probability of the occurrence of BB event and that it can be solved by stabilizing the progress of fleet under non favourable traffic situation. This can be done by reducing the difference between actual journey time and the scheduled journey time of the bus [6]. Achieving results using this method needs control actions regarding speed variations, BH, etc. that turn out to be frustrating for drivers to constantly vary their driving speeds, deviating from their natural driving habits. This leads to a lower compliance of control action.

Therefore, to estimate and identify BB event and solving the event in real time requires a proactive approach and not reactive approach. ML has thus become handy and essential to perform learn-

ing on the database and produce some output while collecting data to react to variable traffic situations.

The objectives of this project are:

- (1) Developing a regression algorithm that can correctly make predictions.
- (2) Implementing a CS to improve the service reliability.

4. IDENTIFICATION OF RESEARCH GAP AND PROBLEM

The studies have provided analysis and frameworks for the problem but some areas still remain untouched or less analyzed. These include both Operation Planning and Operational Control which are:

- (1) Accurate traffic prediction
- (2) Driving habits of drivers
- (3) Incidents of bus breakdown
- (4) Presence of very limited and inconsistent statistical data
- (5) Co-ordination due to changing traffic conditions.

5. CHALLENGES

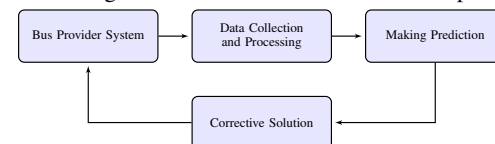
- (1) The major challenge is reducing the discrepancies in the departure of buses from depo according to schedule and realtime running status.
- (2) The UPSRTC leases bus permits to private tenders keeping control with officers and bus conductors which needs an effective co-ordination.

6. EXPECTED IMPACT ON ACADEMICS/ INDUSTRY

This project aims to provide an India centric solution for fleet operations, thereby increasing attention towards betterment of public transport in India. This would also enable passengers to get a real-time and interactive bus transport i.e. estimated times of journey, arrival and departure, time-table, bus location, bus information, etc.

7. UNDERSTANDING BLOCKS

The design of our simulation is based upon the blocks as



Bus Provider System This is the block that contains all information regarding the bus service that includes the schedule of the buses, route information, bus information, etc. This is the data that the system already knows and is used as a marker for the real-time service.

Data Collection and Processing The real time data is collected from the fleet at this stop. For a well prepared data repository, it is easy to add data as it is, but for deviation and noise in data collected, processing of data must be done before adding data to database. It is important for a system that generates huge amount of data to be free from inconsistencies to provide an output in real-time.

Make Prediction Graphical method using linear regression and gradient descent is used to accurately predict the future outcome based on data provided to the system. This is thus helpful in providing a solution to correct fleet operation in case of deviations from their schedule.

Corrective Solution Based upon the predictions made a CS is achieved. It is important to strategically implement the solution so that BB is eliminated without further deviating the fleet schedule other than the buses that most require the correction.

The flowchart for recursion is illustrated in figure 3.

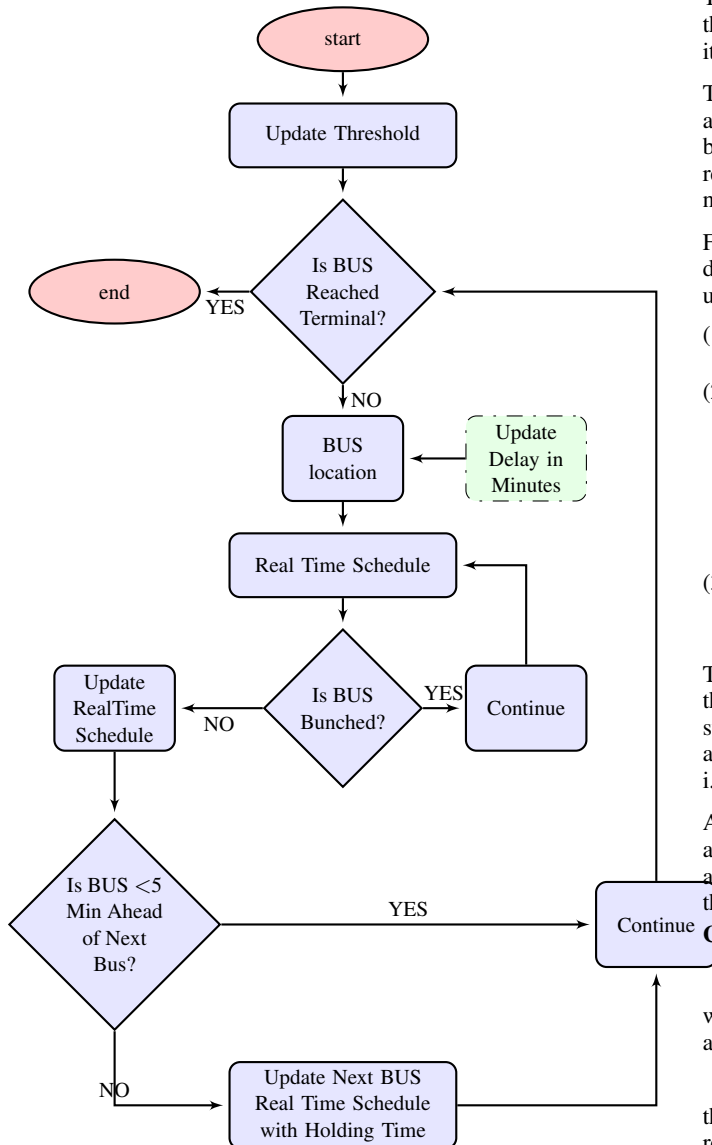


Fig. 3. Simulation Process Flowchart.

8. CODING AND TESTING

BB event occurs by ever changing traffic conditions and hence it becomes difficult to predict this in real time . Since this method is operated on one route therefore , the data present is very small. Also , it is first of a kind bus operation control in India ,the database for our consideration is also limited. The time schedule of the bus depot is collected from UPSRTC .Other data are assumption based to carry out a simulation for the Kaushambi-Meerut route.

It is a high traffic congestion route where the traffic flow can turn bad in a short span of time. Considering this as the main issue, the learning rate for our simulation is given to time. Passengers that board buses on this route are used to heavy traffic and bus delays. The BB event is neglected by passengers and transport providers that has developed a notion of normalcy in the society . However, it could be changed by solving BB.

To do so, by simulating a BB situation over the route, creates a chain of events usually. But in this methodology, the deviated bus tend to generally follow their original schedule by the time it reaches the terminal and other buses to be least affected that come next to the deviated bus, unless the buses are stuck in a traffic jam.

For this, the original schedule of the bus is used as a marker. Any delay in some bus results in the deviation of the bus from its schedule. It could be delayed in cases:-

- (1) Delayed from arrival time, earlier than departure time the bus departs from the stop at the departure time.
- (2) Delayed from both arrival and departure time the bus arrives at the stop and departs from stop after the stoppage time from original schedule gets completed.
 - (a) If the next bus is sufficiently close, the next bus is delayed to a time marker which is 5 min in our simulation.
 - (b) If the next bus is not sufficiently close, there is no change in the schedule of the next bus.
- (3) In case of mechanical failures, the bus will be hugely delayed, in that case, the next buses do not see any traffic congestion and follow their schedule.

The simulation takes use of the system time to align all buses to their schedule. It is used to manage arrival time, departure time, status of bus and for any updations on route, it is used to manage all buses. The use of GPS is not done to counter the false positives, i.e., false alarms for traffic congestion on route.

Arrival and Departure time The time of arrival of bus at certain stop and similarly for departure. Any updation in arrival time of bus will affect the departure time. The stoppage time is also a result of both these time variations.

Case 1 Ideally Stoppage Time

$$O.S.T = (O.D.T)(O.A.T) \quad (1)$$

where, stoppage time is usually fixed as in the original bus schedule and we get

$$O.D.T = (O.S.T)+(O.A.T) \quad (2)$$

the arrival time of bus is usually what varies the most. This equation remains true for times when bus arrival time to a stop is later than the departure time from the stop.

Case 2 In case the bus arrival time is later than scheduled but earlier than scheduled departure time, in that case the stoppage time is reduced and the bus departs at schedule departure time.

$$R.S.T = (O.D.T)(R.A.T) \quad (3)$$

here, departure time is given preference over stoppage time because our aim is to maintain the original bus schedule to overcome bunching.

Case 3 Stoppage time when the bus arrives later than scheduled departure time will again be fixed as in schedule stoppage time as

$$O.S.T = (R.D.T)(R.A.T) \quad (4)$$

here, real time of arrival and departure are used to maintain stoppage time so that the waiting passengers on the stop can be accommodated in the bus for ease in boarding and deboarding.

In order to keep track of the original schedule, the database readily stores the original schedule that is constantly used as a marker for providing solution. The solution is what makes the buses follow their original schedule to the nearest accuracy. Apart from storing the Arrival and Departure time, the stop info, bus info, bus real time schedule are also stored in the database with pre-processed data. For simplicity of understanding the simulation and solution for effective bus control, a simple data set is selected with a few running buses on the route with assumption based data. An instance of schedule of buses on Kaushambi - Meerut route is given below in the figure.

Table 1. INSTANCE OF KAUSHAMBI-MEERUT ROUTE
SCHEDULE OF UPSRTC

Bus	Stop	Arrival Time	Departure Time	K.M
X	MEERUT	-	4:45	0
X	RITHANI	4:59	5:01	7
X	PARTAPUR	5:09	5:11	12
X	BHUDBARAL	5:13	5:15	13
X	SUGAR MIL	5:19	5:20	15
X	GOVINDPURI	5:25	5:27	18
X	MODINAGAR	5:32	5:33	22
X	SIKRI	5:39	5:41	26
X	MURADNAGAR	5:51	5:52	32
X	DUHAI	6:02	6:04	38
X	MORTA	6:12	6:13	42
X	MOHAN NAGAR	6:28	6:30	52
X	KAUSHAMBI	6:40	-	59

The simulation works in two parts which are

- (1) BB detection
 - (2) Implementing a CS.
- (1) For detecting BB, we need to calculate the time to reach the next stop on route for the bus. In general this is checked with the schedule in present system that can tell about the delays at stop arrival of an bus. But, in real-time scenario, this delay can be updated midway at the source of the delay. This system would however require a minimum threshold value that can be considered as non bunching as a delay which is not significant enough to disturb the bus schedule. In our simulation, the threshold value is 5 min. This is constantly matched with the bus arrival time at the upcoming stop. The individual time taken between two consecutive stops would result in giving the total time of journey of the bus, which means that time taken by bus from terminal 1 to stop 1 be T_1 and time taken by bus from stop 1 to stop 2 be T_2 and so on therefore, the total T.O.J of the bus will be:

$$T.O.J = T_1 + T_2 + \dots + T_n$$

where, T_n is the time taken by to reach the last terminal from its earlier stop.

This is helpful in providing a real-time estimated time of journey to the users/passengers via the framework. This would help waiting passengers in managing their journey better and the boarded passengers a near accurate time of reaching their destination. Consider, the threshold value of time delay is reached and/or crossed. This would result in the BB event further down the route when the next bus is following its schedule. This threshold value thus plays an important role in generating CS for BB event. However, this threshold value should also not be relatively low as it can trigger for CS even in situations where the bus can itself correct its schedule in due course. It is therefore implied that a threshold value should consider all the traffic variations possible on the route as a whole and between stops as well to be assigned a value. If poorly valued, the simulation would end up giving multiple varied predictions based on the real-time data and would result in unnecessary correction implementation even if it could have been naturally handled. This is referred to as false positives or false alarms, where non significant enough time delay results in the creation and implementation of a CS.

- (2) Implementing a CS requires the BH strategy as the main solution. However, in some studies conducted earlier, SS is also used as a control strategy to eliminate BB. But, in our case SS is not a feasible solution for BB as :-

- (a) It requires informing the passengers in advance about the stop to be skipped.
- (b) In our case, it is a high frequency route where passengers board / deboard at all stops.

Since passengers activity takes place at every stop, so any stop cannot be skipped on the route even though the bus may have a huge delay from its schedule.

We are now left with the only feasible and effective solution of BH. BH requires a bus to halt at some stop on route. This halting of bus can be intelligently implemented to increase the bus journey time. Considering the cases of delay of bus, BH can provide a solution as

Case 1 Bus delayed from arrival time, but arrives earlier than departure time at stop-the bus is halted to the scheduled departure time of the bus and departed from the stop at the scheduled departure time.

Case 2 The bus arrives later than the scheduled departure from the stop- the bus is halted for the scheduled stoppage time and departs from stop. If the next bus is sufficiently closer to this bus i.e. <5 min time difference between the arrival time of both buses, the next bus is updated to arrive at the stop after 5 min of arrival of the earlier bus (delayed bus).

In both the cases, buses are held to increase the time difference between the buses that come together to provide some regularity of the bus service with respect to time. At some even spaced time intervals a bus reaches a certain stop to provide passengers with bus service. Therefore, when a BH strategy is applied the time of journey of the bus also increases and so does the time taken between consecutive stops by the bus, which will be

$$T_R = T_S + D.T + H.T$$

where, T_R and T_S are real time and scheduled time of bus taken in consecutive stops and D.T is delay time of bus from schedule and H.T is holding time assigned to the bus.

In the simulation, the CS is provided using the regression analysis for linear functions. The regression brings the linear relationship between the bus and the stop at a given time. This linear model is then used to make prediction for the future event of the bus reaching a stop at some other time based on the earlier data given to the framework.

INPUT-

1. The bus stop at which BB occurs - busstop
2. The delay in minutes of the bus to the stop - min
3. Learning rate is given to time

OUTPUT- Predicted CS:

```
def update_fun (busstop , min):
    if min = 0
        #bus is at stop

    while True:
        #update next stop

        if next == 'NA':
            break
        else:
            #no update required

        for s in stops:
            print (busstop)
            #gives values of real time

    if real.arrival != datetime.time(0,0):
        #update arrival time

        if real.departure != datetime.time(0,0):
            #update departure time

    if real.busstop.bus.next == 'NA':
        #bus reached last stop

        pass
    else:
        #update details of rest journey

    h = real.arrival.hour -
        next_real.arrival.hour
    m = real.arrival.minute -
        next_real.arrival.minute
    diff = h*60 + m
    if diff > 0:
        #arrival time difference between
        two consecutive buses

    if diff < 5 and real.departure !=
        datetime.time(0,0):
        #update time of the held bus

    else:
        update_fun (next_busstop ,
            (10+ diff))
```

#recursive function to update all values

return

This algorithm is a self recursive function that is used to update the schedule of the buses when BB event takes place. It gives the CS to the BB event keeping track of the bus relative data of arrival and departure and scheduled arrival and departure at stops. The BH action is applied to the bus sufficiently closer to the earlier bus with a time marker of 5 min. This BH would result in the updation of the schedule of the held bus which is also performed by the recursion algorithm.

This generates an updated schedule of the buses when BB event happens. This updated schedule is then graphically presented with use of linear regression. For sufficiently large trained data, gradient descent can be used to make more accurate predictions, but with our data set and assumption based data linear regression alone also performs well. Gradient descent would straighten the slope to result in perfectly linear prediction, but linear regression alone would give streamlined results, which is adequate for our simulation.

For example- the schedule of buses in our simulation is illustrated in Table 2. At some time instance, say 11:20 am, the real-time

Table 2. SCHEDULE OF BUS IN SIMULATION

Bus	Stop	Arrival Time	Departure Time	Status	Update
A	T1	midnight	10am	Not Started from terminal	
A	S1	10:15am	10:20am	Not Started from terminal	
A	S2	10:45am	10:50am	Not Started from terminal	
A	S3	11:15am	11:20am	Not Started from terminal	
A	S4	11:35am	11:40am	Not Started from terminal	
A	T2	11:55am	midnight	Not Started from terminal	
B	T1	midnight	10:30am	Not Started from terminal	
B	S1	10:45am	10:50am	Not Started from terminal	
B	S2	11:15am	11:20am	Not Started from terminal	
B	S3	11:45am	11:50am	Not Started from terminal	
B	S4	12:05pm	12:10pm	Not Started from terminal	
B	T2	12:25pm	midnight	Not Started from terminal	
C	T1	midnight	11am	Not Started from terminal	
C	S1	11:15am	11:20am	Not Started from terminal	
C	S2	11:45am	11:50am	Not Started from terminal	
C	S3	12:15pm	12:20pm	Not Started from terminal	
C	S4	12:35pm	12:40pm	Not Started from terminal	
C	T2	12:55pm	midnight	Not Started from terminal	

schedule of fleet is illustrated in Table 3. At this time instance a

Table 3. SCHEDULE AT 11:20am WITH NO DELAY IN BUS A

Bus	Stop	Arrival Time	Departure Time	Status	Update
A	T1	midnight	10am	Departure From T1	
A	S1	10:15am	10:20am	Departure From S1	
A	S2	10:45am	10:50am	Departure From S2	
A	S3	11:15am	11:20am	Departure From S3	
A	S4	11:35am	11:40am	15 MINs TO STOP S4	update
A	T2	11:55am	midnight	35 MINs TO STOP T2	update
B	T1	midnight	10:30am	Departure From T1	
B	S1	10:45am	10:50am	Departure From S1	
B	S2	11:15am	11:20am	Departure From S2	
B	S3	11:45am	11:50am	25 MINs TO STOP S3	update
B	S4	12:05pm	12:10pm	45 MINs TO STOP S4	update
B	T2	12:25pm	midnight	65 MINs TO STOP T2	update
C	T1	midnight	11am	Departure From T1	
C	S1	11:15am	11:20am	Departure From S1	
C	S2	11:45am	11:50am	25 MINs TO STOP S2	update
C	S3	12:15pm	12:20pm	55 MINs TO STOP S3	update
C	S4	12:35pm	12:40pm	75 MINs TO STOP S4	update
C	T2	12:55pm	midnight	95 MINs TO STOP T2	update

time delay 20 min is given to bus A for reaching its next stop which creates a BB event, the updated schedule is shown in Table 4. After

Table 4. SCHEDULE AT 11:20am WITH 20 MIN DELAY IN BUS A

Bus	Stop	Arrival Time	Departure Time	Status	Update
A	T1	midnight	10am	Departure From T1	
A	S1	10:15am	10:20am	Departure From S1	
A	S2	10:45am	10:50am	Departure From S2	
A	S3	11:15am	11:20am	Departure From S3	
A	S4	11:35am	11:40am	55 MINs TO STOP S4	update
A	T2	11:55am	midnight	75 MINs TO STOP T2	update
B	T1	midnight	10:30am	Departure From T1	
B	S1	10:45am	10:50am	Departure From S1	
B	S2	11:15am	11:20am	Departure From S2	
B	S3	11:45am	11:50am	25 MINs TO STOP S3	update
B	S4	12:05pm	12:10pm	45 MINs TO STOP S4	update
B	T2	12:25pm	midnight	65 MINs TO STOP T2	update
C	T1	midnight	11am	Departure From T1	
C	S1	11:15am	11:20am	Departure From S1	
C	S2	11:45am	11:50am	25 MINs TO STOP S2	update
C	S3	12:15pm	12:20pm	55 MINs TO STOP S3	update
C	S4	12:35pm	12:40pm	75 MINs TO STOP S4	update
C	T2	12:55pm	midnight	95 MINs TO STOP T2	update

20 min i.e. at 11:40 am the schedule of the bus with delay is illustrated in Table 5. At this time, bus A is behind schedule and the next

Table 5. SCHEDULE OF BUSES AT 11:40am

Bus	Stop	Arrival Time	Departure Time	Status	Update
A	T1	midnight	10am	Departure From T1	
A	S1	10:15am	10:20am	Departure From S1	
A	S2	10:45am	10:50am	Departure From S2	
A	S3	11:15am	11:20am	Departure From S3	
A	S4	11:55am	12:00am	35 MINs TO STOP S4	update
A	T2	12:15am	midnight	55 MINs TO STOP T2	update
B	T1	midnight	10:30am	Departure From T1	
B	S1	10:45am	10:50am	Departure From S1	
B	S2	11:15am	11:20am	Departure From S2	
B	S3	11:45am	11:50am	5 MINs TO STOP S3	update
B	S4	12:05pm	12:10pm	25 MINs TO STOP S4	update
B	T2	12:25pm	midnight	45 MINs TO STOP T2	update
C	T1	midnight	11am	Departure From T1	
C	S1	11:15am	11:20am	Departure From S1	
C	S2	11:45am	11:50am	5 MINs TO STOP S2	update
C	S3	12:15pm	12:20pm	35 MINs TO STOP S3	update
C	S4	12:35pm	12:40pm	55 MINs TO STOP S4	update
C	T2	12:55pm	midnight	75 MINs TO STOP T2	update

buses B and C are running correct on schedule. This will generate a predictive graph at time instance of 11:40.

The blue line in graphs represent the relationship between bus and stop at the scheduled time instance and the green line represents the relationship in real-time. In fig.4, bus A is corrected to solve BB and bus B is allowed to run following its schedule. Had there been a time difference of <5 min between bus A and bus B reaching the same stop, bus B would have been held for upto 5 min at the stop and further update the running schedule of bus B. Since, the prediction appears to be streamlined thus a minimum deviation from the original schedule of the buses.

The graph in fig.5 represents the relationship between buses and stops at multiple time intervals. It gives us the opportunity to calculate the effectiveness of our simulation model. The aim of our simulation was to streamline the deviated buses to their schedule and to achieve this by minimal control actions implementations. This graph reveals that almost all the buses are streamlined barring

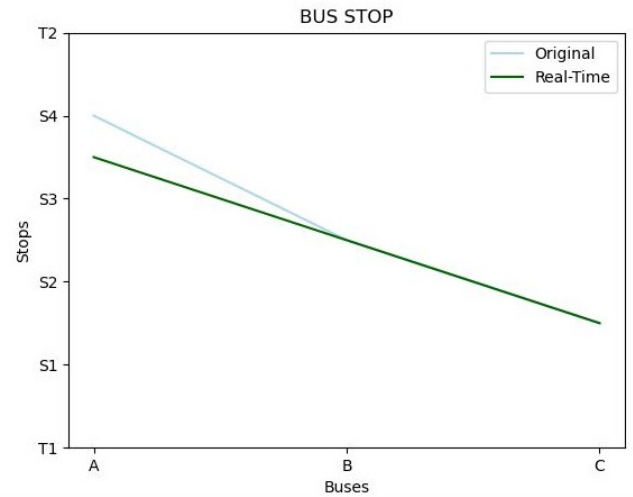


Fig. 4. GRAPH FOR REAL-TIME BUS RELATIONSHIP WITH THEIR LOCATION AFTER DELAY AT 11:40am

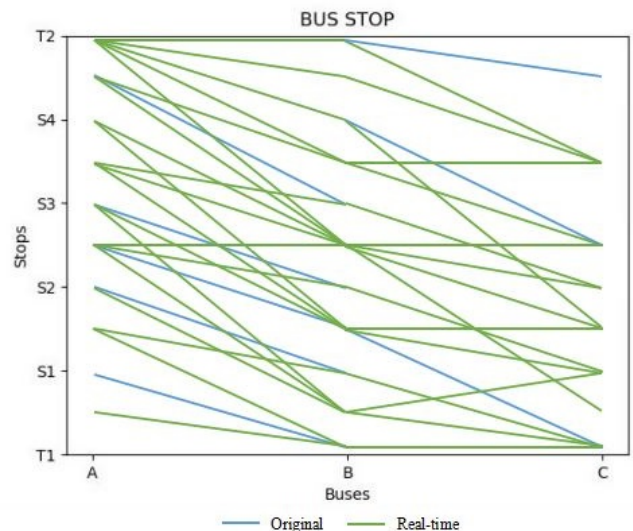


Fig. 5. CONSOLIDATED GRAPH FOR ALL ITERATIONS

aside the event of traffic jam, where our simulation updates a scheduled that resonates the original schedule but with time delay due to traffic jam as illustrated in fig.6. Even in this situation, an accurate prediction can be made about the fleet operation. The overall effectiveness of our simulation is more than two-third i.e. 67 % which is good enough for a high frequency Kaushambi Meerut route.

9. RESULT

The aim of this proposed simulation was to streamline the deviated buses to their schedule and to achieve this by minimal control actions implementations. The graph in fig.5 represents the relationship between buses and stops at multiple time intervals. The evaluation

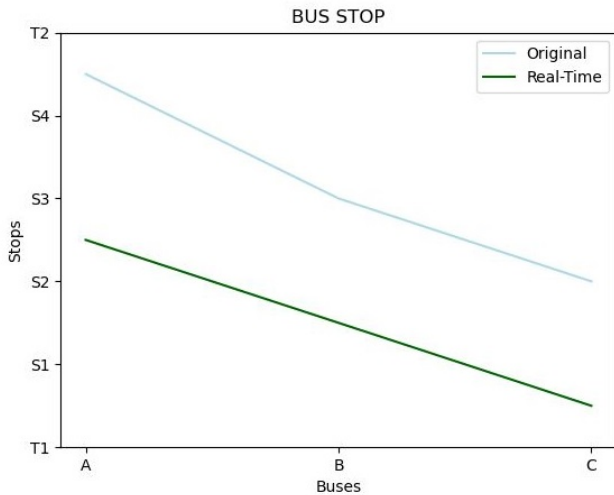


Fig. 6. GRAPH FOR TRAFFIC JAM SITUATION

is performed using the metric as discussed in [1], which is

$$BBreduction = \frac{(BBtripswithoutCS)(BBtripswithCS)}{(BBtripswithoutCS)} \quad (5)$$

It gives the opportunity to calculate the effectiveness of the simulation model. The graph in fig.5 reveals that almost all the buses are streamlined barring aside the event of traffic jam, where the simulation updates a scheduled that resonates the original schedule but with time delay due to traffic jam.

10. CONCLUSION

A framework that helps eliminate BB occurrence is proposed and a simulation for the same is also developed. It is first of its kind for PT system in India where the general attitude of public using PT towards BB is considered as inevitable or a usual feature related to bus travel. The framework makes use of real-time and stored data (historical data) to predict the bunching event. This prediction is then used to generate a solution which gets implemented and correct the BB 67 % times.

The simulation takes use of the data from Kaushambi-Meerut route and assumption based data as test cases for various traffic scenarios that may possibly occur on route. The result shows that the framework is effective in solving BB in real-time and therefore, could be deployed for high frequency bus routes.

11. FUTURE ENHANCEMENT

The future works include carrying out experiments to develop an integrated system that works effectively for both low and high frequency bus routes. There is a scope of re-planning a threshold value considering the route for a system operating over multiple routes. A well planned threshold value would result in a more accurate prediction of bunching, but it should also be sufficiently high to eliminate false positives. The traffic influence on the system can be analysed and an effective solution can be generated to avoid those influences and provide better services during peak traffic hours.

12. REFERENCES

- [1] Matias, L. M., Cats, O., Gama, J., Moreira, J. M., and Sousa, J. F. (2016). An online learning approach to eliminate bus bunching in real-time.
- [2] Ercan, T., Onat, N. C., Tatari, O., and Mathias, J.-D. (2017). Public transportation adoption requires a paradigm shift in urban development structure. *Journal of Cleaner Production*, 142, 17891799.
- [3] Santos, D., Kokkinogenis, Z., de Sousa, J. F., Perrotta, D., and Rossetti, R. J. (2016). Towards the integration of electric buses in conventional bus fleets. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 8893.
- [4] Wakker, K. F. (2015). *Fundamentals of astrodynamics*.
- [5] Xuan, Y., Argote, J., and Daganzo, C. F. (2011). Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. *Transportation Research Part B: Methodological*, 45(10), 18311845.
- [6] Dessouky, M., Hall, R., Zhang, L., and Singh, A. (2003). Real-time control of buses for schedule coordination at a terminal. *Transportation Research Part A: Policy and Practice*, 37(2), 145164.
- [7] Daganzo, C. F. and Pilachowski, J. (2011). Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B: Methodological*, 45(1), 267277.
- [8] Daganzo, C. F. (2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological*, 43(10), 913921.
- [9] Boyd, C. W. (1983). Notes on the theoretical dynamics of intermittent public passenger transportation systems. *Transportation Research Part A: General*, 17(5), 347354.
- [10] Abkowitz, M., Knier, F., Yuh, H., Weagley, R., and Stolka, M. (1987). Electronic transport in amorphous silicon backbone polymers. *Solid state communications*, 62(8), 547550.
- [11] Hassan, S. M., Moreira-Matias, L., Khiari, J., and Cats, O. (2016). Feature selection issues in long-term travel time prediction. *International Symposium on Intelligent Data Analysis*, Springer, 98109.
- [12] Khiari, J., Moreira-Matias, L., Cerqueira, V., and Cats, O. (2016). Automated setting of bus schedule coverage using unsupervised machine learning. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 552564.
- [13] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- [14] Moreira-Matias, L., Ferreira, C., Gama, J., Mendes-Moreira, J., and de Sousa, J. F. (2012). Bus bunching detection by mining sequences of headway deviations. *Industrial Conference on Data Mining*, Springer, 7791.
- [15] Moreira-Matias, L., Gama, J., Mendes-Moreira, J., and de Sousa, J. F. (2014). An incremental probabilistic model to predict bus bunching in real-time. *International Symposium on Intelligent Data Analysis*, Springer, 227238.
- [16] Neapolitan, R. E. (2012). *Probabilistic reasoning in expert systems: theory and algorithms*. CreateSpace Independent Publishing Platform.

13. APPENDIX

13.1 LINEAR PREDICTIONS USING REGRESSION ANALYSIS

It is a statistical process that estimates set of relationships between dependent and independent variables in a function. This estimation is used for making prediction and forecasting for ML applications. The family of linear equations are modelled as Linear Regression. A straight line is represented as:

$$y = mx + c \quad (6)$$

where, y and x are the y and x axis of the graph, m is the slope of the line and c is the point of intercept of line at y - axis. The slope makes it possible to correctly predict the points in positive x - y plane falling on the line.

13.2 GRADIENT DESCENT

It is an iterative optimization algorithm that finds local minima of the function, linear function in our case. It could be best understood as when going uphill/downhill we need to walk on the steepest ascent/descent at each step until we reach the top/foot of the hill. Similarly, out of all possible graphical predictions at each instance gradient descent optimizes a slope that results in the CS.

13.3 TRAFFIC JAM

The condition of traffic jams affect the fleet in near similar fashion i.e. when the buses that get stuck in jam experience near similar time delay until the jam gets cleared. This is because of a high frequency route where thousands of vehicles use the route on daily basis. The challenge occurs when the jam gets cleared and buses seem to be grouped. The simulation results suggest that it is capable of handling these situations as well.