ANALYSIS AND VISUALISATION OF BUS BUNCHING WITH BMTC DATA

Nithya Nadig Shikarpur

BITS Pilani K.K. Birla Goa Campus
Department of Computer Science and Engineering

Guided by

Dr. Tarun Rambha

Assistant Professor
Department of Civil Engineering
Indian Institute of Science, Bangalore

Preface

This report is the summary of the work done from May 2019 - July 2019 during an internship at Center for infrastructure, Sustainable Transportation and Urban Planning (CiSTUP), Indian Institute of Science (IISC), Bangalore under the guidance of Dr. Tarun Rambha.There is a file with the code used during this project that can be used to view the results.

Acknowledgement

I would like to thank **Dr. Tarun Rambha** for his consistent support and invaluable insights during the course of the project. Despite having a busy schedule and not being physically present at all times he was always there to guide me during the internship period. I would also like to thank **Dr. Ipsita Banerjee** for her insights on interpretation of the BMTC bus data. I would also like to thank **Dr. Abdul Pinjari**, chairman CiSTUP for this opportunity. Most importantly I would like to express my gratitude towards all the members of CiSTUP for their constant support during the period of my internship.

Yours sincerely,
Nithya Nadig Shikarpur

Index

ABSTRACT

Bus bunching is a problem that plagues the field of transportation on a daily basis. Once bunched, it is often impossible for buses to break out of the bunching during their trips. In busy metropolitan cities like Bangalore, where a large proportion of people depend on buses for transportation, it is important that buses maintain their schedules. In addition, the heavy traffic congestion doesn't help in reducing bus bunching. This project focuses on capturing and analysing bus bunching instances from the past. This data can then be used to devise new algorithms to reduce bus bunching in the city.

Chapter 1

Introduction

1.1 What Is Bus Bunching?

In public transport, bus bunching refers to the situation when 2 or more buses having evenly spaced schedules end up at the same place at the same time. For this, 2 or more buses have to violate their schedule.
Causes for this could include traffic congestion, extra time taken by passengers to board or deborad the bus or temporary breakdowns of buses while on route.
Effects of bus bunching can get magnified very easily. For instance, when two buses get bunched, the first bus usually gets overcrowded and the second goes near-empty. This becomes a vicious cycle that cannot be broken. This ultimately leads to inconsistent waiting times, not allowing passengers to fully rely on the bus system for their transport.

Figure 1.1: Bus Bunching

1.2 BMTC

Bangalore Metropolitan Transport Corporation (BMTC) is a government agency that controls public transport bus service in Bengaluru. The BMTC hosts ordinary services in addition to special types of services including Suvarna, BIG 10, Vajra etc.

Figure 1.2: BIG 10 bus

BMTC operates across 3 major bus stations named - Kempegowda Bus Station, Shivajinagar and KR Market. It has 34 other bus stations in addition to several intermediary bus stops. BMTC has a ridership of 44.37 lakh (according to the 2017-18 annual report of BMTC).

Given the high number of people dependant on the BMTC for transportation and heavy traffic congestions in the city which can be very conducive to bunching, it is important to find ways to tackle the ineffectiveness brought about due to bus bunching.

1.3 Languages And Libraries Used

1. MySQL (MariaDB) - used to access data from the database
2. Python
   (a) Pandas - data processing library
   (b) modin - used instead of pandas in some instances; its lightweight nature allows faster computations
3. Javascript - used to visualise the data

    (a) Chartjs - library used to plot the real time data on a chart in the simulation

4. HTML - used with Javascript for the simulation

Chapter 2

Data Preprocessing

2.1 Description of Data Used

Following are a list of columns used from the given datasets along with a short description.

1. GPS Data
   (a) ID - unique id for each GPS ping
   (b) DEVICE_ID - unique id for each GPS device installed in a bus
   (c) NO_SATELLITE_IN_VIEW - number of satellites in view for each GPS ping
   (d) LAT - latitude of the ping
   (e) LONGITUDE - longitude of the ping
   (f) IST_DATE - date and time of the ping
2. Ticketing Data
   (a) ticket_id - unique id for each ticket
   (b) schedule_no - of the form <bus number(as displayed on the bus)/trip number of route>
   (c) trip_no - The trip number is 'x' when the given schedule is plying xth number of time in its desired path
   (d) route_no - General format is <Bus_Number><direction>
   (e) ticket_date - date when ticket was issued
   (f) ticket_time - time when ticket was issued
   (g) px_count - number of passengers for whom the tickets were bought

(h) total_ticket_amount - Fare collected for the ticket

3. Vehicles Data

    (a) duty_dt - date of mapping
    (b) vehicle_no - vehicle number
    (c) device_id - unique id of the GPS device in the bus

2.2 Choosing Data To Work With

GPS and Ticketing data were selected from January 1st, 2018. The Vehicles data was from April 1st, 2018 (we have assumed that the vehicle number, GPS device id mapping haven't changed from January 1st, 2018 to April 1st, 2018)
To make things simpler, I have worked only with buses running on the route 248 (Krishna Rajendra Market to Jalahalli Cross). This route was selected as it had the most number of trips on January 1st, 2018 and thus had the highest chance for bus bunching. However this can be extended for several more routes.

2.3 Data Cleaning And Preprocessing

2.3.1 GPS Data

GPS pings were recorded by each GPS device every 10 seconds approximately. However, there were several errors in the data. These errors were numbered as follows:

1. 0 - no error
2. 1 - ping is out of Bangalore
3. 2 - less than 3 satellites in view during the ping
4. 3 - speed between two pings from a particular device is greater than 90 kmph
5. 4 - Readings that have covered non zero distance in zero time
6. 5 - Only one or less reading from the device
7. 6 - GPS ping is more than 20 km away from any of the bus stops on the route (not implemented very efficiently)
8. 7 - latitude of ping is not between -90 and 90 degrees

In the given data set, out of 157985 records, 350 pings weren't in Bangalore and 39 pings covered non zero distance in zero time.

2.3.2 Ticketing Data

This data had all data related to tickets purchased by passengers going in a particular bus. Steps followed to clean data:

1. Data had to be extracted for each device id travelling along the route
2. GPS device ids (GPS data) were mapped to vehicle numbers (ticketing data) using the vehicles document
3. Care had to be taken to consider only vehicles that travel along a particular route during the particular day (this was solely done to keep matters simple)

2.3.3 Binning

Data binning or bucketing refers to the data preprocessing technique in which several data instances are grouped with respect to a certain feature in order to summarise it and make it easier to analyse. This method also removes noisy data, making the data more accurate.

The ticketing and GPS data were binned in intervals of 10 seconds each. This allowed us to compare them easily and use the values simultaneously. The GPS and Ticketing data were binned separately for each device id of all the buses travelling on the given route.

Default values for the following columns were set as stated:

1. Latitude, Longitude - the Latitude and Longitude of the depot of the route based on the direction of the first ticket purchased
2. Other values were set to zero

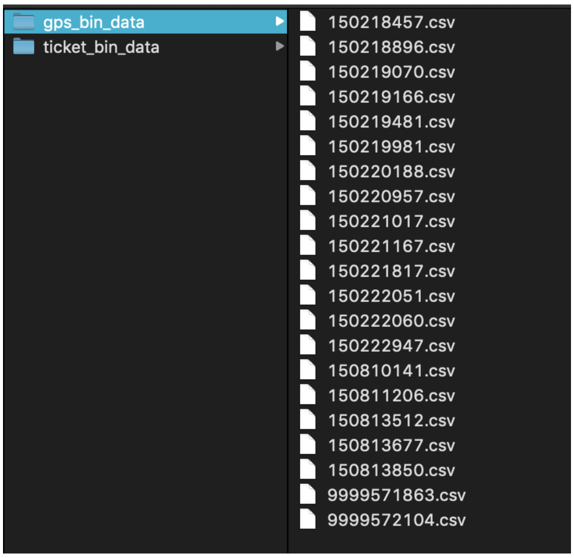


Figure 2.1: File Structure Of Binned Data

Chapter 3

Simulation

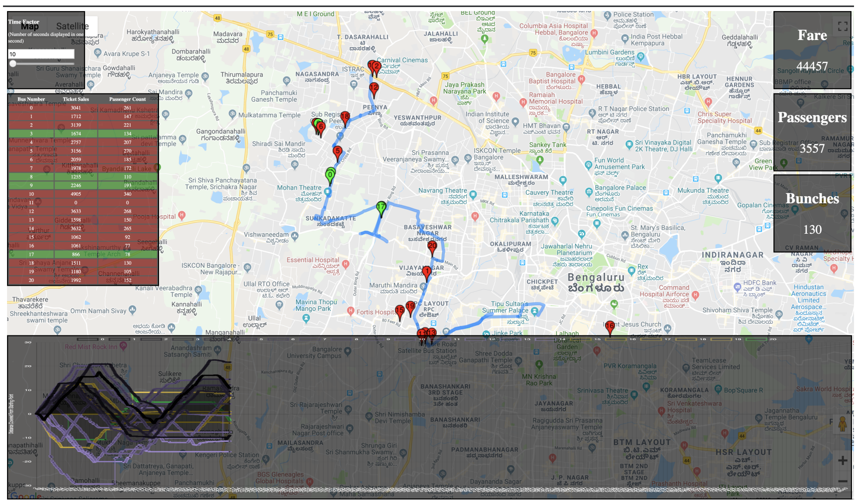


Figure 3.1: Simulation

3.1 Markers And Route

Each marker represents a bus and is numbered in serial order. Buses bunched are green in colour and the other ones are red in colour.
Route of the buses was pulled from the BMTC bus website.

3.2 Route Variables

1. Fare - The total fare collected by all the buses on the route

2. Passengers - The total number of passengers that have boarded any of the buses on the route

3. Bunches - Number of bunching instances
   Here, bunching instance is defined as an instant when 2 buses are travelling in the same direction but are within a certain bunching radius (in this case, 100m). Once a pair of buses undergo bunching in a particular trip, they are not counted again thus preventing double counting.

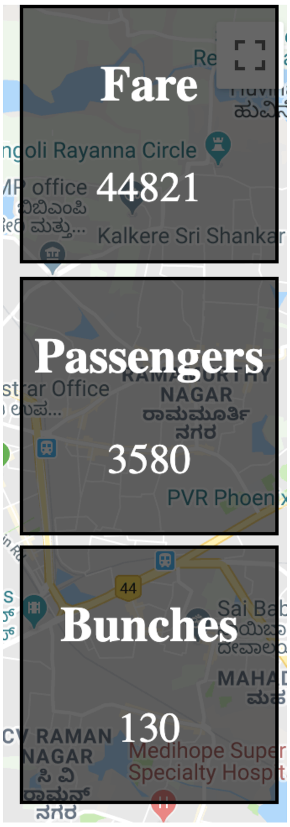


Figure 3.2: Route Variables

3.3 Time Factor

Time factor is defined as the number of seconds from the live data displayed per second in the simulation. This value can range from 10 to 1000.

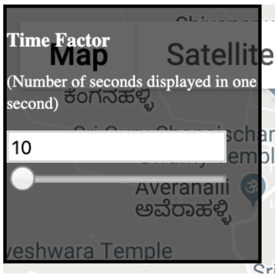
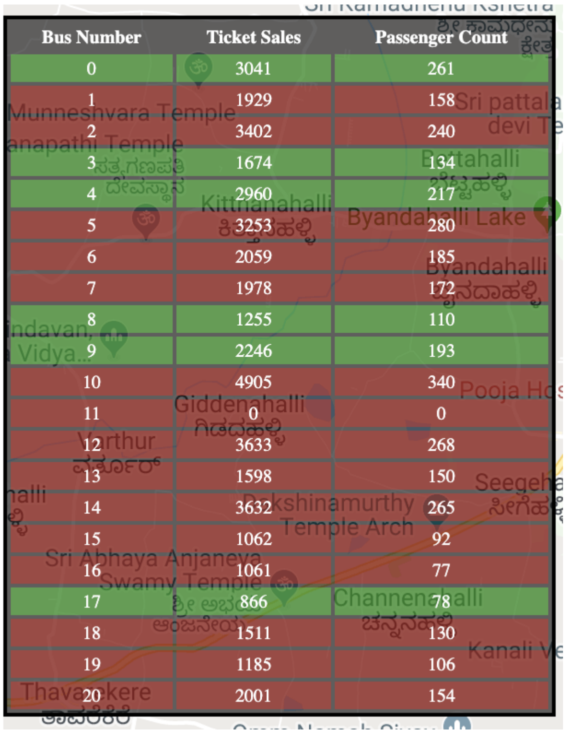


Figure 3.3: Time Factor

3.4 Bus Data

It has the total fare collected and total number of passengers that have boarded the bus. The colour of the row changes based on whether it is bunched or not.



Bus Number	Ticket Sales	Passenger Count
0	3041	261
1	1929	158
2	3402	240
3	1674	134
4	2960	217
5	3253	280
6	2059	185
7	1978	172
8	1255	110
9	2246	193
10	4905	340
11	0	0
12	3633	268
13	1598	150
14	3632	265
15	1062	92
16	1061	77
17	866	78
18	1511	130
19	1185	106
20	2001	154

Figure 3.4: Bus Data

3.5 Space Time Graph

This a graph with the time passed on the X-axis and the distance travelled by the bus on the Y-axis (Distance is positive for the UP direction and negative for the DOWN direction). Each line plotted represents a different bus on the route.
When two lines are parallel, it can be interpreted that they are consistently travelling with an even space between each other. However, when the lines intersect, we can see that they have bunched.

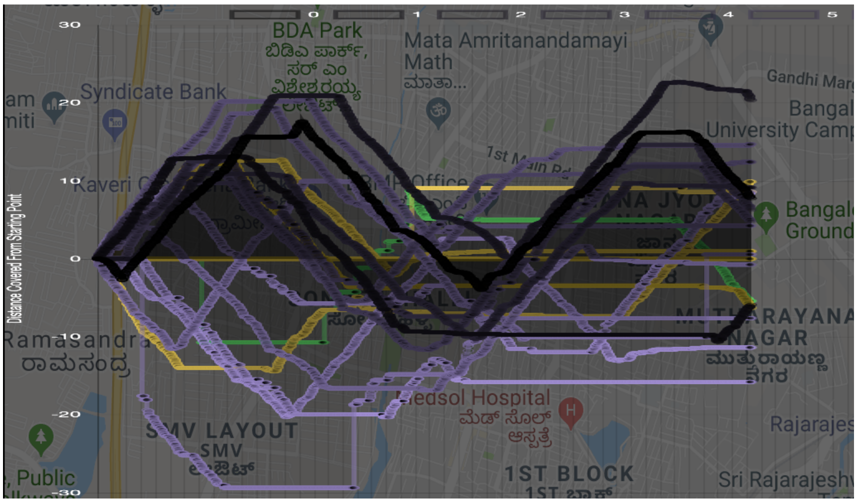


Figure 3.5: Space Time Plot

Chapter 4

Possible Sources Of Error

1. Direction of each bus is pulled from the ticketing data, however there is no indication of when that particular trip ends and the direction is reversed. Hence even if the bus is stationary after a trip it is assumed to be moving in the same direction till direction changes (this could be a problem while defining bunching). This same problem holds for the trip number of the bus as well.

2. GPS device ids and vehicle numbers have been mapped according to newer data. We have assumed that this mapping hasn't changed between January 2018 to April 2018.

Chapter 5

Conclusion

The analysis and visualisation of the BMTC bus data aids us to identify instances of bus bunching. This simulation can be extended to other routes as well to better our knowledge of where exactly bus bunching is occuring. Using this knowledge, algorithms can be designed to reduce instances of bus bunching. Ultimately, this will allow buses to be more reliable and efficient.