Travel time prediction for buses in Rio de Janeiro

Incident recognition and travel time change caused by incidents

Giel Haarsman (s0118907)

Bachelor thesis Civil Engineering

Internship at IFluxo

Supervised by: Warner Vonk (IFluxo) & Tom Thomas (University of Twente)

Enschede and Rio de Janeiro, June 2009

Porta-á-Porta com transporte público

IFLUXO

University of Twente
Enschede - The Netherlands

Table of Contents

Summary

In Rio de Janeiro more people are choosing for individual transport and this results in a decrease in bus passengers and other negative consequences. To let the passengers choose for the bus, IFluxo is developing a service which gives the potential bus passenger information via text message about the departure time of a bus of interest. Three important aspect in this project are:

1) Stakeholder demands and wishes
2) Software for the text message service
3) Estimating the time of arrival of the buses

In this thesis the first and third aspect are viewed. Demands for the public transport operator and the potential passenger have to be matched and the input of one influence the output of the other en the other way around.

The estimation of the arrival time of buses in Rio de Janeiro consists of two parts, the continuous prediction and the incidental prediction. This thesis will start with the analysis of the incidental prediction.

The objective of this research is to develop a table for all possible incidents during bus rides, whereby the table is quantified in order to determine the nature of the incident and as a result of that estimate a reliable time of arrival.

In this thesis the events on a bus ride are viewed, as well visual as with a GPS Travel Recorder. Incidents and irregularities are tried to be recognized and attempted to be distinguished in the data and the influence on the travel time is examined.

Result of the literature research where as follows. One can say that there is done a lot of research on the subject of predicting arrival times, but every research uses other types of data collecting and most researches have their focus on the continuous part of travel time prediction and not on the incidental part. In this thesis, because of the lack of a big GPS database, the focus will be on the change in the arrival time caused by incidents, but as well caused by irregularities. Herby, incidents are non-recurring and irregularities are recurring traffic events.

To recognize the irregularities and incidents in the GPS data and to give the delay caused by them first, all travel times of the different bus lines (2 lines, 2 directions) had to be matched to the same Points of Interest (PoI) which are in this case bus stops. This is called map-matching and this is done in combination with the nearest neighbourhood method. This method chooses the nearest point to the PoI and allocates this point to the PoI. Finally, this will result in average sections times between the points of interest.

Results of the analysis of recognizing the irregularities and incidents in the GPS data, separated from the visual data are negative. It is impossible to recognize the irregularities or incidents in the GPS data, because first, one

incident / irregularity has the same characteristics as another incident / irregularity. Second, the incident / irregularity has within the incident / irregularity different characteristics.

Because incidents and irregularities can not be distinguished in the GPS data, the delays are predicted using both, GPS data and visual data. Only (1) 'a lot of passengers getting in or out the bus', (2) 'road maintenance' and (3) 'congestion' occurred. Incident (1) always occurs in the same section, incident time (time difference with the average section time) is small and fluctuations are minimal. This incident can be included in the continuous prediction. Incident (2) sometimes occurs in the same section, incident time is short and fluctuations are minimal. This can be included in the continuous prediction. However, there are times, appearing in the data, that road maintenance not occur in the same sections and incident times are big, varying from one to two minutes. Incident (3) always occurs in the same section, but incident time is big and is fluctuating from one minute to 7,5 minutes.

Because only three kinds of incident occur in the data an interview is held. This interview gives more insights in the procedures, number of occurrence and incident times of the other incidents. All irregularities can be taken within the continuous prediction. When more data is available the incidents can maybe distinguished and the delay can be predicted, but in this stage it is not possible.

All these results are combined in a table which is displayed at page 37.

1 Introduction

In the western countries people are already choosing for individual transport and this results in a decrease of bus passengers. This phenomenon is starting to occur in developing countries, such as Brazil. To keep the passengers in the bus, the competiveness of the bus, in comparison with the car, needs to be increased.

This is possible in different ways, but Public Transport (PT) operators are thinking differently than the potential passenger. In this way measures are not as affective as they can be.

This match between PT operator and the potential client gives insights in the important role of information provision and the influence of this on the demands and wishes of potential passengers. IFluxo is developing a service that provides the passenger of personal Real Time Travel Information by SMS or via the internet.

Important aspects of this service are software, stakeholders and travel time prediction. In this thesis, as mentioned, the match between stakeholders is discussed, but the main subject is travel time prediction. Hereby reliability and accuracy of the predictions are very important. Incidents have a very big impact on the travel time and thus on the reliability and accuracy.

In this thesis the events on a bus ride are viewed. Incidents and irregularities are tried to be recognized and attempted to be distinguished in the data and the influence on the travel time is examined.

1.1 Objective

The objective of this research is to identify all possible irregularities and incidents during bus rides in Rio de Janeiro and the number of prevalence. This information will be put together in one table, together with the delay belonging to the irregularities and incidents.

1.2 Research questions

To achieve this objective three questions will answered in hierarchic order presented below.

1. How can the match between the interests of the PT operator and the potential passenger be made?
2. What do the terms irregularity and incident refer to and which kind of irregularity and incident will play a part in predicting travel times?
3. What kinds of data sources are necessary to make objective statements about the nature of the irregularities and incidents?
4. How can the collected data be used to estimate indicators with which different irregularity and incident types can be distinguished and predict delays for a given irregularity and incident type?

1.3 Approach

1.3.1 Literature research

Previous researches, terminology, stakeholders and the situation in Rio de Janeiro will be investigated. Findings will be mapped and improved.

1.3.2 Case study

By travelling in a bus, during the execution of this thesis, data is collected in order to give conclusions based on real data. This is done by collecting global positioning and visuals data. The global positioning data is collected with a GPS Travel Recorder which gives the position in altitude, latitude and longitude, the time in seconds and the speed in kilometres per hour. The visual data is collected by sitting in a bus and writing everything down what was happening in the bus, for example: number of passengers getting in or out, time at a bus stop, time at a traffic light, time standing still at road maintenance and congestion time.
Besides this, an interview has to give more clarity about not occurring incidents. The interview is held with Márcio Coelho Barbosa, the director of the federation of bus companies. The first part of the interview was about the match between the public transport operator and the potential passenger. The second part was about the number of times an incident occurs, the procedure followed when an incident occurs and the delay caused by that incident. In this way a global estimated can be made of the delay caused by the incident.

1.4 Outline

To reach the objective of this thesis, first of all a background analysis is done. Previous researches are discussed and the difference between these researches and this research is pointed out. Chapter 3 gives a overview of the two most important stakeholders in this project and the way these stakeholders are connected to each other. In this way one knows what is important for both the potential passenger and the public transport operator. To get a better understanding of the way of thinking and acting in Rio de Janeiro a case description is given in chapter 4. This results in a good understanding of the differences between The Netherlands and Brazil and the bus line the case study is preformed.
In the chapters 5,6,7 and 8 the research questions are answered. First of all, the terms irregularity and incident are clarified, so that one knows what incident and irregularity refer to. In chapter 6 the technical aspects are discussed. Some difficult terms are explained, and analysing preparations, such as map-matching and the nearest neighbourhood theory, are explained. Given these aspects of the analysis, data sources are discussed and divided in 'available' and 'non-available and not useful data'. This will be noted in chapter 7. With the available data sources: the GPS Travel Recorder, the visual data and the interview delays per irregularity or incident are given in chapter 8. At last, conclusions and recommendations are given in chapter 9, presenting the total table with irregularities and incidents with their belonging characteristics.

2 Background

Worldwide Intelligent Transportation Systems (ITS) are gaining popularity, hence also on the area of public transportation, called Advanced Public Transportation Systems (APTS). On of the subheadings within APTS is Real Time Travel Information (RTTI).

In this chapter first some previous research on RTTI will be discussed. Within this research a few important parameters emerge. In the second paragraph these parameters, the technical aspects, will be explained together with other important parameters for the project. Not only technical aspects are important, the business aspects are a part of the project as well. That's why in the last paragraph a small stakeholder analysis is done and stakeholders will be mentioned with their demands and wishes.

2.1 Previous research

Real Time Travel Information is an information system which gives a passenger or potential passenger up-to-date travel information. To give accurate and credible information, the author of 0 *Travel Time Prediction in Rhode Island* says: "the system should be capable of producing accurate and reliable estimates and predictions of the travel times under all possible traffic conditions and circumstances." The author states that these estimations and predictions can be achieved by the use of GPS technologies. Ramakrishna, Ramakrishna, Lakshmanan & Sivanandan (2006) share this opinion and point out the same important parameter, travel time data. They say that the traffic data collection can be done more efficiently with the help of the GPS technologies. This will result in a more reliable and accurate system.

A lot of previous research is done on the subject of RTTI but the way of investigating and the subtopics of the research where different. Hoogendoorn & Van Lint did research on the prediction of travel time and the travel time reliability. They first described tree methods of travel time prediction.

1. Naive methods which use the real (historical) travel times or the instantaneous travel times
2. Model based methods which use a traffic simulation model, simple or difficult.
3. Data driven methods which use statistical techniques and correlate actual travel time to available data.

The advantage of the model based methods is the network width prediction and its possibility to predict under recurrent and non-recurrent conditions (scenarios and incidents). This method, however is very labour and maintenance intensive. The data driven methods do not have this disadvantage and are fast, accurate and robust, but are instead situation specific and can only be used in recurrent situations. Besides this, data driven methods need enough historical data (Hoogendoorn & Van Lints, 2008)

The author of *Travel Time Prediction in Rhode Island* also names some of the advantages of the data driven methods and did research on two kinds of networks, namely Neural Networks and a Fuzzy-Neutral Network. The data in this project is collected by loop detectors, license plate matching and road watching. Therefore no GPS data was involved.

In the research of Ramakrishna et al. (2006), GPS data were used to develop a model to predict arrival times of buses. Again the Neutral Network model and a Multiple Linear Regression model are discussed and Ramakrishna et al. (2006) also name the advantages of a data driven method.

A striking issue in these researches is the lack of considering incidents in the prediction. The author of *Travel Time Prediction in Rhode Island* says: "The unexpected information in the traffic patterns is treated as noise" (p.3). Besides this, Hoogendoorn & Van Lint say the influence of incidents and weather conditions fall outside the scope of time travel prediction, because they can not be predicted. On the other hand they suggest that the model based method and the data driven method can be combined, because that way all advantages are included and incidents can probably be predicted.

Because the above mentioned researches do not include the recognition of incidents, two researches are mentioned below.

In the research of Sheu (2004) about the characteristics of incidents and their detection on freeways, data is collected from point detectors and it contains count and occupancy. These data provide different characteristics of incidents, such as: time-varying lane-changing fractions and queue lengths in blocked lanes, the lanes blocked due to incidents, and duration of incident, etc.. These are the normal ways of identifying incidents, with the traditional data collecting methods. In the research of Kamran & Haas (2007), newer ways of data collecting are used, namely the collection of GPS data of vehicles. In this research all vehicles hold a GPS device. First, a road segment with an abnormal pattern is recognized. Second, the road segment is divided into smaller sections and third, abnormal behaviour of a vehicle is detected. Characteristics of this kind of detecting are velocity, direction and the distance to the source. However the research method of the above mentioned research is different of the method used in this thesis. The amount of vehicles holding a GPS device is bigger, because in the research of Kamran & Haas (2007) all vehicles holds a device and in this thesis only some probe vehicles.

Recapitulatory, one can say there has been done a lot of research on the subject of RTTI, but no, known by the author, research on the influence of incidents on travel time. To give an accurate and reliable prediction of the travel time it is important, that the prediction model contains an incident part. Also on the subject of recognizing incidents a lot of research has been done, but again the previous research method is different to the method used in this thesis. In this thesis the influence of incidents on the travel time is examined. Other than previous researches, this research uses a small amount of GPS data and visual data to identify incidents and predict delays. In chapter 5 more research is done on incidents and in this chapter the term incident is explained. In this thesis first, incidents patterns are tried to be recognized in the data, whereby first the focus is on common incidents. Second the possible delay (the difference between the Estimated time of arrival and the real time of arrival) in the travel time are considered.

3 Stakeholders in Public transport

Besides making the system work, thus the technical aspects discussed in the previous chapter, it is important to know what stakeholders want, thus the commercial aspects. The RTTI system IFluxo is developing has two important stakeholders from whom they are depending, namely the bus passenger and the bus companies. First, their wishes, important demands and the match between these are listed, based on literature. Second, a new way of matching the interest is presented and in the last paragraph this will be elaborated.

3.1 Interest stakeholders and matching the interest

3.1.1 Interest Public Transport Operators

The public transport operator in Brazil is focused on making profit. They will achieve this by making a high turnover and lower costs. The results of a conversation about the bus companies with W. Vonk and R. Hulleman on the 13th of April 2009 are as follows. Bus companies are, of course, focused on making profit by selling tickets. On the other hand they look at the costs of letting a bus ride. An important outcome of the conversation was the way bus companies look at costs and profits. Only the visible direct costs are important for the Brazilian bus companies. For example, tire wear, fuel costs and labour costs. They do not think about non-visible costs, for example, a better bus schedule, by which with a smaller number of buses, the same frequency is reached. IFluxo can provide these insights.

These findings are confirmed in a report of Duarte & De Souza Jr (2003). In their research about improving the public transport quality in Brazil they wrote about the decrease in passengers taking a bus and the decrease in comfort and the correlation between these two aspects. Because of the decrease in demand, bus companies keep the same buses for over more that 10 years and wages of bus employees are kept low. This will result in a decrease of comfort and safety, which will at the end result in a decrease of bus passengers.

3.1.2 Interest bus passenger

The bus passenger and potential bus passenger have important demands when it comes to bus journeys. Comfort, travel time, waiting time, information and costs are the most important ones. By developing a SMS-service, the comfort and costs will not really change. Only the information service and the total time a passenger needs for his journey will change. Bovy & Van Goeverden (1994) suggest the travel time factor (TTF) includes all these different times such as time needed for walking to the bus, waiting time, real travel time, and the time it will take to get to your destination. The TTF expresses the travel time ratio of public transport / car. If this ratio is below two, more people will choose to travel with the bus.

Besides this, RTTI can be of use to both frequent and less frequent passengers (Bovy & Goeverden, 1994). The choice of transportation is not affected dramatically by the frequency of travelling. Vivier, Willequet, Mezghani, Dauby, & Marlier (1999) state that passengers find information important, but according to the author of 0*Openbaar vervoer moet leren luisteren* the information given has to be simple and compact and above all correct! One can say that for passengers a low TTF and simple, compact and correct information are, in this case, the most important demands.

3.1.3 Matching the interest

To come to a bus ride with passengers in the bus, there has to be a match between the operator and the passenger. Ongkittikul (2006) uses the figure shown in 'Figure 1' to show the match between these two stakeholders. Hereby he equals the output of the PT operator to the input of the potential passenger.

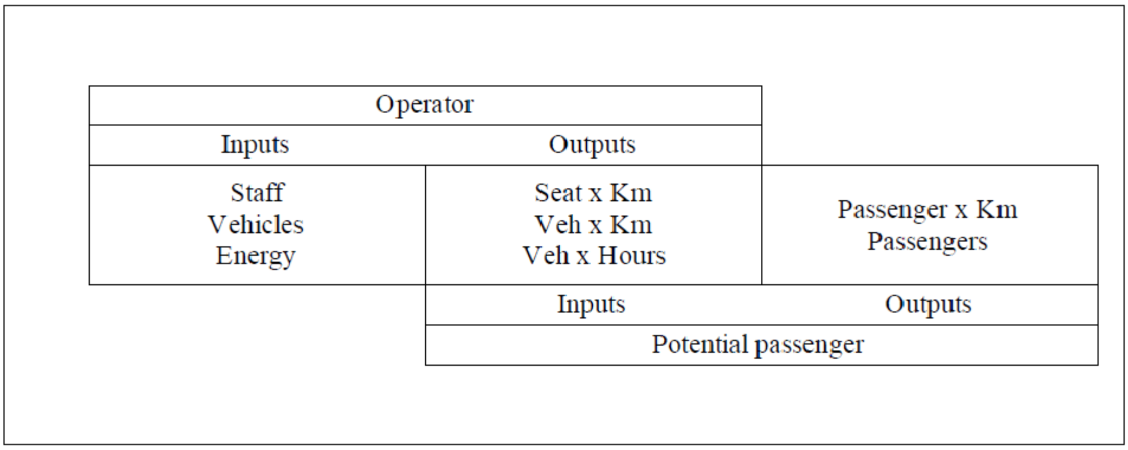


Figure 1: Public Transport system by Ongkittikul

Ongkittikul figure is incomplete. He equals different parameters, which are not quite the same. In the next two paragraphs a new figure is presented and elaborated in a better way.

3.2 New way of mapping public transport system

The public transport operator (PT operator) puts a service in the market using input. Simply said, the PT operator has buses and employees (PT operator costs). These inputs lead to an output: A number of buses, during a certain time period on a part of the road network. The potential passenger has a certain input, demands and wishes about his journey (profit potential passenger). If there is a match between this input and output, the potential passenger will change into a passenger, the passenger will make a ride with the bus (profit passenger). To make this trip, he needs to pay the ticket price. This ticket price is the outcome, as well for the passenger (costs), as for the PT operator (profit). This is put into Figure 2 which is located besides.

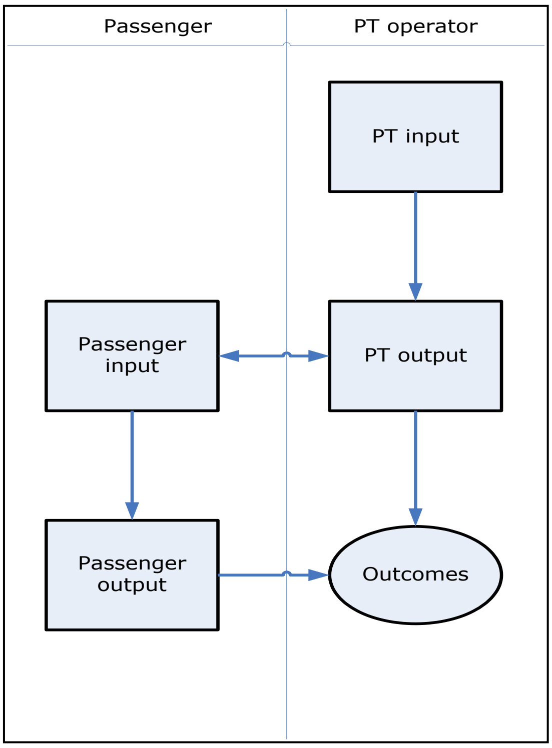


Figure 2: New PT system

3.3 *Elaboration of public transport system*

Figure 2 can be elaborated into a more detailed figure. Below, each part of the figure is explained and finally presented in Figure 3.

3.3.1 Input Public Transport Operator

The PT operator has a certain amount of buses (fleet) and a certain number of employees. The input consists of fleet and employees. Fleet can be divided into buses, maintenance and fuel. Important parameters on this subject for the PT operator are the age of the bus, tire wear and the amount of fuel used.

3.3.2 Output Public Transport Operator

As said before, the output of the PT operator is a certain number of buses in a certain time period on a part of the road network. The buses drive with a frequency, speed and acceleration and contain a number of seats and employees. These buses cover a number of kilometres of the total road network. Besides this, some buses are equipped with extra functions, such as air-conditioning, a low entry or a lift for disabled passengers. Last but not least, the PT operators can give information about this service. This information can be provided on the internet or by the employees.

3.3.3 Input Potential passenger

The input of the potential passenger consists of their demands and wishes. The availability is an important factor, which consists of time accessibility and place accessibility. Besides this, during the trip one wants a safe and comfortable trip with an as low as possible travel time. While waiting for the departure of this trip, the passenger wants a good sheltered waiting place. Finally, he wants a good service: information and help during the ride are examples.

3.3.4 Output potential passenger

The output of the potential passenger depends on the decision he makes. If he wants to make the trip, he pays the ticket price and makes a bus trip for a certain time, on a certain bus and for a number of kilometres.

3.3.5 Outcomes

The total outcomes are a number of passengers multiplied by the ticket price.

3.3.6 Match input passenger and output PT operator

The PT wants a certain outcome, namely as much as possible profit. To reach this goal, he needs to level his outputs with the inputs of the potential passenger. The smaller the differences between these two, the bigger the chance a match will occur and a passenger will get in the bus and pay the ticket price. To understand these matches between in- and output, two examples are worked out below, namely the PT operator output ‘frequency’ and ‘information’.

First, frequency effects the time accessibility. Frequency is traditionally an important parameter in public transport. As shown in Figure

3: **Elaboration new PT system**frequency is connected to time accessibility. When the frequency is higher, the time accessibility for the potential passenger is better.

Second, information is very important in this thesis. With information a lot of input parameters of the potential passenger can be influenced. In the previous example one can see that frequency can influence the time accessibility, but also information can influence time accessibility. In this way the subconscious time accessibility is adapted. If the potential passenger gets a text message about the departure time of his bus, he can calculate his departure time at his home. In this way he thinks the bus is always at a right time, for the passenger, at the bus stop and the waiting time is minimized. Besides this, information can influence the comfort by seat reservation or the service for the potential passenger, because other information can be send by text message.

Below the above mentioned information is listed in Figure 3.

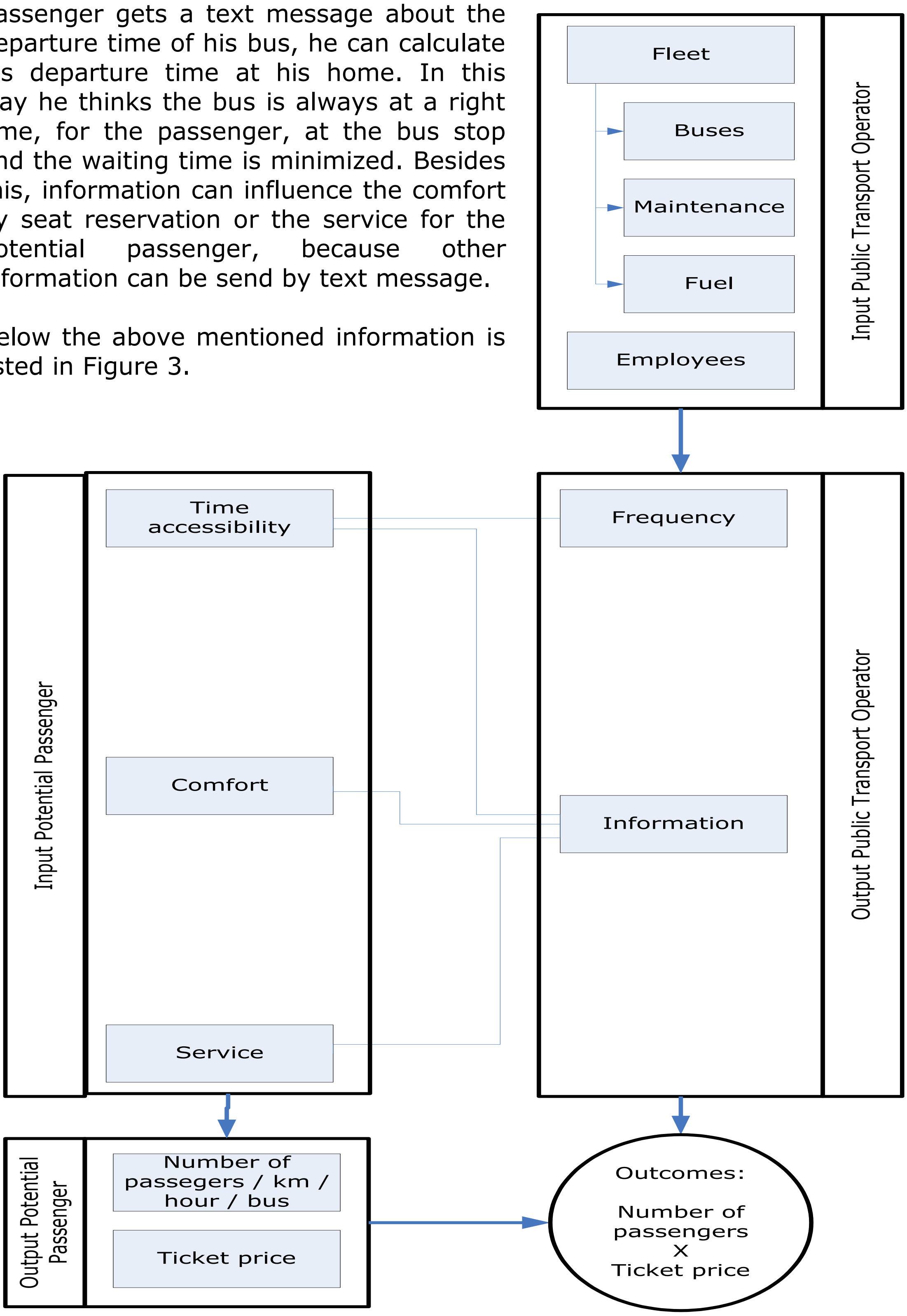


Figure 3: Elaboration new PT system

The ‘Associação nacional das empresas de transportes urbanos’ (the national association for urban transport) in Brazil has done research on the quality and results of the PT operators in urban areas. They come up with the same results as Duarte & De Souza Jr (2003), but use more concrete parameters. Fleet age, tariff and number of passengers per km or vehicle are important indicators for the loss of interest of the passenger in public transport. It is important for the PT operator to change these indicators in his advantage. The new way of mapping the problem as presented in Figure 3 and not like, for example, Ongkittikul (2006) is to get insight in other ways of lower costs and increase profits. For example, the above mentioned link between ‘information’ and ‘time accessibility’ is another way of improving the time accessibility of the passenger and maybe even a way of lower the costs for PT operator. This is due the information supply which increase the time accessibility, so the PT operators may decrease the number of buses per hour, increase the number of passengers in a bus and lower de costs and higher the profit.

After presenting the schedule and the different way of looking at the problem in the interview, Mr. Barbosa corroborated the way of mapping the problem. Hereby he gives two examples of differences in interest. First, the PT operator invests a big amount of money in new buses and not in the bus driver, while research point out that 78% of the passengers considers the interaction with the bus driver as most important. Second, passengers consider problems the PT operator does not know about, or does not want to listen to. 22 % of the passengers once experienced a problem in the bus, where a defect on the stop button (so the bus driver will not stop), or damaged banks are the most mentioned. These examples illustrate again the importance of a correct match between the two most important stakeholders.

4 Case description

There are a lot of differences between The Netherlands and Brazil. Below first the traffic situation of Brazil is explained by mentioning the general traffic conditions, the bus network and the specification of the bus lines the data is collected. Second and last the preparations made for analysing the difference in ETA made by irregularities and incidents are explained. The hardware and software are mentioned and the method of map-matching and calculating the average travel time are explained.

4.1 Traffic situation in Brazil

4.1.1 Traffic picture

The traffic, especially during rush hour, is very intensive. Almost every street is an one-way street and has two to five lanes. The network design is based on only motorized vehicles and pedestrians. Motorized vehicles always have priority and they use it. There are almost no bicycle paths and the few cyclists present have to cycle on the road.

The quality of the roads is often low. Pits and holes appear very often and roads are not plain, so because of this, drivers have to dodge a lot. Besides this, drivers can pass other drivers on both sides and acceleration and deceleration are intense. The driver really has to fight for his place on the road!

4.1.2 Bus network structure

In Rio de Janeiro there are 49 different bus companies who all want to transport potential bus passengers (Secretaria Municipal de Transportes, 2006). The results of this thesis are based on the collected data of line 456 and 457 of bus company 'Viação Acari'. Besides these legal companies there are a lot of illegal 'companies' with smaller vans, see Figure 4.

Figure 4: Illegal van

Figure 5: Turnstile

The buses of the legal companies have a high entry which is unfriendly for older and disabled persons. A probable time saving measure is the presence of a conductor. This allows the passengers to get in fast without paying directly to the driver. With a small amount of passengers getting in the bus, the bus already can drive away while the passengers pay to the conductor, see Figure 5: Turnstile

Line 456 and 457 are the bus lines from the neighbourhoods Botafogo to Méier. The lines cross different parts of Rio de Janeiro and face different kinds of roads. In 'Appendix B: Road types' the different road types are given. Lines 456 and 457 consist of 'Via Expressa (80km/h)' (Blue) and 'Via Arterial (60 km/h)' (Red) roads, see Figure 6: Road type. So the lines have double functions and so different section velocities and bus stop densities.

Figure 6: Road type

5 Incidents and irregularities

In the previous chapters background information is given. Previous researches are discussed, technical terms are explained, a new form of matching interest of stakeholders is presented and the situation in Rio de Janeiro is pointed out. The things missing in previous literature but very important for this thesis, are an explanation of the terms irregularity and incident and the labelling of the irregularities and incidents that can possibly happen. The next chapter will discuss this by answering the first research question, namely:

What do the terms irregularity and incident refer to and which kind of irregularity and incident will play a part in predicting travel times?

The outline of this chapter will be as follows. First, the definition of the terms, used in this thesis, is given after which the different occurring irregularities and incidents are listed. At last the conclusion is given which contains a first version of the decision tree.

5.1 Definition irregularity and incident

The word "incident" does not have one official definition, neither in the daily literature nor in the transportation literature. In the daily literature Van Dale defines an incident as "an unforeseen event" or as "a disturbing occurrence". In the transportation literature there are a few slightly different definitions. The Traffic Incident Management Handbook for example defines an incident as "any non-recurring event that causes a reduction of roadway capacity or an abnormal increase in demand". Accidents, broken vehicles and also special events (soccer games and concerts) fall within this definition. Another definition, given by the Traffic Management Data Dictionary, is "an unplanned randomly occurring traffic event that adversely effects normal traffic operations". From their perspective they make a difference between planed and non-planed events, for example, respectively roadwork and accidents (Balke, Fenno & Ullman, 2002). Within the scope of this project the definition for incident has to be narrowed down.

Incident: "Any non-recurring traffic event that adversely effects the estimated time of arrival of the bus."

Irregularity: "Any recurring traffic event that adversely effects the estimated time of arrival of the bus."

'*Non-recurring traffic event'* includes, what Balke, Fenno & Ullman (2002) considers as 'planned events', for example roadwork. In this project the incident 'roadwork' is not considered planned, because of the absence of information. Perhaps after some days of data collecting 'roadwork' is a recurring traffic event. More about this in paragraph 7.2.2.

Recurring traffic event: All recurring events as well in demand as in traffic situations are included.

'Adversely effects the estimated time of arrival of the bus' is not only an increased ETA, but also a decreased one.

5.2 Possible irregularities and incidents

Below a list is given of the different irregularities and incidents that can take place on or around a bus ride and which affects the ETA. Just as a normal bus stop, rush hours or traffic lights, irregularities have a recurring character. Normally, these irregularities are included in the 'continues prediction of ETA' and considered as noise. In this thesis, they considered as an irregularity and an attempt to recognize them in the GPS data and to say something about the change in ETA is made. The irregularities and incidents are based on personal experiences in Rio de Janeiro and are shown in Table 1: Allocation of incidents.

Irregularities : Recurring	Incidents: Non-recurring
• Normal bus stop • Rush hours • Traffic lights • Passengers getting in/out wherever they want • A lot of passengers getting in / out at bus stop • Bus is not stopping at bus stop • Passengers asking information to the bus driver • Extreme demand	• Road maintenance • Broken bus • Accident • Absence of bus driver • A disabled person have to get in the bus • No conductor in the bus • Broken turnstile • Roadblock • Assault

Table 1: Allocation of incidents

Most of the irregularities and incidents speak for themselves, but a short explanation is given in Appendix D: Incidents explained.

5.3 Conclusion

The central question in this chapter was:

What do the terms irregularity and incident refer to and which kind of irregularity and incident will play a part in predicting travel times?

The incident in paragraph 5.2 all fall within the definition of the terms irregularity and incident and stated below.

Incident: "Any non-recurring traffic event that adversely effects the estimated time of arrival of the bus."

Irregularity: "Any recurring traffic event that adversely effects the estimated time of arrival of the bus."

One can expect that the irregularities have a less drastic effect on the ETA then the incidents. Besides this, the expectation is that irregularities are not recognizable in the data, because the irregularities are less extreme than the incidents and because they appear often (recurring). Thus the irregularities can be included in the continuous prediction. More about this in chapter 7 and 8.

6 Technical aspects

Besides the GPS information and the Estimated Time of Arrival (ETA) of buses another technical aspect of RTTI is important in this project, namely the way of reaching the passenger to provide him or her with the personal RTTI. To give a complete picture of the project this is also mentioned under technical aspects.

6.1 The use of GPS, Geographic Information System (GIS), and Map-Matching

To explain the use of GPS, GIS and Map Matching within the project and this thesis, first a short definition and explanation of the terms is given.

6.1.1 Global Positioning System (GPS)

GPS is a satellite-based radio navigation system. The GPS satellite constellation consists of 32 satellites, arranged in six orbital planes. Each satellite carries a high precision atomic clock and broadcasts messages at regular and known time instants. Each message includes an identity number and the location of the satellite. A receiver on the ground decodes the message and uses the signal propagation time to calculate a pseudo range. In order to determine its position, the receiver needs to know both the pseudo ranges to the satellites as well as their positions. Simultaneous observation of at least four satellites permits determination of the 3D coordinates of the receiver. The 3D coordinates contain latitude, longitude and altitude (Fjellström & Andersson, 2004). Besides these parameters, the receiver gives date, time, speed and an indexing number.

6.1.2 Geographic Information System (GIS)

GIS captures, stores, analyzes, manages, and presents data that is linked to a location (Wikipedia, 2009a). One point in a GIS is named a Point of Interest (PoI). This PoI is an important point on the map (Wikipedia, 2009b). In this case it can be for example a bus stop, a traffic light, etcetera. In this thesis all Points of Interest are bus stops. Between these points average travel time is known and travel time can be predicted.

6.1.3 Map-Matching and Nearest Neighbourhood Method

In this paragraph the principle of map-matching is explained using the nearest neighbourhood method. First, a theoretical explanation is given and second a case based example. Last, the hardware and software used is mentioned.

Theoretical map-match

The map-matching principle consists of allocating some collected GPS coordinates to the nearest PoI, the nearest bus stop. For example, the GPS points called 'P' and 'Q' will be map-matched to PoI 'A', because it falls within the preliminary determined area of PoI 'A' and GPS points 'R' and 'S' are map-matched to PoI 'B'. As one can see, GPS points 'X' are not in the area and will not be matched to neither PoI 'A' nor PoI 'B'. See 'Figure 7: map-match'.

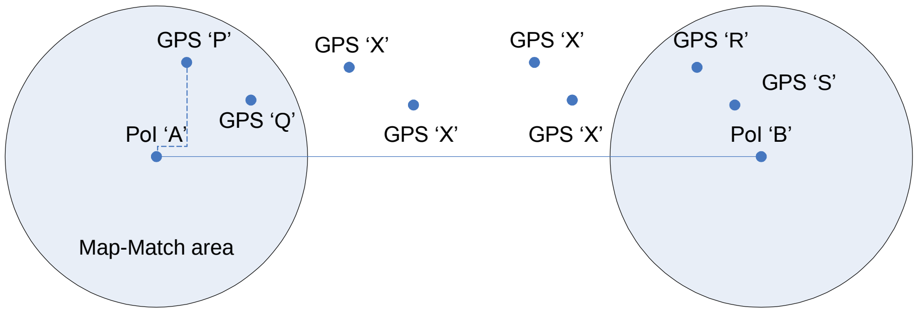


Figure 7: map-match

Now GPS points P, Q, R and S are selected, the nearest point will be linked to the PoI where it is map-match to. Thus, in 'Figure 7: map-match', GPS point 'S' will be linked to PoI 'B', because it is closest to the PoI. Al data from GPS point 'S' will be used to calculate travel times between PoI.

Case based example

To give a better explanation, tables of a case example will be used to show how the process of map-matching and nearest neighbourhood calculation is done.

To start with, two tables are necessarily. One table with al the GPS data, that is data of every second of a random journey. This table contains the latitude and longitude coordinates of every second of the journey, Table 2: GPS data The second table contains the latitude and longitude of the Points of Interest, the bus stops, Table 3: PoI data.

INDEX	DATE	TIME	LATITUDE	LONGITUDE
1	2009/4/8	11:35:03	22,946827	43,180979
2	2009/4/8	11:35:05	22,947013	43,181228
3	2009/4/8	11:35:06	22,947032	43,181140
4	2009/4/8	11:35:07	22,947071	43,181115
5	2009/4/8	11:35:08	22,947024	43,181106
6	2009/4/8	11:35:09	22,947020	43,181093
7	2009/4/8	11:35:10	22,947005	43,181081
8	2009/4/8	11:35:11	22,946991	43,181070
9	2009/4/8	11:35:12	22,946991	43,181067
10	2009/4/8	11:35:13	22,946992	43,181058
11	2009/4/8	11:35:14	22,946990	43,181050
12	2009/4/8	11:35:15	22,946985	43,181047
13	2009/4/8	11:35:16	22,946986	43,181047
14	2009/4/8	11:35:17	22,946984	43,181045
......	......	......	......	......

Table 2: GPS data

INDEX	LATITUDE	LONGITUDE
1	**22,941313**	**43,182282**
2	22,933734	43,184360
3	22,917403	43,194276
4	22,916743	43,195195
......	......	......

Table 3: PoI data

The first step one has to take, is to determine an area around the PoI. In this case the area is chosen to be + and – 0,002000 in the latitude and longitude of the PoI. This is equal to approximately 200 meter. Thus, for PoI 1 (Black in Table 3) all GPS point of Table 2 with a latitude between 22,939313 and 22,943313 and a longitude between 43,180282 and 43,184282 get the match with PoI 1 of Table 3 (black).

The second step is to calculate the distance between the matches. Thus, calculate the distance between the latitude of Table 2 and the latitude of Table 3 and of course the distance between the longitudes. Using these two, the Euclidian distance, the real distance to the Point of Interest can be calculated. In 'Table 4: Step 1 and step 2', the first two steps are visualized. The names 'longitude' and 'latitude' are the coordinates from the GPS data from Table 2, 'poilatitude' and 'poilongitude' are the coordinates from the PoI from Table 3 and 'distance latitude', 'distance longitude' and 'real distance' are the calculate distances as mentioned at step 2.

INDEX	DATE	TIME	LATITUDE	LONGITUDE	poilatitude	poilongitude	distance latitude	distance longitude	real distance
......	......	......	......	......	......	......	......	......	......
106	2009/4/8	11:36:49	22,941557	43,180879	22941313	43182282	244	1403	1424
107	2009/4/8	11:36:50	22,941474	43,180904	22941313	43182282	161	1378	1387
108	2009/4/8	11:36:51	22,941394	43,180917	22941313	43182282	81	1365	1367
109	2009/4/8	11:36:52	22,941328	43,180929	22941313	43182282	15	1353	1353
110	2009/4/8	11:36:53	22,941272	43,180935	22941313	43182282	41	1347	1348
111	2009/4/8	11:36:54	22,941232	43,180940	22941313	43182282	81	1342	1344
112	2009/4/8	11:37:28	22,940402	43,182052	22941313	43182282	911	230	940
113	**2009/4/8**	**11:37:29**	**22,940384**	**43,182176**	**22941313**	**43182282**	**929**	**106**	**935**
114	2009/4/8	11:37:30	22,940346	43,182196	22941313	43182282	967	86	971
115	2009/4/8	11:37:31	22,940310	43,182211	22941313	43182282	1003	71	1006
116	2009/4/8	11:37:32	22,940254	43,182276	22941313	43182282	1059	6	1059
117	2009/4/8	11:37:33	22,940169	43,182387	22941313	43182282	1144	105	1149
118	2009/4/8	11:37:34	22,940097	43,182434	22941313	43182282	1216	152	1225
......	......	......	......	......	......	......	......	......	......

Table 4: Step 1 and step 2

Step three is to select the nearest GPS point from Table 5, thus with the shortest distance to the Point of interest. If this is done with all the Points of Interest, the travel time between the Points of Interest can be calculated. This is shown in Table 5. As one can see, all data in the black hatched rows in the tables correspond to each other.

index	DATE	TIME	LATITUDE	LONGITUDE	poilatitude	poilongitude	Real distance	traveltime
113	**2009/4/8**	**11:37:29**	**22,940384**	**43,182176**	**22,941313**	**43,182282**	**935**	**00:02:19**
252	2009/4/8	11:39:48	22,933766	43,184491	22,933734	43,184360	135	00:02:35
338	2009/4/8	11:42:23	22,917380	43,194137	22,917403	43,194276	141	00:01:22
420	2009/4/8	11:43:45	22,916772	43,195210	22,916743	43,195195	33	00:00:42
462	2009/4/8	11:44:27	22,914762	43,196019	22,914744	43,195983	40	00:02:49
......	......	......	......	......	......	......	......	......

Table 5: Step 3, select nearest point and calculate travel time

Combining all travel times from the different tables (different days) together will give an average travel time between two Points of Interest and a average travel time for the whole journey. The result of this is given in chapter 8, where they also are analysed.

Hardware and software

In order to collect the data for the database, hardware is necessary. Using a GPS Travel Recorder (GPS-TR) this data is collected (QSTARZ BT-Q1200 Ultra Solar Travel Recorder). The parameters collected by the GPS-TR are the following: date, time, latitude, longitude, altitude and speed. By several actions, shown in Appendix E: Export data to MS Access, using Microsoft Office Excel and Microsoft Office Access, a database containing all this data is established in Microsoft Office Access.

Using another GPS-TR, with the ability of pointing out your own PoI, another table is constructed with only PoI (QSTARZ BT-Q1000P Travel Recorder).

6.2 Estimated Time of Arrival (ETA)

The Estimated Time of Arrival is important to get the right information about the arrival times of busses to the costumer. ETA can be divided in two important aspects, which are the continuous prediction and the incidental prediction. However, these two aspects should be taken together because they interact with each other and both influence the estimated time of arrival. Therefore it is important to work on one aspect of ETA but to keep the other aspect in mind all the time. In this thesis the focus will be on incidental prediction. The definition of incidental prediction will be given in paragraph Definition irregularity and incident. The information mentioned above is combined in a schematic map in Figure 8: Aspects of ETA.

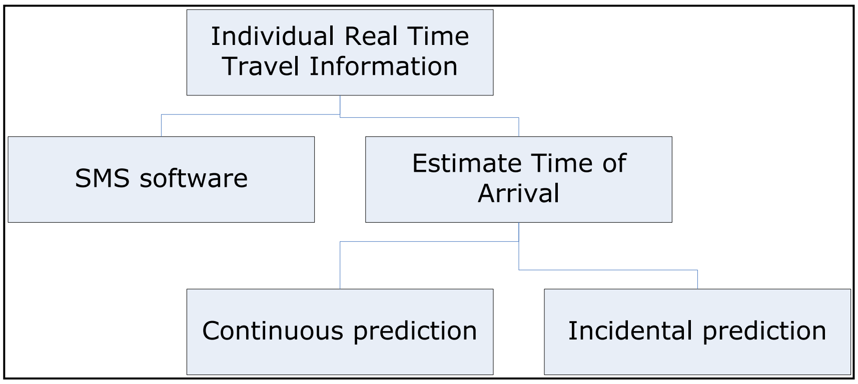


Figure 8: Aspects of ETA

In general, the average travel time between two Points of Interest is in the database. Thus when a bus is passing PoI 'A', the travel time needed tot get to point PoI 'B', 'C' etcetera can be estimated through summing the different average travel times over the section(s). Of course there are different scenarios which will differ by, for example day types, weather type or time.

6.3 Short Messenger Service (SMS)-Service / Internet-Service

For the passenger individual RTTI is simple. He receives an SMS telling him when his bus or tram is due to arrive at the bus stop where he wants to board. He can subscribe for this service via internet (regular journeys/times for instance commuting to work) or via SMS (incidental journeys) (Vonk & Thomas, 2007).

Via internet the concept will look like www.busms.com and by SMS the passenger has to send the same information but packed in a SMS. An example of the interface is given in Appendix A: Interface www.busms.com

7 Data collection and reduction

In the previous chapter the terms irregularity and incident are explained and irregularities and incidents are listed and allocated. To make statements about these irregularities and incidents, different data sources and consequential parameters are necessary. The next chapter will give an answer to the second research question, namely:

What kinds of data sources are necessary to make objective statements about the nature of the irregularities and incidents?

The outline of this chapter will be as follows. In previous research about incident detection, different data sources and parameters are used for detecting these incidents. In the first paragraph these data sources and parameters are mentioned and the way of using indicators or patterns is explained. Secondly, the difference between the method of detecting incidents of previous literature and the method used in this thesis is explained, after which the possible indicators, which can be used in this thesis, are mentioned. Because of the lack of sufficient data about the traffic situations and bus related incidents, such as: accidents, broken buses and roadblocks, an interview with bus companies is presented as data source. This will give a good impression about the procedures followed in case of these traffic situations and bus related incidents. Concluding, the answer to the second research question is given.

7.1 Literature

The use of a certain indicator depends on the way of data collecting. Below a list is given of data collecting methods and belonging parameters.

- Data from inductive loops (speed, flow rates, occupancy)
- Data from traffic control and ramp metering systems (phases, green times, speed, occupancy, presence).
- Data from automatic vehicle identification (AVI) or video detection systems (speed, travel times)
- Data from probe vehicles, instrumented vehicles driving in the traffic flow that send position at regular time intervals to a monitoring system (travel times, routes)
- Historical and forecast weather data
- Temporal information (season, day of week, hour of day, special occasions)
- Infrastructure related information such as planned road works and lane closures and ramp closures.
- Data from incident detection systems (lane closures, severity, expected duration)
- Anecdotal data (*Travel Time Prediction in Rhode Island*)

Sheu (2004) used inductive loops to collect lane traffic counts and the occupancy numbers per lane. In this way he could characterize incidents by time-varying lane-changing fractions and queue lengths in blocked lanes, the lanes blocked due to incidents, and duration of incident. When an incident takes place on a certain lane, cars will change lane and the occupancy on the lanes upstream will be larger than downstream. In this way he was able to recognize the pattern of an incident. Sheu (2004) hereby adds the comment about similar patters caused by recurrent traffic jams and the incident and the difficulty to distinguish them.

As mentioned in paragraph 'Previous research', Kamran & Haas (2007) used GPS data of vehicles. They first used the indicator speed to determine a road segment, and afterwards different indicators were used to locate the incident, such as speed, direction and position. In this research, an important comment is added, namely the importance of having a lot of data. If the data is scarce, traffic incidents can be difficult to detect. Irregularities occur more often, but have a less characteristic character so detecting them in the data is also difficult.

7.2 Parameters

The difference between the previous mentioned research and this thesis is the way of collecting data. This difference also means a difference in the use of parameters. In this thesis data is collected by GPS-TR which is used in a single bus. Patterns and conditions between two points on a road can not be compared and no occupancy number of the road can be collected. Below two paragraphs discuss different data sources. The first paragraph names the available data, the second paragraph the unavailable and not useful data sources.

7.2.1 Available data in this stage

There are two data sources available in this stage of the project. These are the GPS Travel Recorder and a interview with the director of the federation of bus companies.

GPS-TR

As mentioned, only one GPS-TR is used while sitting in a bus. This GPS-TR provides us directly and indirectly with parameters for irregularity and incident detection. Below a table shows an example of the direct collected data, Table 6: Direct possible parameters, and the indirect parameters, Table 7: Indirect possible parameters. Acceleration is the speed difference between two seconds. In 'Appendix F: Calculate acceleration' the way of calculating the acceleration is shown. 'Bus stop' is pointed out by another GPS Travel recorder and is map-matched to a point. In paragraph 6.1.3 this is explained. The 'Traveltime' is the time needed to go from 'busstop' A to 'busstop' B etc. and the 'digital distance' is difference in distance between the latitude and longitude of the bus stop and the latitude and longitude of the real GPS points, which is also explained in 6.1.3.

Direct possible parameters								
INDEX	**DATE**	**TIME**	**LATITUDE**	**N/S**	**LONGITUDE**	**E/W**	**ALTITUDE**	**SPEED**
number	**yyyy/mm/dd**	**hh:mm:ss**	**Digital Degrees**		**Digital Degrees**		**Meters above sea level**	**km/h**
1393	2009/5/6	01:23:25	22,912598	S	43,236575	W	-53,33	32,360386
1394	2009/5/6	01:23:26	22,912645	S	43,236541	W	-53,42	29,476582
1395	2009/5/6	01:23:27	22,912675	S	43,236533	W	-53,62	24,636806
1396	2009/5/6	01:23:28	22,912651	S	43,236565	W	-53,82	14,496928
1397	2009/5/6	01:23:29	22,912641	S	43,236607	W	-53,96	10,474814
1398	2009/5/6	01:23:30	22,912632	S	43,236648	W	-53,98	8,076952

1399	2009/5/6	01:23:31	22,912618	S	43,236689	W	-53,79	6,125635
1400	2009/5/6	01:23:32	22,912609	S	43,236697	W	-53,43	4,476992
1401	2009/5/6	01:23:33	22,91261	S	43,236709	W	-53,12	4,472149
1402	2009/5/6	01:23:34	22,912616	S	43,236713	W	-52,84	4,633297

Table 6: Direct possible parameters

	Indirect possible parameters			
INDEX	**ACCELERATION**	**BUSSTOP**	**TRAVELTIME**	**DISTANCE**
number	**km/h/s**	**Letter**	**hh:mm:ss**	**Digital Degrees**
1393	-2,883804	A		
1394	-4,839776			
1395	-10,139878			
1396	-4,022114	B	00:00:03	0,000053
1397	-2,397862			
1398	-0,004843			
1399	0,161148			
1400	-0,825749	C	00:00:06	-0,000035

Table 7: Indirect possible parameters

Interview PT operator

During the ten weeks of data collection almost only irregularities happened. The incidents did not occur very often but they are nonetheless important, if not most important. The incidents have the most extreme impact on the ETA. To map these incidents and their effects, Márcio Coelho Barbosa, the director of the federation of bus companies, is interviewed. This is done to get an impression of the procedures and ways of behaviour in case of incidents. This is not objective data but it will give a good insight in the effects of incidents on the ETA.

7.2.2 Non-available and 'not useful' data

The above collected data is objectively and therefore can give an objective statement about what is happening. All this data is available. Below two kinds of data are mentioned, intern and extern data. Intern data means: Data directly collected from the bus itself. This data is not available in this stage of the project. Extern data means: Data collected from objects other then the bus. This data is not useful for the project because an objective statement about the change in ETA can not be made.

Intern

Kind of data	**Information about**	**Example**
Machine to machine data	Engine activity	Engine is not running: incident happened
The persons counter on the turnstile	Busyness in the bus	If it is busy the exit time is (probably) longer or at least the chance of more people getting out is bigger
Emergency button	Assault or not	Assault committed: ETA is increases

This data is not available, at the time of writing this report but it can be used in the future to help predicting the ETA of buses.

Extern

Kind of data	**Information about**	**Example**
Data about the weather forecast	Kind of weather	Rain: ETA is probably increasing
Tow truck message	Broken bus	Send message to tow truck: bus is broken
Maintenance report	Maintenance on road	Maintenance: ETA probably increases

The above mentioned data is not useful if one wants to predict the time of arrival of a bus objectively. Data about the weather will probably increase the ETA but one can not say with how many seconds or minutes. A message send to a tow truck company is not objective information. In this case, machine to machine data is more objective. A maintenance report of a road says something about a probable increase of the ETA, but the increase in ETA varies per intensity and one can not say how many seconds or minutes it will increase.

The objectivity of the above mentioned data is not sufficient and not accurate. One only can say something about the chance of occurrence of an increasing ETA, but not about the quantitative increase in the ETA. For an accurate and reliable service, this data can not be used. On the other hand, if one wants to say something about the possibility the ETA will change, this data can be used.

7.3 Conclusion

The central question in this chapter was:

What kinds of data sources are necessary to make objective statements about the nature of the irregularities and incidents?

There are four important findings relating this question. First, there is the GPS-TR, which provides different parameters directly, namely: Date, time, latitude, longitude, altitude and speed. Indirectly it provides the parameters: acceleration, bus stop or not, travel time and distance. The second finding is that there are data sources which are useful but not available yet, namely: Machine to machine data, persons counter on the turnstile and the emergency button. Third, the interview is a data source necessary to make statements about the nature of incidents. This data source will provide insight that are not objective, but will give a good understanding of the incidents and the number it occurs. The fourth and last finding is that extern data sources, not directly coming from the bus, are of no use for making objective statements about the nature of the incidents. This is because of the lack of accurate and quantitative data. In this case, one can only say, maybe or perhaps an incident will occur and the ETA will increase.

Thus, only the possibility of occur will be bigger. Besides this, IFluxo wants to be independent of other data sources.

To achieve the objective all findings are put together in a first version of the table. In this table, finally, all information such as the delay, the data source needed and the occurrence of the irregularities and incidents is displayed. In the tables below first the irregularities (recurring) and the incidents (non-recurring) are displayed. Recognizable, says something about the data needed to recognize the irregularities or incidents. In the next chapters this table will be completed.

Recurring situations	Recognizeable
Normal bus stop	yes
Congestie	no
Traffic light	no
Passengers getting in/out wherever they want	no
A lot of passengers getting in / out at bus stop	no
Bus is not stopping at bus stop	yes
Passengers asking information to the bus driver	no
Good / bad weather	maybe if data available

Figure 9: Irregularities

Non-recurring events	Recognizable
Road maintenance	when capasity is reached maybe through delays
Broken bus	if data available
Accident	if data available
Absence of bus driver	yes, when bus stay in garage and if data available
A disabled person in the bus	if data available
No conductor in the bus	no, looks like normal long bus stop
Roadblock	yes, direction is different
Assault	if data available
Extreme weather conditions	if data available
Extra demand	no
Non-recurring congestion	maybe by delay

Figure 10: Incidents.

In the next chapter the GPS-Travel Recorder, visual data and the interview will be used as a data source. The parameters in these data sources will be used to give indicators which consequently will be used to distinguish irregularities and incidents and predict delays for the given irregularity and incident.

8 Results

In the previous chapter two data sources are selected, namely the GPS-Travel Recorder and an interview. With these data sources the next chapter will give an answer to the third research question, namely:

How can the collected data be used to estimate indicators with which different irregularity and incident types can be distinguished and predict delays for a given irregularity and incident type?

The outline of this chapter will be as follows. Paragraph one describes first, the attempts to distinguish irregularities and incidents in data. The process of the incidents and irregularities is monitored, as well within one incident as between incidents and irregularities. In this way irregularities or incidents can be distinguish in the data without using visual data. Second, the change in section time caused by irregularities and incidents is discussed. A new average section time is calculated and the delay caused by each incident or irregularity is analysed. In this way one can say more about the delay a irregularity or incident causes. The second paragraph lists the most important outcomes of the interview with Márcio Barbosa, and finally the third research question will be answered.

8.1 Data

In this paragraph first, the attempts to distinguish irregularities and incidents in data are described and second, the change in section time caused by the different irregularities and incidents is examined.

8.1.1 Distinguish

In the collected data, coming from the GPS-TR, not all irregularities and incidents occur, only the ones listed below.

- Road maintenance
- Passengers getting in / out wherever they want
- A lot of passengers getting in / out at bus stop
- Bus is not stopping at bus stop
- Passengers asking information to the bus driver.
- Partially roadblock

Besides this, also the common bus stop, congestion and traffic lights can be seen in the data.

It appears to be impossible to distinguish irregularities or incidents in the data, because of two causes. First, there is the difference within the irregularities or incident themselves. Second, there are the similarities between irregularities or incidents.

The difference within the irregularities or incidents themselves is big. Not one irregularities or incident or common occurrence is similar. For example, a traffic light characterizes itself by slowing down (part 1), standing still (part 2) and pulling up (part 3). Thus, acceleration is a high negative number and speed decreases, speed is zero and acceleration to, acceleration is a high positive number and speed increases. However, part 1 can be very quick or very slow, this differs from the situation. If a driver sees a red light far ahead, he can react gradually and acceleration would be low. In another situation, acceleration would be high because the traffic light suddenly jumps red. Part 2 can be long or short and part 3 can also be quick or slow, because of different situations.

Secondly, irregularities or incidents look similar in the data. Replace the above words "traffic light" and "red" for "passenger" and "put their hand up / push the stop button" and one got the description of a bus stop. In 'Appendix I: Similar data' these similarities are shown on the basis of data. Five examples are given. Example 1: Short bus stop, Example 2: long bus stop, Example 3: half stop at traffic light, Example 4: long traffic light and Example 5: short traffic light.

8.1.2 Delays

In paragraph 6.1.3 the principles of map-matching and the nearest neighbourhood method are explained. These tools are used to calculate the travel times between the Points of Interest, the bus stops. This is done in MS Access with the queries listed in 'Appendix G: Map-matching with Nearest Neighbourhood '.

Before analysing the change in travel time caused by an irregularity or incident, the data is first filtered to exclude different extremes.

The first filter was time zone, only journeys in the same time zone are selected. The time zone of the morning is 08:30 – 09:30 and the time zone of the afternoon is 16:30 - 19:30. Extreme weather did not occur in the data. Besides this, two extremes with section times of 26:57 (09-05-13) and 14:48 (09-04-14) are filtered out. These times both occur in the journey 1-457. The explanation for these huge sections times is a wrong allocation. After map-matching those in journey 1-456, no extreme section times and journey times are seen.

The days without visual data are included into the average section time and journey time.

Last, the first two stops and last two stops are excluded from analysing. This is done, because of the inaccuracy on the first and last bus stop. At these stops, a lot of GPS coordinates are collected with different times and since the average waiting time of all journeys is 5:57 minutes, the section time and journey time can fluctuate with almost six minutes.

Difference line 456 and 457

These filters result in Table 8 which give an impression of the minimum, average and maximum journey times per line. Besides this, the number of days and the number of existing visual forms are presented in this table.

	1-456	**2-456**	**1-457**	**2-457**
Minimum	26:30	38:35	20:52	29:37
Average	32:08	41:28	23:12	39:06
Maximum	39:36	48:54	24:59	48:18
Number of days data	16	9	7	9
Number of visual forms	8	8	2	1

Table 8: Journey times

As one can see, the difference in journey time of 456 and 457 in the morning (1-456, 1-457) is large, exactly 8:56 minutes. This difference is caused by a difference in bus route. In the morning 457 passes the Maracana stadium on the north side and 456 on the south side, this means that 456 takes a longer route because it goes through a residential area and picks up passengers, see ‘Appendix J: Extra route section line 456’. This results in an extra journey time of 8:54 minutes, which is 28 % of the total journey. The number of passengers getting in and out the bus on this section of the journey is respectively 6 and 5, which is respectively 15 and 18 % of the total journey. Concluding one can say that line 456 picks up 15 and drops 18 % of his passengers in this part of the journey, which result in a 28% longer journey time. This is unfavourable compared to line 457. The numbers are presented in Table 9.

Total journey time	32:08 minutes
Section time	8:56 minutes
Total number of pick-ups	32 passengers
Total number of drop-offs	44 passengers
Section pick-ups	6 passengers
Section drop-offs	5 passengers
% section time	28%
% pick-ups	15%
% drop-offs	18%

Table 9: Difference between line 1-456 and 1-457

Irregularities or incident delays

As mentioned in paragraph 8.1.1, only some irregularities or incidents occurred during the data collection period. These were: ‘a lot of passengers getting in or out the bus’, ‘road maintenance’ and ‘congestion’. First, the irregularities or incidents are excluded from the section travel times (in the data), after which their section travel time is compared to the average section travel time. In the sequel this time is called ‘incident time’. The outcomes will be presented per line.

Line 2-457 (Meier – Botafogo)

- Only one visual form is present of line 2-457, thus there is no analyse done on this line.

Line 1-457 (Botafogo - Meier)

- In the journeys of line 1-457 only road maintenance occurred two times. In this road maintenance, two of the four lanes are closed and as one can see in Table 10 the average incident time is 12 seconds. This minimal deviation is the result of a sufficient capacity with even two lines. No extreme incident time did occur.

	Minimum	Average	Maximum	Measurements
A lot of passengers >5	-	-	-	-
Road maintenance	00:00:07	00:00:12	00:00:17	2
Congestion	-	-	-	-

Table 10: Incident times 1-457

Line 2-456 (Meier – Botafogo)

- The average incident time of ‘a lot of passengers’ is 34 seconds. The places these irregularities and incidents occur were always on the same bus stops, thus this can be included in the continuous prediction.
- Road maintenance occurred four times, all in different sections. The incident times differ from one minute to two minutes.
- There are a lot of congestion cases on this line and the incident times vary a lot. There are two sections where the congestion always occur, thus this can be included in the continuous prediction. Although the big varieties can give problems.

	Minimum	Average	Maximum	Measurements
A lot of passengers >5	00:00:03	00:00:34	00:01:57	12
Road maintenance	00:01:03	00:01:30	00:02:02	4
Congestion	00:01:00	00:02:59	00:07:32	8

Table 11: Incident times 2-456

Line 1-456 (Botafogo - Meier)

- As said with line 2-456, the irregularity ‘a lot of passengers’ occurs often at the same points and the incident time is small. It can be included in the continuous prediction.
- Road maintenance even had negative numbers on this line. Intensity did not reach the capacity of the open roads.
- Also congestion had negative numbers, which means the irregularity not cause a delay. Besides this, again, congestion occurs in the same sections. Because incident times are negative, one can conclude these days are favourable days compared to the average. Nothing can be said about this, because of the half of the days no visual data is available.

	Minimum	Average	Maximum	Measurements
A lot of passengers >5	00:00:14	00:00:26	00:0:44	5
Road maintenance	-00:01:43	-00:00:22	00:00:35	8
Congestion	-00:02:10	-00:01:11	-00:00:27	4

Table 12: Incident times 1-456

8.2 Outcomes interview

As mentioned in paragraph 7.2.1, Márcio Coelho Barbosa, the director of the federation of bus companies, is interviewed. In the first part, the focus was on the match between the PT operator and the potential passenger, as presented in paragraph 3.3.6. The second part focused on the different incidents. Hereby each incident was discussed on the basis of tree questions, namely:

- Procedures for the bus driver. What happens with the bus and also with the passengers?
- How long does it take, as well for the bus as for the passengers?
- How often does an incident occurs?

Below, the most important outcomes are listed of the different incidents.

When a bus can not continue his ride because it is broken, passengers have to get out. The driver and the exchanger lead the passengers into a different bus which is on the same line and the passengers do not have to pay. The bus gets on a tow truck and is dragged away. It differs per line how many times a bus get broken. For example, a busy line with a lot of paying passengers gets newer buses than a non-profitable line. This will result in more broken buses on the non-profitable lines.

The procedure with accidents is the same as with a broken bus. The bus also stops and will not continue his way, because an accident report has to filled in, otherwise the driver has to pay for the caused damage.

The absence of bus drivers is a big issue nowadays in Brazil. Not because of, as thought in advance, sickness or because he overslept himself, but because of the competition of other companies who need drivers.

To prevent assaults in buses, cameras are filming in the bus. The number of assaults in buses is decreasing. There is no panic button in the bus and even when there will be one, drivers will not use it because of the own safety.

Total roadblocks occur not very often and if the occur there is no fixed way for the driver to get around the roadblock.

Nowadays in Brazil not many buses have an elevator for disabled persons. In 2014 all buses in Rio de Janeiro need to have an elevator for disabled persons otherwise they do not get a new license as PT operator. Thus, nowadays not many disabled persons get in a bus, but this will increase in the future.

The trend is to implement Rio Card totally in the buses. Rio card is a card one can use to pay automatically in buses. In this way no exchanger is needed and getting in and out is maybe even faster.

In cases of extreme demand, such as the ending of a match in Maracana soccer stadium, of the ending of a concert at the Sambadrome a lot of people want to get on the bus in a short amount of time. In these cases extra buses are put one the line, but it is still not sufficient.

8.3 Conclusion

The central question in this chapter was:

How can the collected data be used to estimate indicators with which different irregularity and incident types can be distinguished and predict delays for a given irregularity and incident type?

The question consists of two parts. First, there is a distinction of irregularities and incidents trough indicators in de GPS data. Second, there is a prediction of delays for these irregularities and incidents.

Distinguishing irregularities and incidents in the GPS data is in this stage not possible because of two causes. First, there is the difference within the irregularities and incident themselves. A bus stop can be long or short or can have a big or small acceleration. Second, there are the similarities between irregularities or incidents. A traffic light looks the same as for example a bus stop.

Predicting delays for the irregularities and incidents is done based on GSP and visible data. Only (1) ‘a lot of passengers getting in or out the bus’, (2) ‘road maintenance’ and (3) ‘congestion’ occurred. Irregularity (1) always occurs in the same section, incident time is small and fluctuations are minimal. This irregularity can be included in the continuous prediction. Incident (2) sometimes occurs in the same section, incident time is short and fluctuations are minimal. This can be included in the continuous prediction. However, there are times, appearing in the data, that road maintenance not occur in the same sections and incident times are big, varying from one to two minutes. Irregularity (3) always occurs in the same section, but incident time is big and is fluctuating from one minute to 7,5 minutes.

Because only three kinds of irregularities or incidents occur in the data an interview is held. This interview gives more insights in the procedures, number of occurrence and incident times of the other incidents.

The first version of the table with information about the irregularities and incidents, presented in the previous chapter is supplemented below. Supplemented is the frequency, the number of times an incident or irregularity occurs on line 456 or 457, and the delay after detection, severe, moderate or unknown.

Non-recurring events	Recognizable	Frequency	Delay after detection, and ETA change
Road maintenance	when capasity is reached maybe through delays	normal	depends on intensity. From data: 1 till 2 minutes
Broken bus	if data available	incidental	severe
Accident	if data available	incidental	severe
Absence of bus driver	yes, when bus stay in garage and if data available	incidental	unknown
A disabled person in the bus	if data available	incidental	moderate
No conductor in the bus	no, looks like normal long bus stop	incidental	moderate
Roadblock	yes, direction is different	incidental	severe
Assault	if data available	incidental	severe
Extreme weather conditions	if data available	incidental	yes, and increasing when time expire
Extra demand	no	incidental	severe
Non-recurring congestion	maybe by delay	incidental	severe

Figure 11: Incident

Recurring situations	Recognizeable	Frequency	Delay after detection, and ETA change
Normal bus stop	yes	very often	no
Congestie	no	normal	severe, fluctuating extremely from 1 to 7,5 minutes
Traffic light	no	very often	no
Passengers getting in/out wherever they want	no	normal	no
A lot of passengers getting in / out at bus stop	no	depends of linenumber and time zone	no
Bus is not stopping at bus stop	yes	often	no
Passengers asking information to the bus driver	no	normal	moderate
Good / bad weather	maybe if data available	normal	unknown

Figure 12: Irregularities

9 Conclusion

The objective of this research was to develop a table for all possible irregularities and incidents during bus rides, whereby the table is quantified in order to determine the nature of the incident and as a result of that estimate a reliable time of arrival. On basis of the answers of the four research questions and the personal experiences during this thesis the table is developed. First, this table is presented and explained. Second, recommendations for further research will be presented as well as recommendations for IFluxo for continuing the project.

9.1 Conclusion

Below, two tables are presented. Figure 13 contains all incidents and Figure 14 contains all irregularities.

Figure 13 contains all incidents that can happen. The incidents are discussed on a number of characteristics. These are: 'recognizable', 'frequency of occurrence', 'number of time a month', 'delay after detection and ETA change', 'name' and last 'if an extra SMS is necessarily'.

All incidents are not recognizable in the available data and, except for 'road maintenance', all incidents did not occur. Nevertheless, most incidents can probably be recognized when other data sources are available. This other data sources are two different sources. First, Machine-2-Machine data from the engine, panic button and the lift for disabled persons. They can tell if the engine is not running or when the panic button or lift is used. Second, real time weather data can be used to predict extreme weather and possible delays.

The delay after detection and the change in ETA are mostly outcomes of the interview. Only of 'road maintenance' one can say the change in ETA is one till two minutes, but it depends of the intensity.

Finally, the incidents are named and there is decided if an extra SMS has to be send to the passenger. Most incidents have a certain impact on the ETA, thus the passenger needs a message about the change in the arrival time of the bus. The names are divided in, busy, delay, incident and weather. 'Busy' for times delays are caused by busyness, 'delay' when the delay is not extreme but significant, 'incident' when delays are extreme and 'weather' when the delay is caused by the weather.

The recurring situations, irregularities, have a high frequency. They occur often and are a normal event on a bus journey. There is no significant delay and changes in journey time caused by the recurrent situations can be considered as noise in the data. Only bus stop situations can be recognized in the data, because the bus stops are Points of Interest and thus their coordinates are known. Finally, congestion can cause a significant change in the ETA, because of the extreme fluctuation in incident time.

Non-recurring events	Recognizable	Frequency	Delay after detection, and ETA change	Name	Extra sms
Road maintenance	when capasity is reached maybe through delays	normal	depends on intensity. From data: 1 till 2 minutes	delay	depends on delay
Broken bus	if data available	incidental	severe	incident	yes
Accident	if data available	incidental	severe	incident	yes
Absence of bus driver	yes, when bus stay in garage and if data available	incidental	unknown	delay	depends on delay
A disabled person in the bus	if data available	incidental	moderate	busy	yes
No conductor in the bus	no, looks like normal long bus stop	incidental	moderate	-	no
Roadblock	yes, direction is different	incidental	severe	delay	yes
Assault	if data available	incidental	severe	incident	yes
Extreme weather conditions	if data available	incidental	yes, and increasing when time expire	weather	yes
Extra demand	no	incidental	severe	busy	depends on delay
Non-recurring congestion	maybe by delay	incidental	severe	delay	yes

Figure 14: Incidents

Recurring situations	Recognizeable	Frequency	Delay after detection, and ETA change	Name	Extra sms
Normal bus stop	yes	very often	no	-	no
Congestie	no	normal	severe, fluctuating extremely from 1 to 7,5 minutes	delay	depends on delay
Traffic light	no	very often	no	-	no
Passengers getting in/out wherever they want	no	normal	no	-	no
A lot of passengers getting in / out at bus stop	no	depends of linenumber and time zone	no	-	no
Bus is not stopping at bus stop	yes	often	no	-	no
Passengers asking information to the bus driver	no	normal	moderate	-	no
Good / bad weather	maybe if data available	normal	unknown	weather	maybe

Figure 13: Irregularities

9.2 Recommendations

Below five important recommendations are given to develop a better service.

- To recognize incidents and predict associated changes in the arrival time of buses, extra data sources should be used. Machine-2-Machine data should certainly be used. If this data source is used, incidents can be recognized more easily and much faster. The data collected by using M2M data is also directly coming from the bus and thus can be used to give objective statements about the nature of the incident and the delay caused. More research has to be done on the use of real time weather data. This data is not directly coming from the bus, but maybe can say something about the coming delays.
- All coordinates of the whole bus route has to be in the database and not only PoI. In this case one can say if the bus leaves the route. This may indicate to a road block or just the bus driver who chooses another route to travel.
- When sending information to passengers this information has to be correct. The rather have no information at all then incorrect information. Thus it is important to send correct information. In case a incident happens and the cause and duration of the delays is not certain, send the passenger a message with this information.
- Finally, there has to be done more research on absence of GPS data. During this research a gap in the GPS happened several times. From just a few seconds, till the half of a journey. When there is no data send to the database, no prediction can be made. A good example of this event is the Santa Barbara tunnel. Here no GPS data can be collected. When an error in the data occurs or no GPS data is collected, the passenger also has to know he can not expect a prediction.

References

Author: Unknown. *Travel Time Prediction in Rhode Island*. (year: unknown)

Balke, K.N., Fenno, D.W. & Ullman, B. (2002). *Incident management performance measures*. The Texas A&M University System College station, Texas transportation institute.

Bovy, P.H.L & Van Goeverden, C.D. (1994). *De rol van kwaliteit in het openbaarvervoersysteem gemeten aan de verplaatsingstijdsfactor*. TU Delft, Faculty of Civil Engineering Department of Infrastructure and traffic. Contribution to public colloquium 1994, Utrecht.

CROW. (14-04-2009) *Categorie indeling van wegen.* http://www.crow.nl/verkeerstekens/t63_p0_m7_i4168.htm.

Duarte, P. & De Souza Jr., D.I. (2003). *Suggestions toward improvement of public transport service quality*. Unpublished thesis. Universidade Estadual do Norte Fluminense Laboratório de Engenharia de Produção

Fjellström, J. & Andersson, D. (2004). Vehicle Positioning with Map Matching Using Integration of a Dead Reckoning System and GPS. Unpublished manuscript. Linköping University, Sweden.

Hoogendoorn, S. & Van Lints, H (2008). Reistijdvoorspellingen en reistijdbetrouwbaarheid. *NM Magazine, 2008, 3*.

Kamran, S. & Haas, O.C.L. (2007). A Multilevel Traffic Incidents Detection Approach: Identifying Traffic Patterns and Vehicle Behaviours using real-time GPS data. *Proceedings of the Intelligent Vehicles Symposium, IEEE, Istanbul, Turkey, 13-15th June 2007*. pp. 912 – 917

Openbaar vervoer moet leren luisteren. (2008). *Mobility-Linq 7-2008.* pp. 16-18

Ramakrishna, Y., Ramakrishna, P., Lakshmanan, V. & Sivanandan, R. (2006). Bus Travel Time Prediction Using GPS Data. *Transportation Land Use, Planning, and Air Quality - Proceedings of the 2007 Transportation Land Use, Planning, and Air Quality Conference* 320, pp. 124-133

Secretaria Municipal de Transportes. (2006). *Plano Diretor de Transportes da Cidade do Rio de Janeiro*.

Sheu, J.B. (2004). A sequential detection approach to real-time freeway incident detection and characterization. *European Journal of Operational Research, 157.* pp. 471–485

Steehouder, M. e.a. (1992). *Leren Communiceren*. Groningen: Wolters-Noordhoff

Van Dale onlinewoordenboek. (21-04-2009). *Incident*. www.vandale.nl/vandale/.

Verschueren, P., Doorewaard, H. (1995). *Het Ontwerpen van een Onderzoek*. Utrecht: Lemma

Vivier, J., Willequet,S., Mezghani, M., Dauby, L. & Marlier, S. (1999). The Customer is the centre of interest. *Transport Public International - 4/99*. pp. 24-41

Vonk, W. & Thomas, T. (2007). *Development of Individual Real Time Travelers Information for Public Transport and SMS Serverbox to finance the Individual RTTI*. Unpublished manuscript. IFluxo.

Wikipedia. (21-04-2009 a). *GIS*. http://en.wikipedia.org/wiki/GIS.

Wikipedia. (21-04-2009 b). *PoI*. http://nl.wikipedia.org/wiki/POI.

Wikipedia. (06-05-2009 c). *Streekbus*. http://nl.wikipedia. org/wiki/ Streekbus.

10 Appendices

10.1Appendix A: Interface www.busms.com

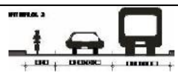
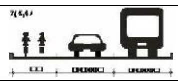
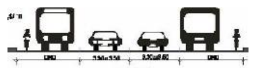
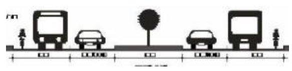
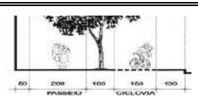
Back to text: Short Messenger Service (SMS)-Service / Internet-Service

10.2Appendix B: Road types

Table B presents the minimum and maximum widths of the infrastructure for the circulation of the bicycle in the existing road system. The measures for arterial roads and express roads are the same.
Table C, D and E show the implementation and the dimensions of the infrastructure for the bicycle in various configurations of the road in respectively: arterial or express roads, collector roads and local roads. Although this tables give a first impression of the measures it will be necessary to determine the measures in more details.

Infra-estrutura cicloviária	Medidas mínimas	Medidas Máximas
Faixa compartilhada veículo/bicicleta	4,00m	4,20m
Ciclofaixa unidirecional em via local	1,50m	2,00m
Ciclofaixa unidirecional em via coletora	1,70m	2,00m
Ciclofaixa bidirecional	2,50m	2,50m
Ciclovia bidirecional	2,50m	-

Table B: measures for the bicycle infrastructure implemented in the different road types (source: Gondim [12])

Classificação Viária	Configuração de via.	Configuração da pista	Desenho	Largura mínima para veículos e bicicletas Largura faixa rolada é 3,30m (entre parênteses a largura desejável)
Via Expressa (80 km/h)		Ciclovia em direção lateral		
Via Arterial 60 km/h	Sentido unico	compartilhada		3,30 x n + 0,90 x2 (3,50 x n + 0,70 x 2)
		ciclofaixa unidireciional		3,30 x n + 1,70 (3,50 x n +1,70)
		ciclofaixa bidireciional		3,30 x n + 2,50 (3,50 x n + 2,50)
	Dois sentidos de tráfego	faixa compartilhada		3,30 x n + 0,90 x2 (3,50 x n + 0,70 x 2)
		ciclofaixa unidireciional		3,30 x n + 1,70 x 2 (3,50 x n + 1,70 x 2)
		ciclovia bidireciional		3,30 x n + (2,00 +2,50) x 2 (3,50 x n + (2,50 +2,50) x 2)
	Com canteiro central	faixas compartilhadas		3,30 x n + 5,00 + 0,90 x 2 (3,50 x n + 5,00+ 0,70 x 2)
		ciclovia central		3,30 x n + 2,50 + 1,20 x 2 (3,50 x n + 2,50 + 1,20 x 2)
		ciclovia bidirecional		3,30 x n +(2,50+2,00) x 2 (3,50 x n + (2,50+2,50) x2)
	Calcadas com ciclofaixas	Calcada com ciclofaixa [1]		1,50 (a) + 2,00 (p)+1,0 (m) +1,50 (CF)

[1] (incluindo afastamento (a) de meio fio e parede, faixa para passeio (p) e mobiliário urbano (m)

Table C: implementation of the bicycle infrastructure in an express and arterial road (source: adapted from Gondim [12])

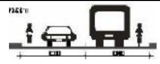
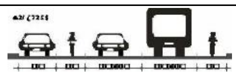
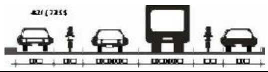
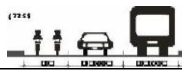
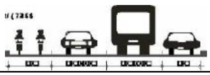
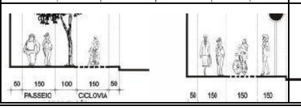
Classificação Viária	Configuração da pista	Desenho	Largura mínima para veículos e bicicletas
Via Coletora (largura de faixa 3,20m) 40 km/h	compartilhada em ambos os lados da via		4,20 + 4,20 = 8,40
	compartilhada de um lado da via e ciclofaixa do outro		4,20 + 3,20 + 1,70 = 9,10
	ciclofaixa nos dois lados da via		(3,20 x 2)+(1,70 x 2)= 9,80
	estacionamento de um lado e ciclofaixas em ambos os lados		2,40 + (3,20 x 2)+(1,70 x 2) =12,20
	ciclofaixas e estacionamentos em ambos os lados da via		(2,40 x 2)+(3,20 x 2)+(1,70x2)= 14,60
	ciclofaixa bidirecional em um lado da via		2,50+(3,20x2) = 8.90
	ciclofaixa bidirecional de um lado e estacionamento do outro		2,50 + (3,20 x 2)+ 2,20 = 11,10
	Calcada com ciclofaixa [2]		1,0 (a) + 1,5(p) + 1,0 (m) + 1,5 (CF) = 5,0m 0,5 (a) + 1,5(p) + 1,5 (CF) + 1,5 (ponto de ônibus)

Table D: implementation of the bicycle infrastructure in a collector road (source: adapted from Gondim 12])

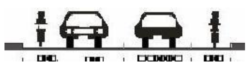
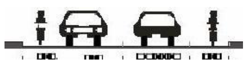
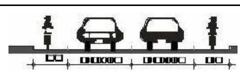
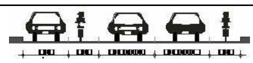
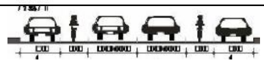
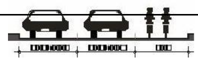
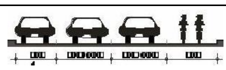
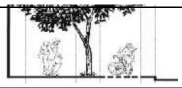
Classificação Viária	Configuração da pista	Desenho	Largura mínima para veículos e bicicletas (largura de faixa 2,70m)
Via Local 30 km/h	compartilhada em ambos os lados da via		4,00 + 4,00 = 8,00
	compartilhada de um lado da via e ciclofaixa do outro		4,00 + 2,70 + 1,50 = 8,20
	ciclofaixa nos dois lados da via		(2,70 x 2)+(1,50 x 2)= 8,40
	estacionamento de um lado e ciclofaixas em ambos os lados		2,20 + (2,70 x 2)+(1,50 x 2) =10,60
	ciclofaixas e estacionamentos em ambos os lados da via		(2,70 x 2)+(1,50 x 2)+(2,20x2)= 12,80
	ciclofaixa bidirecional em um lado da via		(2,70 x 2)+2,50 = 7.90
	ciclofaixa bidirecional de um lado e estacionamento do outro		2,50 + (2,70 x 2)+ 2,00 = 9,90
	Calcada com ciclofaixa [3]		0,9(a) + 1,5(p) + 0,7 (m) + 1,2 (CF) = 4,30m

Table E: implementation of the bicycle infrastructure in a local road (source: adapted from Gondim 12])

Back tot text: Bus network structure

[2] (incluindo afastamento (a) de meio fio e parede, faixa para passeio (p) e mobiliário urbano (m)
Ou de volta de ponto de ônibus excluindo mobiliário urbano e afastamento do meio fio e incluindo ponto.

[3] (incluindo afastamento (a) de meio fio e parede, faixa para passeio (p) e mobiliário urbano (m)

10.3Appendix C: Analysing form bus trip

date: 22-5-2009		bus line: 456		from: Botafogo			to: Meier			page: 1	
		Time: 8:50		Weather: good			Number of persons in bus: 13			Waiting time: 8:45 minuten	
time		Events									
begin time	end time	Traffic light stop	Road block	Bus stop	In	Out	Stop	Road maintenance	Accident	Broken bus	Remaining
00:00:00	00:00:29			25;27	0	1					
00:00:30	00:00:59										
00:01:00	00:01:29										
00:01:30	00:01:59			35;40	0	2		54			2 banen van de 4
00:02:00	00:02:29							27			
00:02:30	00:02:59	50									
00:03:00	00:03:29	10									
00:03:30	00:03:59										
00:04:00	00:04:29										
00:04:30	00:04:59										
00:05:00	00:05:29										
00:05:30	00:05:59	37;48			0	3					VRI bushalte om de hoek
00:06:00	00:06:29	15									
00:06:30	00:06:59	47									
00:07:00	00:07:29										
00:07:30	00:07:59										
00:08:00	00:08:29	10;23			0	2					VRI bushalte op 200 mtr
00:08:30	00:08:59	42									
00:09:00	00:09:29										

10.4Appendix D: Incidents explained

- Broken bus

  During a bus ride, the bus gets engine problems and can not continue its route. Passengers have to get out of the bus and the bus is taken out of service.
- Accident
  - With the bus

    The bus gets involved in an accident. Passengers have to get out and the bus is taken out of service.
  - In front of the bus

    An accident happened in front of the bus and therefore the road capacity decreases. It will take a longer time for the bus to pass this part of the track.
- Road maintenance
  - At bus stop

    The bus is not able to stop at the bus stop so it takes longer for the passengers to get in. Besides this, there is less road capacity.
  - In bus track

    There is maintenance in the track of the bus, so there is less road capacity. It will take a longer time for the bus to pass this part of the track.
- Absence of bus driver

  The bus driver is not coming to work so the bus can not start at its normal schedule.
- Passengers getting in/out wherever they want

  Passengers get in and out at places between the bus stops. This causes deceleration of the bus and so it takes longer to pass this part of the track.
- A lot of passengers getting in / out at bus stop

  The bus has to stop for a longer time than normal at the bus stop, because more than 5 passengers are getting in or out.
- Bus is not stopping at bus stop

  Because no passenger wants to get in or out of the bus the bus can hold its normal speed. This affects the ETA of the bus at the next bus stop.
- No conductor in the bus

  People have to pay the bus driver instead of the conductor. This will take more time.
- A disabled person have to get in the bus

  Some buses can transport disabled persons, but it will take more time then normal.
- Broken turnstile.

  People have to get in the back of the bus.
- Passengers asking information to the bus driver.

  It will take longer for the bus driver longer because he needs to answer the questions.
- Roadblock

Something is blocking the road for a certain period, for example a tumbrel.

- Assault

There is an assault on the bus, as well on the bus conductor or a bus passenger.

Back to text: Possible irregularities and incidents

10.5Appendix E: Export data to MS Access

Software: MS Access
Import gps data

1. open MS Access
2. file – new – empty database
    a. pop up: new database (save)
3. file – externe gegevens ophalen – import
    a. pop up: import
        i. type: text file
        ii. select: TABEL 1
4. import wizard
    a. next
    b. check: 1st row contains fieldnames
    c. next
    d. next
    e. click: advanced
        i. change all types of the field to *text*
        ii. Change the symbol for decimal to the symbol point (.)
        iii. Ok
    f. Next
    g. Check: no primary key
    h. Next
    i. Name the file
    j. Finish
    k. Click; OK

Back to text: Map-Matching and Nearest Neighbourhood Method

10.6Appendix F: Calculate acceleration

1. create column
    l. 1 time click table of interest
    m. Click design
    n. Change column SPEED in SPEED1
    o. Make column SPEED2 and ACCELERATION, type: text
    p. Close and save
    q. Copy second field SPEED1 till last field SPEED1 to SPEED2
2. calculate acceleration
    a. open query's under object
    b. make a new query by clicking new
    c. click ok
    d. click close
    e. click SQL and paste the next text
    f. UPDATE [table name] SET ACCELERATION = (SPEED2-SPEED1)/1000000;
    g. Save the query and run it
    h. Click 2 times ok

Back to text: In this paragraph first, the attempts to distinguish irregularities and incidents in data are described and second, the change in section time caused by the different irregularities and incidents is examined. Distinguish

10.7Appendix G: Map-matching with Nearest Neighbourhood method

Stap 1:

```
SELECT [090416-2-457].* INTO mapmatch
FROM [090416-2-457], [090417-2-457 BS]
WHERE [090416-2-457].latitude>[090417-2-457 BS].latitude-2308 And
[090416-2-457].latitude<[090417-2-457 BS].latitude+2308 And [090416-
2-457].longitude>[090417-2-457 BS].longitude-2308 And [090416-2-
457].longitude<[090417-2-457 BS].longitude+2308;
```

Stap 2:

```
SELECT mapmatch.*, [090417-2-457 BS].latitude AS poilatitude, [090417-
2-457 BS].longitude AS poilongitude INTO jointabel
FROM mapmatch INNER JOIN [090417-2-457 BS] ON
(mapmatch.longitude<[090417-2-457 BS].longitude+2308) AND
(mapmatch.longitude>[090417-2-457 BS].longitude-2308) AND
(mapmatch.latitude<[090417-2-457 BS].latitude+2308) AND
(mapmatch.latitude>[090417-2-457 BS].latitude-2308);
```

Stap 3:

```
alter table jointabel
add column distancelatitude INT, distancelongitude INT
```

stap 4:

```
UPDATE jointabel SET jointabel.distancelatitude = ABS(jointabel.latitude-
jointabel.poilatitude), jointabel.distancelongitude = ABS(jointabel.longitude-
jointabel.poilongitude);
```

Stap 5:

```
alter table jointabel
add column realdistance INT
```

stap 6:

```
UPDATE jointabel SET jointabel.realdistance =
SQR(((jointabel.distancelongitude*jointabel.distancelongitude))+((jointabel.
distancelatitude*jointabel.distancelatitude)));
```

Stap 7: nog niet goed

```
SELECT MIN(realdistance) AS mindistance, poilatitude, poilongitude INTO
eindtabel
FROM jointabel
GROUP BY poilatitude, poilongitude;
```

Stap 8:

```
SELECT jointabel.* INTO eindtabel2
FROM jointabel, eindtabel
WHERE jointabel.poilatitude=eindtabel.poilatitude And
jointabel.poilongitude=eindtabel.poilongitude And
jointabel.realdistance=eindtabel.mindistance
order by index
```

stap 9: dubbele afstand en poilatitude tabel

```
SELECT poilatitude, count(*) AS [count], max(time) AS maxtime INTO
verwijderen
FROM eindtabel2
GROUP BY poilatitude
HAVING count(*)>1;
```

Stap 9,1: verwijder dubbele

```
DELETE *
FROM eindtabel2
WHERE exists(select * from verwijderen
     where verwijderen.poilatitude = eindtabel2.poilatitude and
verwijderen.maxtime <> eindtabel2.time);
```

stap 9,2: nieuwe kolommen creëren, trajecttime en time 2

```
alter table eindtabel2
add column trajecttime TIME, time2 TIME
```

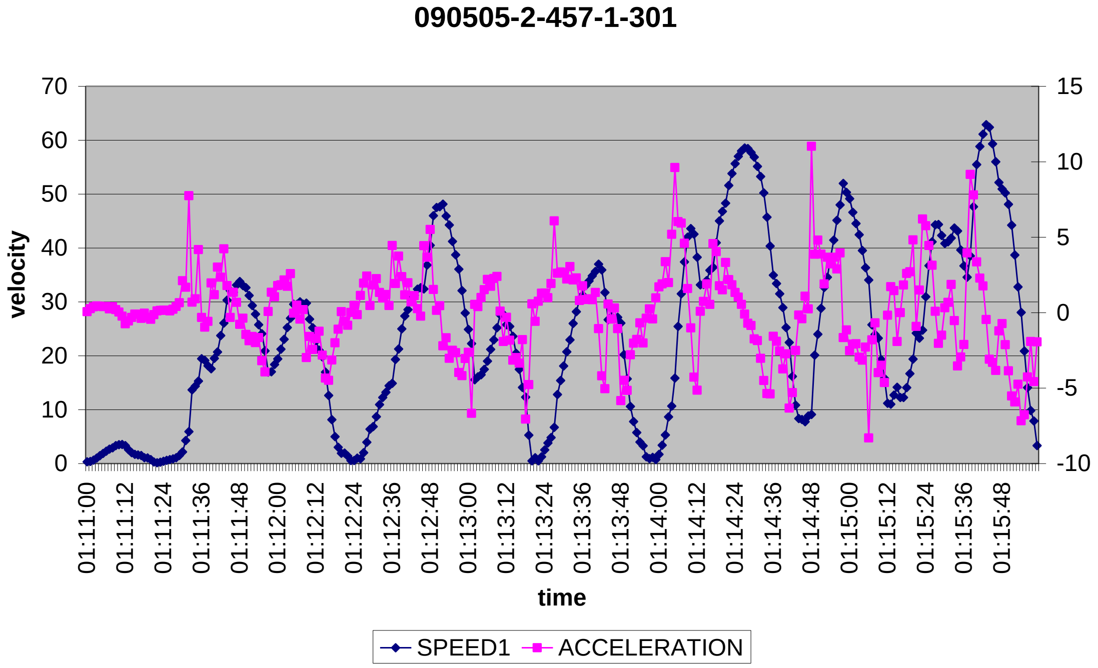
hier tussen 'time' type veranderen naar TIME en records handmatig kopiëren van 'time' naar 'time 2'. Hierbij field 2 tot eind kopiëren van 'time' en plakken in field1 van 'time2'.

stap 9,3: tussentijden

```
UPDATE eindtabel2 SET eindtabel2.trajecttime = (eindtabel2.time2-
eindtabel2.time);
```

Back to text: Map-Matching and Nearest Neighbourhood Method

10.8 Appendix H: Graphic velocity and acceleration



Back to text: Distinguish

10.9Appendix I: Similar data

Example 1: Short bus stop

TIME	SPEED1	ACCELERATION
20:39:37	24.259354	-11,062928
20:39:38	13.196426	-7,71996
20:39:39	5.476466	-3,890963
20:39:40	1.585503	-0,944561
20:39:41	0.640942	-0,230576
20:39:42	0.410366	-0,237425
20:39:43	0.172941	0,600871
20:39:44	0.773812	0,160583
20:39:45	0.934395	0,223408
20:39:46	1.157803	0,41138
20:39:47	1.569183	0,241746
20:39:48	1.810929	0,306639
20:39:49	2.117568	0,230889
20:39:50	2.348457	0,293383
20:39:51	2.641840	1,980009
20:39:52	4.621849	0,753542
20:39:53	5.375391	0,910605
20:39:54	6.285996	6,229098
20:39:55	12.515094	3,904755
20:39:56	16.419849	3,089178
20:39:57	19.509027	3,750922
20:39:58	23.259949	3,098909
20:39:59	26.358858	3,769358

Example 2: long bus stop

Example 3: half stop at traffic light

TIME	SPEED1	ACCELERATION
20:36:16	22.080805	-0,747303
20:36:17	21.333502	-0,397541
20:36:18	20.935961	-1,071703
20:36:19	19.864258	-1,76261
20:36:20	18.101648	-2,136071
20:36:21	15.965577	-0,854741
20:36:22	15.110836	-4,525557
20:36:23	10.585279	-2,947873
20:36:24	7.637406	-0,263888
20:36:25	7.373518	0,873946
20:36:26	8.247464	1,800669
20:36:27	10.048133	1,270312
20:36:28	11.318445	2,675312
20:36:29	13.993757	4,130249
20:36:30	18.124006	3,383633
20:36:31	21.507639	1,433281
20:36:32	22.940920	4,13179
20:36:33	27.072710	2,621559

TIME	SPEED1	ACCELERATION
20:37:25	44.359097	-0,701787
20:37:26	43.657310	-5,182735
20:37:27	38.474575	-4,895702
20:37:28	33.578873	-0,827641
20:37:29	32.751232	-1,81526
20:37:30	30.935972	-4,553827
20:37:31	26.382145	1,838446
20:37:32	28.220591	5,313242
20:37:33	33.533833	4,141693
20:37:34	37.675526	0,909534
20:37:35	38.585060	-1,640045
20:37:36	36.945015	-2,798332
20:37:37	34.146683	-5,048893
20:37:38	29.097790	-7,987498
20:37:39	21.110292	-5,288582
20:37:40	15.821710	-6,224699
20:37:41	9.597011	-5,333691
20:37:42	4.263320	-2,674621
20:37:43	1.588699	-0,986136
20:37:44	0.602563	-0,018865
20:37:45	0.583698	0,255743
20:37:46	0.839441	-0,041927
20:37:47	0.797514	0,226673
20:37:48	1.024187	-0,076959
20:37:49	0.947228	0,020342
20:37:50	0.967570	-0,312973
20:37:51	0.654597	-0,344719
20:37:52	0.309878	-0,28423
20:37:53	0.025648	0,080503
20:37:54	0.106151	0,062436
20:37:55	0.168587	0,199811
......	......	......
20:38:20	0.574418	0,8216
20:38:21	1.396018	1,726543
20:38:22	3.122561	1,700953
20:38:23	4.823514	2,056159
20:38:24	6.879673	5,812382
20:38:25	12.692055	3,711553
20:38:26	16.403608	3,070063
20:38:27	19.473671	4,279207
20:38:28	23.752878	1,741837
20:38:29	25.494715	-1,520943
20:38:30	23.973772	-1,254824
20:38:31	22.718948	5,076065
20:38:32	27.795013	6,508988
20:38:33	34.304001	6,042736
20:38:34	40.346737	-1,057419

Example 4: long traffic light

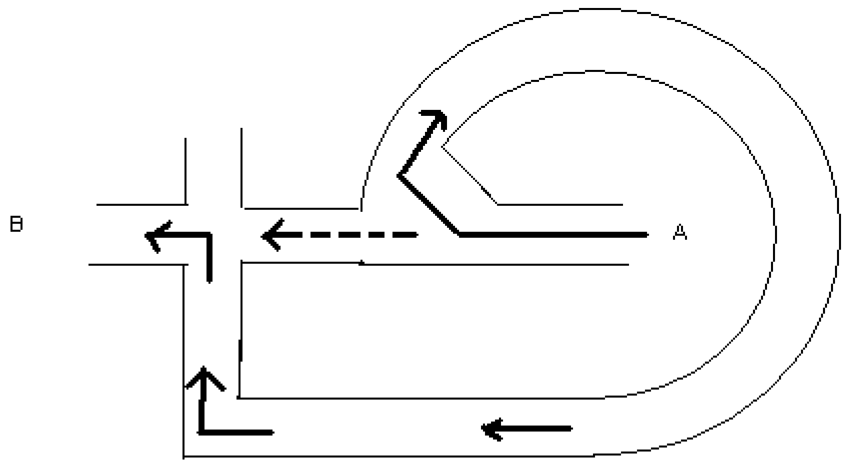
TIME	SPEED1	ACCELERATION
20:40:05	30.607738	-2,485801
20:40:06	28.121937	-2,996979
20:40:07	25.124958	-6,427269
20:40:08	18.697689	-6,691233
20:40:09	12.006456	-4,490141
20:40:10	7.516315	-1,761307
20:40:11	5.755008	-3,356939
20:40:12	2.398069	-1,025708
20:40:13	1.372361	-0,877245
20:40:14	0.495116	-0,236787
20:40:15	0.258329	0,045313
20:40:16	0.303642	0,202803
20:40:17	0.506445	-0,042302
20:40:18	0.464143	-0,021745
20:40:19	0.442398	0,129636
20:40:20	0.572034	0,112064
20:40:21	0.684098	-0,401381
20:40:22	0.282717	0,077884
20:40:23	0.360601	0,299342
20:40:24	0.659943	0,282787
20:40:25	0.942730	0,045163
20:40:26	0.987893	0,040877
20:40:27	1.028770	0,126144
20:40:28	1.154914	-0,129083
20:40:29	1.025831	-0,448676
20:40:30	0.577155	0,104106
20:40:31	0.681261	-0,332021
20:40:32	0.349240	0,011019
20:40:33	0.360259	-0,173451
20:40:34	0.186808	0,061043
20:40:35	0.247851	0,199193
20:40:36	0.447044	-0,101523
20:40:37	0.345521	0,228589
20:40:38	0.574110	-0,119523
20:40:39	0.454587	0,176381
20:40:40	0.630968	1,036612
20:40:41	1.667580	1,286911
20:40:42	2.954491	2,006548
20:40:43	4.961039	1,822438
20:40:44	6.783477	6,738849
20:40:45	13.522326	4,184566
20:40:46	17.706892	5,349318
20:40:47	23.056210	4,516491
20:40:48	27.572701	3,417753
20:40:49	30.990454	3,64768
20:40:50	34.638134	-0,473961
20:40:51	34.164173	1,812626

Example 5: short traffic light

TIME	SPEED1	ACCELERATION
20:26:56	36.142662	0,896446
20:26:57	37.039108	-0,514297
20:26:58	36.524811	-1,995281
20:26:59	34.529530	-2,138955
20:27:00	32.390575	-3,301855
20:27:01	29.088720	-3,2134
20:27:02	25.875320	-2,094196
20:27:03	23.781124	-3,625887
20:27:04	20.155237	-2,435312
20:27:05	17.719925	-2,197376
20:27:06	15.522549	-1,432766
20:27:07	14.089783	-0,376953
20:27:08	13.712830	-3,903974
20:27:09	9.808856	-2,620983
20:27:10	7.187873	-2,749355
20:27:11	4.438518	-1,207016
20:27:12	3.231502	-0,459602
20:27:13	2.771900	0,107173
20:27:14	2.879073	0,11348
20:27:15	2.992553	-0,470443
20:27:16	2.522110	0,093486
20:27:17	2.615596	0,355537
20:27:18	2.971133	0,142294
20:27:19	3.113427	0,576941
20:27:20	3.690368	1,080844
20:27:21	4.771212	2,042768
20:27:22	6.813980	2,257152
20:27:23	9.071132	3,07037
20:27:24	12.141502	3,57614
20:27:25	15.717642	4,453701
20:27:26	20.171343	3,544912
20:27:27	23.716255	1,636692
20:27:28	25.352947	-0,052274

Back to text: Distinguish

10.10 Appendix J: Extra route section line 456



Back to text: Delays