UNIVERSITY OF NAIROBI
SCHOOL OF COMPUTING AND INFORMATICS

Real Time Traffic Monitoring Using On-board GPS Data

BY

Jared Okoyo Onsomu
P58/76479/2012

Supervisor
Prof. Peter Waiganjo

September, 2014.

Project submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science of the University of Nairobi

Declaration

This project, as presented in this report, is my original work and has not been presented for any other university award. Materials of work done by other researchers are mentioned by clear references.

Signature:____________________________ **Date:_____________________________**

Mr. Onsomu Jared Okoyo

P58/76479/2012

The Project has been submitted in partial fulfillment of the Requirements for the Degree of Master of Science in Computer Science at the University of Nairobi with my approval as the University Supervisor.

Signature:____________________________ **Date:_____________________________**

Prof. Waiganjo Peter Wagacha

School of Computing and Informatics

Acknowledgement

I wish to sincerely thank the following people for their contributions towards the successful completion of this project.

First, the **Almighty God** for the strength and good health He granted me throughout the duration of the project.

I am greatly indebted to my supervisor **Prof. Peter Waiganjo** for his good guidance and direction, advice, strong support, highly-valued criticisms, deep insights and for providing resourceful materials that contributed to the quality of this project.

For praying for me and supporting all my endeavors, I give all my appreciation to my late Dad **Mr James Onsomu** and Mum **Mrs. Rusaria Onsomu**.

May God bless you all!

Abstract

The high levels of congestions in the city of Nairobi have exceeded the slow increments in transportation infrastructure supply in many. This dramatic increase in traffic volume is causing various social, environmental and economic problems. In dense urban areas, adding capacity through construction of new or expanding existing infrastructure is difficult due to lack of space and prohibitive costs. A more viable approach to cope with the congestion problem is to monitor traffic congestion, understand the causes of its formation and development, and use the aforementioned knowledge in traffic management systems and transportation planning to mitigate traffic congestion.

To study the congestion problems, a sample transit route was used (46 Kawangware). Four Matatus plying this route were fitted with GPS gadgets for traffic data collection (GPS position, speed and timestamp). A System was developed to receive and process traffic data in real time. The developed System provided interfaces to monitor average speed and travel time for the whole route in real time; hence one is able to gauge overall congestion levels at any given time. The algorithm for calculating average speeds was developed in such away to take care of the missing data as the gadgets used to collect data could not cover the whole route at any given time. Collected data was analyzed to identify congestion hotspots and traffic patterns of the route.

The results obtained from the study indicate that weekends experience less congestion than the rest of the weekdays. Tuesdays reflect highest level of congestion with an average travel time of 1 hour for distance of 10.39 km, while during weekends the average travel time is 25 minutes. The results also show an upward congestion levels trend between Monday and Friday. This reflects the rhythm of city life: traffic congestion on Fridays is typically higher than that on Mondays. Finally the study identified congestion hotspots for route under study which include route segment between Olendume to Yaya, Hurligham, Kenyatta Avenue to NHIF and Haile Selassie.

Table of Contents

List of Figures

List of Tables

List of Acronyms

GPS – Global Positioning System

AVL – Automatic Vehicle Locator

GRP – Gross Regional Product

GSM - Global System for Mobile

GPRS - General Packet Radio Service

Chapter One

1.1 Introduction

Traffic congestion continues to be one of the major problems in various cities in the world. The dramatic increase in traffic volume in Nairobi is leading to massive congestions causing various social, environmental and economic problems. Congestion are often caused by or made worse by traffic incidents. A sudden traffic surge immediately after special events or some special location (‘Hot Spots’) in the city can create substantial traffic congestion in the area related with these Hot Spots.

Traffic congestion provides real challenges to the mobility and safety of our daily travel. Congestion may be alleviated by providing timely and accurate traffic information so that motorists can avoid congested routes by using alternative routes or changing their departure times. In general, the public tends to think more in terms of travel time rather than volume in evaluating the quality of their trips. Accordingly, hot spot information is a key element for avoiding congestion as well as for making informed driving route decision (Shoaib Kmaran, 2007).

IBM conducted a Survey to gauge drivers’ perception of how traffic affects them based on factors such as stress, anger, health, and performance at work or school. Approximately 8,000 drivers, from 20 cities around the world, were surveyed in their native languages with Nairobi being one of them. The result indicate Mexico City showed up as the ‘most painful’ city for commuting, while Montreal, London, and Chicago came out the ‘best’. In Nairobi, 35% of drivers reported that they have spent three hours or more in traffic, and in Moscow, over 45%. 91% of all respondents found themselves stuck in traffic with the maximum delay reported to be about 1.3 hours when averaged across all cities. The results of the survey were compiled into Commuter Pain Index, which ranks the emotional and economic toll of commuting in each city as shown in the diagram below (Gyimesi, et al., 2011).

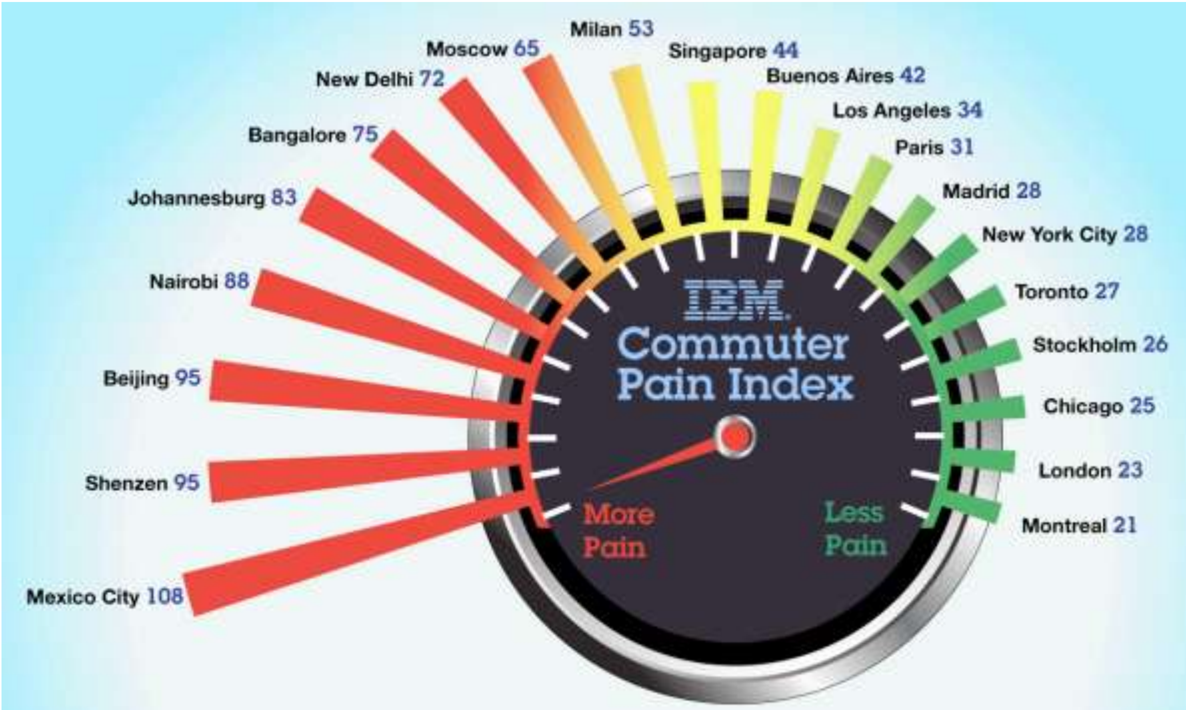


Diagram 1 Commuter Pain Index (Source: IBM 2011 Global Commuter Pain Survey)

Economics of Traffic Congestion

The impact of severe traffic congestion on regional economies is significant. Removing today's gridlock by building infrastructure and adding capacity to create free-flow traffic conditions throughout a region could boost productivity for workers by as much as 30 percent in highly congested areas.

The tables below show the effect of congestion on gross regional product (GRP) in eight cities in America and show the economic benefits cities can expect from making the transportation investments needed to solve congestion (Hartgen & Fields, 2009).

	2003 Urban Area		Expected 2030 Urban Area		Cost to Remove LOS F Congestion by 2030, ($B)
Region	Pop, K	Travel Time Index (TTI)	Pop, K	TTI	
Charlotte	725	1.31	1,185	1.62	$2.9
Salt Lake City	877	1.28	1,251	1.59	$1.2
Denver	2,050	1.40	3,210	1.80	$10.0
Atlanta	2,924	1.46	5,009	1.85	$13.1
Seattle	2,946	1.38	3,963	1.79	$4.8
Detroit	3,939	1.38	4,277	1.50	$24.1
San Francisco	4,120	1.54	4,968	1.86	$29.2
Dallas	4,312	1.35	7,014	1.73	$26.4

.

Table 1 ***Summary of Regional Traffic and Congestion Statistics (source: Hartgen & Fields, 2009)***

Urbanized Area	Location with Largest Benefit from Congestion Reduction	Potential Annual Increase in GRP With Free-Flow Traffic Conditions Around Location ($B)	Potential Increase in Tax Revenues Over 20 Years as a Result of Free-Flow Conditions and Increased GRP ($B)	Estimated Cost to Remove Severe Regional Traffic Congestion by 2030 ($B)
Atlanta	Major Suburb	$15.4	$21.5	$13-15
Charlotte	Major Suburb	$22.5	$31.5	$3-5
Dallas	Major University	$46.0	$64.4	$26-30
Denver	Major Mall	$38.5	$53.8	$10-15
Detroit	Major Suburb	$7.3	$10.3	$25-30
Salt Lake City	Major Suburb	$0.7	$1.0	$1-2
San Francisco	Major Mall	$10.7	$15.0	$29-40
Seattle	Major Suburb	$13.4	$18.8	$5-10

Table 2 ***Sample of Increased Economic Productivity and Tax Revenues with Free-Flow Conditions (source: Hartgen & Fields, 2009)***

Traffic congestion and air pollution

In most cities in the world, the traffic congestion is the largest contributor of urban air pollutants, as well as to high levels of carbon monoxide and hydrocarbons, among other substances. These high levels contribute to various respiratory and cardiovascular illnesses. Various epidemiological studies have clearly linked transport-related contaminants to asthma, bronchitis, heart attacks, and strokes. As shown in Table 3, even though the level of different air pollutants in cities of developing cities are generally higher than those of developed ones, still very few cities overall stay below the recommended levels by the World Health Organization (WHO) .

Air Pollution in Selected Cities:

	PM10 (µg/m3)	SO2 (µg/m3)	NO2 (µg/m3)
WHO Air Quality Guidelines:	**20**	**20**	**40**
Beijing	89	90	122
Delhi	150	24	41
Tokyo	40	18	68
Seoul	41	44	60
Mexico City	51	74	130
Bangkok	79	11	23
London	21	25	77
New York	21	26	79
Paris	11	14	57
Shanghai	73	53	73
Santiago	61	29	81
Sao Paulo	40	43	83

Table 3 ***Air Pollution in Selected Cities (Data Source: World Bank, 2009)***

Approximately 15% of Nairobi's households own cars. Research done by Salon and Eric (2012) shows that without policies that make non-motorized transport safer and public transport service better, car ownership and use will increase sharply as the city's residents become wealthier, further congesting already-overloaded roadways.

Timely and accurate detection of congestion hot-spots is an essential part of any successful advanced traffic management system. If any Hot Spots are detected quickly, they can be cleared swiftly, resulting in less traffic congestion, and then more appropriate traffic management strategies can be applied to provide better service to the road users. So it is desired that the Hot-Spots related traffic performance to be measured so that the traffic flow can be improved.

In recent years, we are witnesses of very rapid development of the navigation technology, and especially in area of the satellite navigation.

The satellite navigation Global Position System (GPS) has become very popular, and many people are using this navigation to plan trips and routes. With standard GPS device, we can determine the current speed of the movement and the geographical location. This information among other can be stored in logs. We can later use these logs for processing. The GPS navigation inherits some inaccuracy when determining geographical location. This is because of the obstacles that can block the satellite signal that the GPS device is receiving. In the past, the inaccuracy was about 100 meters, but in nowadays this inaccuracy is about 10-20 meters (Nermin Mudzelet, 2007).

Main advantage of the GPS navigation is it´s accessibility and low cost of implementation. In the past, if we wanted to collect information about the traffic condition we needed to install and maintain expensive equipment into roads such as loop detectors.

1.2 Problem Statement

Traffic congestion is an on-going and old problem that is not showing any signs of improvement in cities. It has detrimental impacts on economy, health and environment as evidenced by delays, which may result in late arrival for employment, meetings, and education, resulting in lost business, disciplinary action or other personal losses. Wasted fuel which increases air pollution and carbon dioxide emissions contributing to global warming, health and environmental problems owing to increased idling. Traffic jam comes with stress and frustrations which may lead to accidents. Finally, Blocked traffic may interfere with the passage of emergency vehicles traveling to their destinations where they are urgently needed.

Building new roads to address congestion problem is not the only solution other methods using current technology are needed to better manage traffic. Technology will provide a platform for visualization of historical and real time traffic data which is currently missing. The provision of accurate travel time information to the public is critical as they can make informed decision in terms of which alternatives routes to use and what is most appropriate day of the week and time to travel.

1.3 Research Objectives

The objectives of this research are as follows:

1. Build real time map-matching GPS data system for traffic monitoring
2. Determine vehicle average movement speeds on different road segments
3. Estimate travel time from one point to another
4. Analyze traffic flow patterns based on temporal and spatial correlations to identify hot spot areas

1.4 Research Questions

The following are the research questions:

1. How to determine vehicle speeds from Matatu GPS device data?
2. What algorithm can be used to compute average speeds and travel time?
3. What methods can be used to analyze and visualize vehicle traffic patterns?

1.5 Justification

Traffic Monitor System provides both real-time and historical data about the traffic conditions prevailing on the road network. This information is be used for quick reaction measures, such as, advising commuters to take alternative routes. In the long run, however, this information can be used to plan a better road network by identifying areas of frequent congestion and building alternative routes and also can guide in policy review, formulation and implementation processes geared towards reduction of congestion.

1.6 Scope of work

The study concentrated on providing real time average speed and travel time along Kawangware town route and provided interfaces for traffic congestion analysis, as the prior knowledge of these areas is very important for better traffic management strategies. GPS data was used for this purpose.

1.7 Definition of Terms

Hot-Spots In any city hot spots are the locations or those areas of the city, which are regularly congested and where the flow of traffic is very slow that is traffic is moving very slowly or traffic is completely stopped. These areas could increase the severity of congestion. Hot spots can be of two types; regularly and occasionally.

GRP: is the sum of the value of all goods and services produced annually by a state or urban region.

GNP: is sum of the value of all goods and services produced annually in the national economy.

Free-flow speeds: is the term used to describe the average speed that a motorist would travel if there were no congestion or other adverse conditions (such as bad weather).

Traffic Congestion: is a condition on road networks that occurs as use increases, and is characterized by slower speeds, longer trip times.

1.8 Assumptions

The following assumptions were considered:

1. Internet connection to the GPS server will be available.
2. The GSM network will be stable and always available.
3. Real-time tracking will be a few minutes behind the normal global time.
4. GPS General Packet Radio Service (GPRS) messages are relayed almost immediately and that there is no big lag.

Chapter Two

2 Literature Review

In this chapter, focus on previous work related to this project.

2.1 Automatic Vehicle Location (AVL) Systems

Automatic Vehicle Location (AVL) Systems are computer-based vehicle tracking systems. They are also referred to as Automatic Vehicle Monitoring (AVM) Systems or Automatic Vehicle Location and Control (AVLC) Systems (Casey & Labell, 1998). They are used in transit, trucking fleets, police cars, ambulances, and for military purposes, and their use in transit continues to grow (Casey & Labell, 1998).

Figure 1below shows the schematic display of an AVL (Casey, et al., 2000).

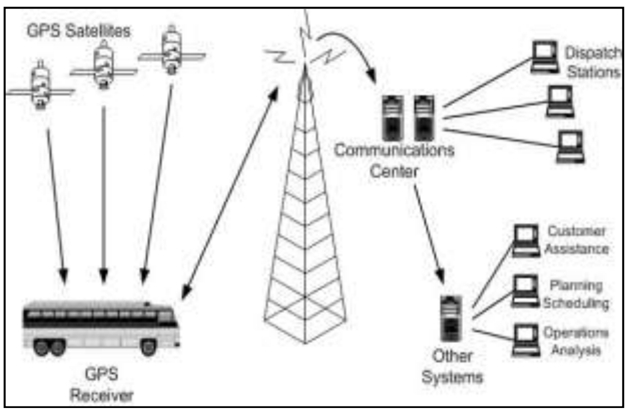


Figure 1 Schematic Display of an AVL System (Source: Casey, et al.,

2.1.1 Benefits of AVL Systems

The benefits of AVL systems are as follows:

1. They collect real-time information which can be provided to the public,
2. They improve schedule reliability
3. They reduce operating and maintenance costs
4. They improve service efficiency
5. They enhance safety and security

The benefits of AVL systems have been well documented. Schedule adherence improved by 23% in Baltimore, 12.5% in Kansas City, 8.5% in Hamilton, Ontario, and 4.4% in Milwaukee after AVL installation (Johnson & Thomas, 2000). Operating costs were reduced by $500,000 per year in Kansas City, and $45,000 per year in London, Ontario (Transit, 1996). Paratransit ridership increased by 17.5% and paratransit passenger waiting time decreased by 50% in Winston-Salem, North Carolina (Transit, 1996).

2.1.2 Uses of AVL Systems

The first use of AVL technology in transit was in London, England in the late 1950s, and the first use in the United States was in Chicago in the late 1960s (Turnbull, 1993) A number of transit systems in North America and abroad began to plan for and implement AVL systems during the 1980s (Turnbull, 1993). According to the research of Casey in 2002, 322 transit agencies are operating, implementing, or testing AVL systems (Casey, 2002). The number of transit agencies used AVL systems increased by four hundred percent as compared to earlier studies in 1995 (Casey, 2002). An increasing number of transit agencies are planning to install AVL systems because the cost of AVL systems has rapidly dropped (Turnbull, 1993).

2.1.3 Technologies of AVL Systems

AVL systems consist of two technologies, location technology and data transmission technology. Location technology is used to measure actual real-time position of each vehicle, and data transmission technology is used to relay the information to a central location (Casey & Labell, 1998).

2.1.3.1 Location Technology

Location technology includes dead-reckoning, signpost and odometer, global positioning systems (GPS), differential GPS (DGPS), etc. For transit vehicle, a single location technology is usually insufficient for determining position. For instance, tall buildings block the signals and result in multi-path errors. Therefore, the primary location technology is supplemented with another location technology (Casey & Labell, 1998).

Dead-reckoning is the most self-determined form of location technology. The transit vehicle determines its own location without the help of external technologies. First, the transit vehicle is told its starting point. The vehicle measures the traveled distance from the starting point by reading the odometer. Then the vehicle determines the traveled direction by compass headings. Dead-reckoning location technology is seldom used by itself because the equipment has to be reset frequently from a known location. Dead-reckoning is usually supplemented by one of other location technologies like signpost or GPS (Casey &

Labell, 1998). It is relatively inexpensive, but the accuracy degrades with distance traveled (Okunieff, 1997).

Signpost and odometer uses a series of radio beacons placed along the bus routes. The beacons send a low power signal and the signal is detected by a receiver on the transit vehicle. Then the transit vehicle reports its position to dispatch according to the traveled distance, which is taken from the odometer (Casey & Labell, 1998). This technology requires low in-vehicle cost and it is well established and proven. However, additional installation is required with route changes and it cannot track the vehicle when a bus is off-route (Okunieff, 1997).

Global positioning systems (GPS) determine the position using the signals which are transmitted from up to 24 satellites. GPS works anywhere the satellites reach, and it is much more robust than other location technologies. However, satellite signals do not reach underground and they can be interrupted by tall buildings or foliage. Where these problems happen, signpost and odometer can supplement the GPS (Casey & Labell, 1998).

The U.S. Department of Defense (DOD) intentionally degraded the accuracy of GPS for safety reason. To correct this interruption, an additional (differential) correction was added and this system is called Differential GPS (DGPS) (Casey & Labell, 1998).

2.1.3.2 Data Transmission Technology

Position information, regardless of which location technology is adopted, is usually stored on the transit vehicle for some period of time. The information is relayed to the dispatch center in raw form or is processed on-board the vehicle. The two most common data transmission technology are polling and exception reporting (Johnson & Thomas, 2000).

With polling technology, the computer at the dispatch center asks each bus for its location at regular intervals. The accuracy of location is a function of how often the transit vehicle is polled. In addition, the number of radio frequencies which are available in urban areas is limited. Due to this reason, many transit agencies have chosen another technology, exception reporting (Casey & Labell, 1998).

Under the exception reporting method, each bus reports its location at a few specified locations or when the bus is found to be off-schedule by some pre-defined tolerance. Exception reporting requires each transit vehicle to know not only its position but also its scheduled position. Many agencies combine the polling and exception reporting methods (Johnson & Thomas, 2000).

2.2 Global Positioning Systems (GPS)

GPS is a satellite-based navigation system which is funded and controlled by the U.S. Department of Defense (DOD) (Okunieff, 1997). Even though it was intended for military use, the system has been available for civilian applications world-wide since the 1980s (Dana, n.d.). The GPS consists of 24 satellites (see figure 2) and transmits the estimated position, velocity, and current time to GPS receivers. To compute position, velocity, and current time, signals from at least four satellites are used.

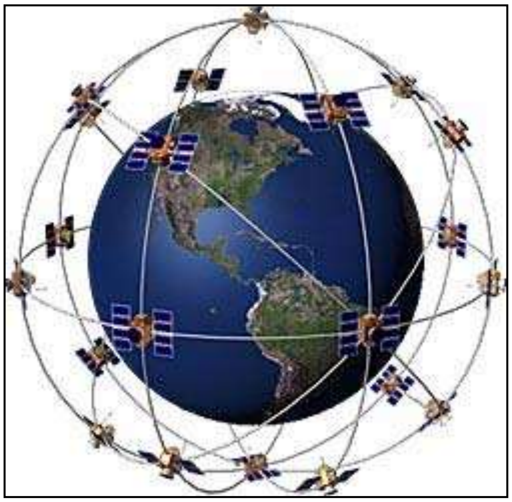


Figure 2Twenty Four Satellites of GPS (Source: Okunieff, 1997)

2.2.1 Accuracy of GPS

GPS has two positioning services, Precise Positioning Service (PPS) and Standard Positioning Service (SPS). PPS is used by authorized users such as U.S. and Allied military while SPS is used by civilian users worldwide (Dana, n.d.). For security reasons the DOD intentionally degraded SPS accuracy. The accuracy of PPS was within 22 meters, and the accuracy of SPS was within 100 meters (Dana, n.d.). To improve the accuracy of SPS, an additional correction (differential) signal was added, and is called Differential GPS (DGPS) (Center, 2014). The accuracy of DGPS was better than 10 meters (Okunieff, 1997).

The SPS accuracy was dramatically improved when the US military removed the intentional degradation to the signal on May 1, 2000 (Center, 2014). Currently the accuracy of PPS and SPS are the same. The current accuracy of GPS is between 10 and 20 meters, and that of DGPS is between 3 and 5 meters (Johnson & Thomas, 2000).

2.2.2 Use of GPS in Transportation

While traditional methods of data collection techniques in transportation are burdensome, time consuming, and error prone, GPS provides better accuracy, consistency, automation, and easier integration between collected data and the data based on GIS (Quiroga & Bullock, 199). Because of the advantages of GPS, a number of studies on data collection using GPS have been conducted (Quiroga & Bullock, 199). They used GPS to collect travel time, speed, route choice, and travel surveys. They have shown that the use of GPS for collecting data is easier and more accurate than traditional methods (Quiroga & Bullock, 199).

2.3 Travel Time Prediction Models

The accurate prediction of link travel time is critical to ITS transit applications. With the development of Advanced Travelers Information Systems (ATIS), the importance of the short-term travel time prediction has increased markedly (Chien & Kuchipudi, 2002). A number of prediction models, including historical data based models, regression models, time series models and neural network models, have been developed over the years by various transit agencies.

2.3.1 Historical Data Based Models

Historical data based models predict travel time for a given time period using the average travel time for the same time period obtained from a historical data base. These models assume that traffic patterns are cyclical and the ratio of the historical travel time on a specific link to the current travel time reported in real-time will remain constant (Smith & Demetsky, 1995). The procedure requires an extensive set of historical data and it is difficult to install the system in a new setting (Smith & Demetsky, 1995). Real-time models assume that the most recently observed transit travel times will stay consistently into the future. Chen et al developed a prediction algorithm that combined these two approaches. First a historical data base was used to obtain estimated travel time. This time was subsequently adjusted as real-time location data are obtained.

2.3.2 Regression Models

Regression models are conventional approaches for predicting travel time (Abdelfattah & Khan, 1998). Regression models predict a dependent variable with a mathematical function formed by a set of independent variables. To establish a regression model, the dependent variables need to be independent. This requirement limits the applicability of the regression model to the transportation areas because variables in transportation systems are highly inter-correlated.

2.3.3 Time Series Models

Time series models assume that the historical traffic patterns will remain the same in the future. The accuracy of Time series models is a function of the similarity between the real-time and historical traffic patterns (Chien, et al., 2002). Variation in historical data or changes in the relationship between historical data and real-time data could significantly cause inaccuracy in the prediction results (D'Angelo, et al., 1999). D'Angelo used a non-linear time series model to predict a corridor travel time on a highway (D'Angelo, et al., 1999). He compared two cases: the first model used only speed data as a variable, while the second model used speed, occupancy, and volume data to predict travel time. It was found that the single variable model using speed was better than the multivariable prediction model.

2.3.4 Kalman Filtering Models

Kalman filtering models have been used extensively in travel time prediction research (Chien, et al., 2002). Chen and Chien (Chien & Kuchipudi, 2002) and Wall and Dailey (Wall & Dailey, 1999) used Kalman filtering techniques to predict auto travel time. The Kalman filtering model has the potential to adapt to traffic fluctuation with time-dependent parameters (Chien, et al., 2002). These models are effective in predicting travel time one or two time periods ahead, but they deteriorate with multiple time steps (Park & Rilett, 1999). Park and Rilett compared neural network models with other prediction models including Kalman filtering techniques to predict freeway link travel time. While the average mean absolute percentage error (MAPE) of neural network models changed from 8.7 for one time period to 16.1 for 5 time periods, that of Kalman filtering changed from 5.7 to 20.1 (Park & Rilett, 1999).

2.4 Existing Solutions

Several solutions have been proposed to monitor and reduce the traffic jam problem, some of which are AVL based and others rely on other advanced technologies. In this section, we summarize some of the existing work and compare it against the deigned model.

2.4.1 Access Kenya

The Access Kenya launched a free online traffic camera service onto their website portal. The update service consists of Internet Protocol (IP) cameras that capture traffic movement and transmit visuals of the same via the Access Kenya website portal, live and in real time.

The Group has installed 23 such cameras within the Nairobi's Central Business District and its environs as well as 2 cameras in Mombasa.

Weakness

This solution does not give numeric speed values, hence making it difficult to analyze data and the cost of installing and maintaining IP cameras is high.

2.4.2 Ma3Route

Ma3Route is a mobile/web/SMS platform that crowd-sources for transport data and provides users with information on traffic, Matatu directions and driving reports. The Application gives users live update on what is happening on major roads in Nairobi. The updates rely mainly on live feed sent by end users either via the App or Twitter. Traffic updates can be viewed as either text feed or via Google maps integrated in the Application.

Weakness

Traffic updates provided by the application relies on the goodwill of people to post traffic situations on road. The updates are provided in different data formats thus it is difficult to automatically analyze the data.

2.4.3 The Sydney Co-ordinated Adaptive Traffic System (SCATS)

SCATS is one of the most widely used Adaptive Traffic System in Australia and the world. It was developed by the Roads and Traffic Authority (RTA) of New South Wales, Australia in the 1970s (Zhao & Tian, 2012). As of November 2011, more than 3,700 traffic lights within New South Wales were connected, monitored and controlled using the SCATS network (Roads & Maritime Services, 2011). The control system involves the use of inductive loops beneath the surface of the road immediately before the intersection stop line of a road. The induction loops are used to detect the presence of a vehicle in addition to measuring the degree of saturation and traffic flow over a set cycle (Samadi, et al., 2012).

The data collected via the induction loops is gathered within the local controller situated at each intersection which is then transmitted to a regional computer. The data is then analyzed and assessed by the regional computer in order to calculate the most appropriate cycle lengths, splits and offsets for the network of local controllers within vicinity of the regional computer. The signal timings are then re-transmitted back to the local controllers to implement the appropriate light changes at the series of intersections (Dineen & Cahill, 2001). Figure 3 illustrates the general physical architecture of SCATS.

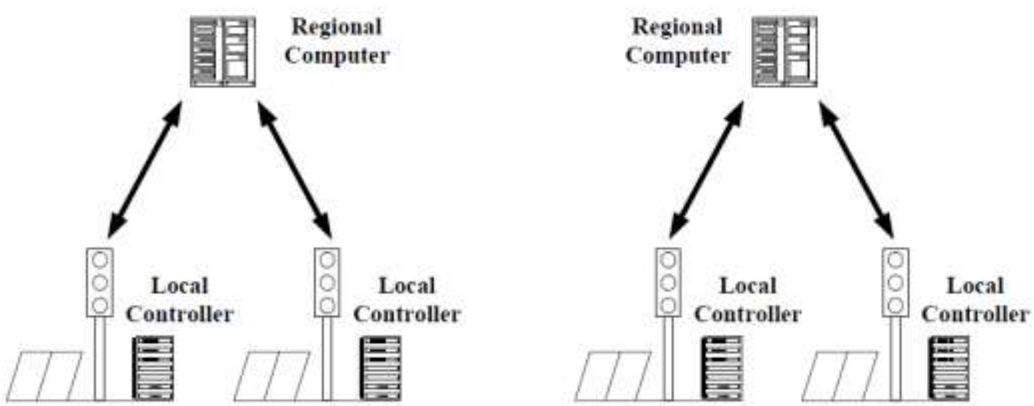


Figure 3 The Physical Architecture of SCATS (Source: Dineen & Cahill, 2001).

Weakness

SCATS relies on stationary monitoring units (induction loop), to gather traffic data. The data gathered by this detector is incomplete and does not provide a clear picture of the nature of traffic. It is impossible to infer from the gathered data any useful information on travel times or the volume of congestion.

2.4.4 GuideME

GuideME, is a system that monitors the highways and calculates the average speed of the passing cars before storing the data in a database. The user can then access real-time updates with exact numerical values of the highways' average speed via a web portal or SMS.

The solution uses two Radio Frequency Identification (RFID) readers on each monitored highway, and installs passive RFID tags on the passing cars. The two RFID readers are integrated with a General Packet

Radio Service (GPRS) device that sends data to a web server. The server runs a calculating algorithm and a database to store the data for querying.

2.4.5 Image Processing

Image processing has been used for a long time to monitor traffic (Mo, et al., 2009). In this solution, the data is first provided in a video sequence. The video is then processed to locate the observed vehicles in each frame of the video. Several image processing techniques are applied; such as segmentation, filtering and edge detection. This allocates the vehicle contour within the frame observation zone. To calculate the speed of a vehicle, the conditional centre of the vehicle must be found in the contour area. By labeling and counting the contours, we can calculate the number of moving vehicles in the observed zone; therefore, we can estimate the level of road's congestion in that area.

Weakness

This solution depends solely on implementing a large number of expensive Closed Circuit Television (CCTV) cameras around the city; not to mention the expensive equipment required to process the data and the long time taken to process the frames.

2.4.6 CarTel

CarTel System was developed by Massachusetts Institute of Technology (MIT). The objective of CarTel was to recommend better routes for the drivers. The system created a mobile-sensing network by installing a very small computer device on moving vehicles. The device calculated the speed of its vehicle at certain intervals and sent it via wireless Internet to a database. The web portal provided geo-spatial data visualization maps that represented the data received from the sensors on the cars.

Weakness

CarTel depends on Wi-Fi networks that the vehicles pass by to send the data. This method cannot work in Nairobi as number of free Wi-Fi spots is not many. Our proposed solution uses GPRS to send the data; which ensures that data can be virtually sent anytime and anywhere.

2.5 Overall System Architecture

Figure 4 depicts the overall system architecture showing how the system components are interlinked. The System components include the following:

1. Data collection – AVL gadgets fitted on vehicles are used for data collocation.
2. Communication – GPRS is used for communication AVL gadgets and the server that is continuously listening to configure port.
3. Transform and Load data – received raw data is converted into the required format and loaded into the database.
4. Data Analysis – Loaded data is analyzed to identify average travel time from one point to another, congestion hotspots and the general traffic pattern of a particular route

The results obtained from the system have impact on Economy, Social issues, Policy formulation, Air pollution and Road Safety.

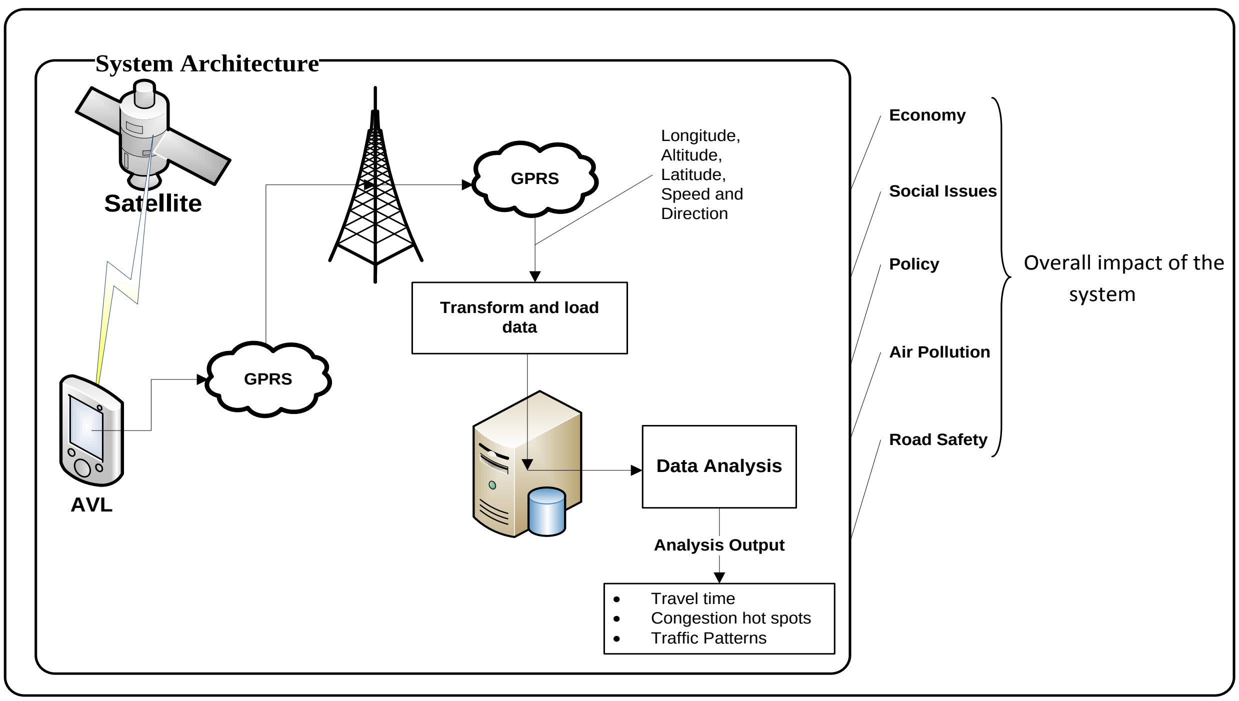


Figure 4 System Architecture (Source: Author)

Chapter Three

3 Methodology

3.1 Introduction

This chapter provides the methodology that was used in system analysis, design, implementation and testing.

3.2 Agile Methodology

The development phase used agile methodology which is a software development methodology typically used in rapid development and implementation. It is an incremental, repetitious means of managing projects; particularly in the field of software development. These iterations give project managers many opportunities to evaluate and change the project during its lifecycle as well as keeping the end user informed and involved in the development.

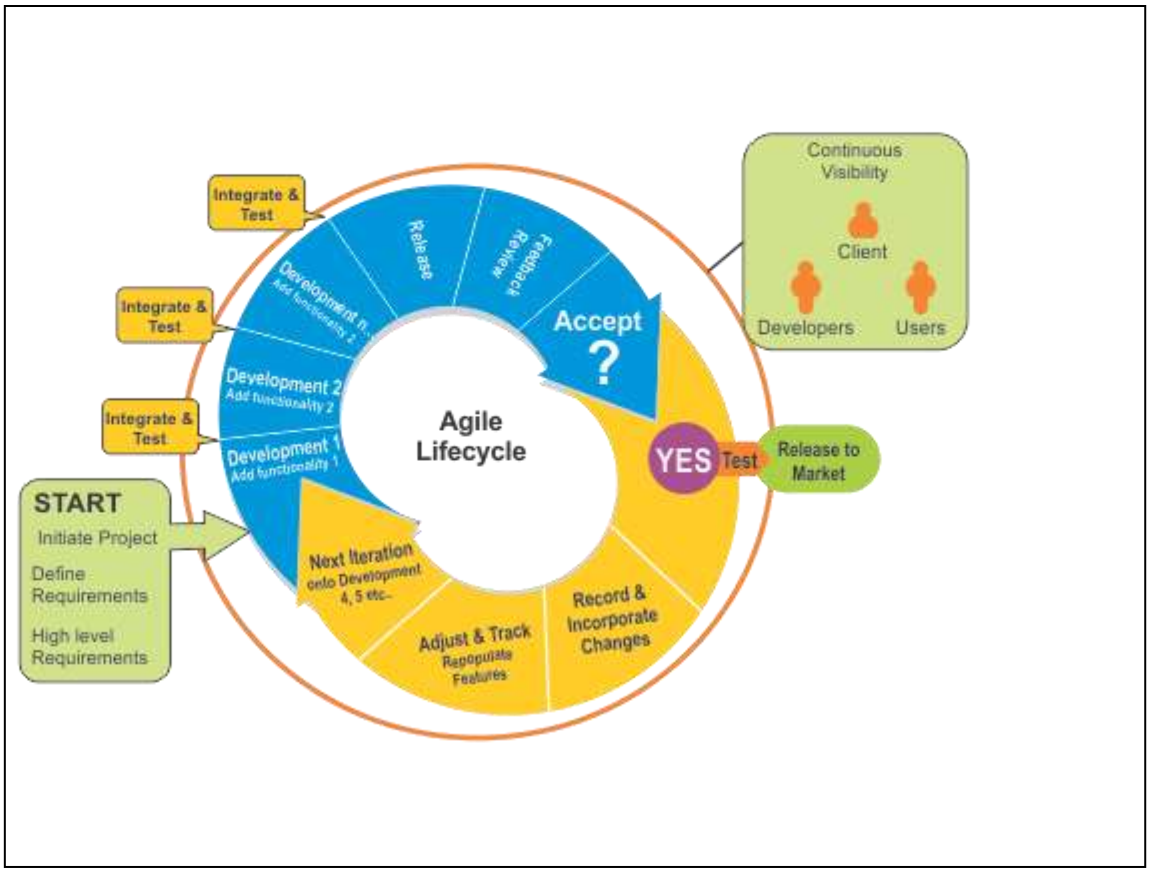


Figure 5: Agile Methodology Diagram (Source: Williams 2007)

3.2.1 Why Agile Methodology

According to Williams (2007), software development often has too much change during the time that the team is developing the product to be considered a defined process. Williams (2007) further states that a set of predefined steps may not lead to a desirable, predictable outcome because software development is a decidedly human activity: requirements change, technology changes, people are added and taken off the team, and so on.

3.3 Data Collection

AVL gadgets were used as the primary data collection tool. The area of interest in this study was route 46 which starts from Railways bus station in town to Kawagware Market. According to Quiroga and Bullock (1999) Analysis, the route was broken down into not more 1.5 KM segment length to better quantify the performance of congestion management measures. Four vehicles were fitted with AVL system for data collection. In this system, GPS receiver was interfaced with a GSM modem and placed in the vehicle. Data was collected for every 10 seconds from 6:00 AM to 8:00 PM for a period of eight weeks. Real time communication of the data was carried out using GPRS. The collected data were stored using Structured Query Language (SQL) database on a central server. The collected data from the AVL units included the ID of the AVL unit, latitude and longitude of the location at which the data entry was made, speed and the corresponding time stamp.

The figure below illustrates diagrammatically how data was collected.

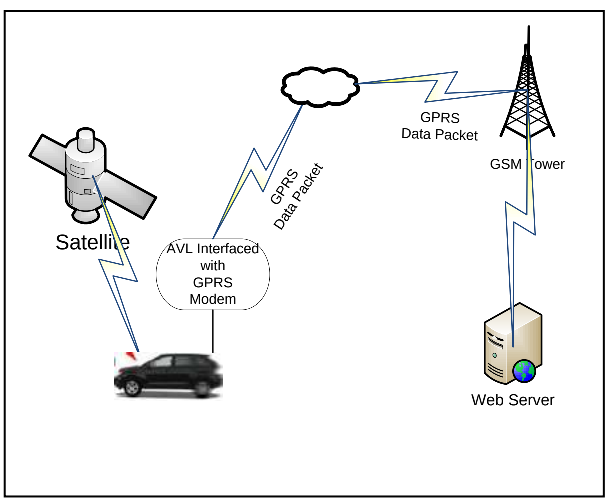


Figure 6: Data Collection Structure (Source: Author)

3.4 AVL Gadget String Format

The AVL trackers sends data as one sentence starting with $$ and ending with ##. Sample data sentences received by the GPS Data Receiver are as follows:

$$1000000001???&A9955&B051304.000,A,0116.8618,S,03645.5808,E,0.00,202.38,280412,,,A*79|0.9|&C0000011111&D0026:164&E10000000&Y00180000##

AVL trackers' manufacturers choose which parts of the NMEA 0183 standard to support. The NMEA 0183 data stream consists of a series of sentences delimited by a newline character. Each sentence begins with a six character identifier, the first character of which is always "$". The NMEA 0183 standard defines dozens of sentences, but only a fraction applies directly to GPS devices. The AVL gadgets used in this project uses the $GPRMC sentence that contains all the information required for our analysis, which include: time, date, position, speed, and unique identifier data.

The table below outlines an interpretation of the sample data sentence:

Data	Description	Parameter
1000000001	Device unique identifier	Maximum of 15 characters
A9955	Device board ID	Manufacturer specific
051304.000	Time	hhmmss.ddd
A	Validity of the fix	A = Valid
0116.8618	Latitude	ddmm.mmmm
S	Latitude hemisphere	S = Southern
03645.5808	Longitude	dddmm.mmmm
E	Longitude hemisphere	E = Eastern
0.00	Speed, knots.	s.s
202.38	Course in degrees	h.h
280412	Date	ddmmyy
'$' ',' '?' '&' and '#'	Separator to the various data parameters	

Table 4: An interpretation of the GPS data sentence

3.5 Data Analysis

3.5.1 Travel Time Calculation

Travel time calculation was based on GPS data collected above which has speed, course, position, time and date of the vehicles movement in the road network. Each captured record was treated as a single GPS observation. In this approach relation among GPS points was not considered, just how each GPS point influence the average speed of the road segment. This approach is referred to as Point-Based approach. The reason for using this approach was because it is very hard to get coverage of all segments in the road network. By coverage, we mean that some segments do not have any GPS observations. If there are no values of the average speed for some segments, average time cannot be calculated. Thus average time from one point to another cannot be calculated if that segment is included in the route under consideration.

The average time for a single segment was calculated with formula: the length of the segment divided by the segment average speed.

The segment length was calculated based on the Spherical Law of Cosines as shown in the formula below:

$$D = 1.852 * 60 * ARCOS (SIN(L1) * SIN(L2) + COS(L1) * COS(L2) * COS(DG))$$

Where:

- ✓ L1 = latitude at the first point (degrees)
- ✓ L2 = latitude at the second point (degrees)
- ✓ G1 = longitude at the first point (degrees)
- ✓ G2 = longitude at the second point (degrees)
- ✓ DG = longitude of the second point minus longitude of the first point (degrees)
- ✓ DL = latitude of the second point minus latitude of the first point (degrees)
- ✓ D = computed distance (km)

Spherical Law of Cosine was used because it incorporates six trigonometric functions in its computation hence results in more accurate distance value the can be as small as one metre.

3.5.2 Hot Spot Identification Algorithm

The following are steps that were used for identification of hot spots:

Step 1:

To analyze the traffic behavior efficiently, the road was segmented and after the segmentation normal average speed was assigned to each road segment depending on the type of the road. According to Joseph, et al., n.d. speed limit of 30km/h will be designated for residential areas and 60km/h and above for major arterial roads.

Step 2:

Calculate the average speed of the vehicles in a certain direction in segment which has most no. of the GPS data. Current_average_speed = (1/ N) ∑ vehicle_speed Where N is the number of vehicle in that segment and summation runs over N.

Step 3:

If Current average speed is less than Normal average speed, for any segment, then mark the segment for further analysis, and move to the next step.

Step 4:

Determine the current average speed of the road segments in front and behind of the marked segment, if the average speed in the 'front' segment is much higher than the 'marked' segment then a blockage within the slowest is likely. Since there are possibilities, that segments could include normal stoppage points such as traffic lights, junctions and roundabouts.

Step 5:

Compare the average speed of the vehicles in the neighboring segment with that of the marked segment. The average speed in segments in front and behind the marked segment is found in order to detect if there is an incident in the slowest marked segment or there is just a general congestion. If there is just a general congestion in the segment (as it consist of a place or is close to a place, where traffic is usually congested, for e.g. traffic lights, junction or level crossing) the average speed will be similar in adjoining segments. However, if there is a hot spot causing a blockage in the marked segment then the average speed of the segments in front of that blocked/marked segment will be higher and with less number of vehicles.

3.6 System Testing

To ensure veracity and completeness of the Real Time Traffic Monitoring System, a number of tests were conducted. Testing was iterative and incremental in nature. Of concern in this testing were both the functional and non-functional requirements which were identified during the analysis stage. The tests were performed on server before deploying the application to production server.

3.6.1 Module Testing

Each module was tested on its own. This was done to ensure that every unit/module works properly on its own.

3.6.2 Regression Testing

As coding progressed, new changes were sometimes incorporated into a module. The new changes and the entire module had to be tested again to ensure that the changes did not affect the functioning of previously verified code.

3.6.3 System Testing

After all the modules were completed, tested and found to function properly, they were integrated. The entire system was then tested to ensure there was cohesiveness and that the system performed according to specification.

Chapter Four

4 Analysis, Design and Implementation

4.1 Requirement Specification and Systems Analysis

4.1.1 Functional Requirements

This requirement defines functions of a given system or its components, describing an activity or process that the system must perform. The following are traffic monitor modules:

i. **GPS data Receiver module** – it is used for data collection, the module continuously listens for client connections read data and save onto a database table.

ii. **Transformation and loading module –** it is used for checking validity of data received, convert data into the required format as follows:

   a. Coordinates from degree minutes, second format to decimal format this is imported as is the required for mapping on Google Map.

   b. Time to UNIX timestamp

   c. Speed from knots to km/hr. 1knot = 1.83km/hr

   Finally load data to the master table.

iii. **Process transformed data** – this module computes the following parameters in real time:

   a. Vehicle movement direction

   b. Vehicle current segment

   c. Over-speeding SMS notification to the administrator

   d. Average speed for each segment and the whole route.

   e. Detect congestions in subsequent segment based on the current vehicle segment and sends traffic jam SMS alert to the drivers.

iv. Reports module – this module generate the following reports:

a. Route hourly average speed per week day

b. Segment hourly average speed per week day

c. Congestion level analysis report

d. Over speeding report per vehicle in a given time range

e. Over speeding report per week day in a given time range

4.1.2 Non-Functional Requirements

A non-functional requirement specifies systems' properties and constraints.

i. An easy to use and navigate Graphical User Interface.

ii. Simple and easy to interpret graphs.

4.2 System Design

4.2.1 Architectural System Design

This depicts from a broader perspective of how the various components of the system are interlinked. The system was designed in a 4-tier architecture consisting of the Data Layer, the Data Transformation Layer, the Business Logic Layer and the User Interface Layer as shown below:

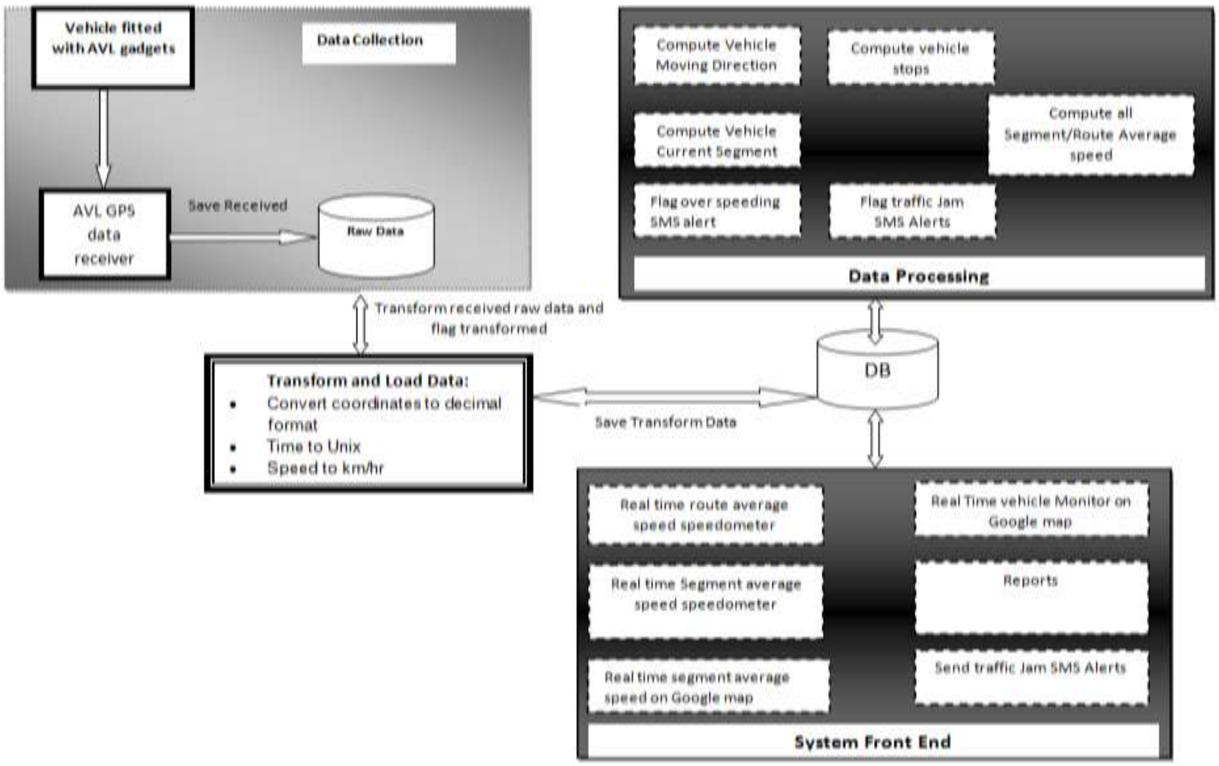


Figure 7: Over-all System Architecture (Source: Author)

4.2.2 System Flowchart Design

A flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analyzing, designing, documenting or managing a process or program.

4.2.2.1 Overall System Flowchart

The overall system flowchart in figure 8 shows the main steps of the system. The first step in the flowchart is data collection from GPS devices amounted on matatus in real time. After data is received it is transformed into various parameters including longitude, latitude, timestamp and speed. Finally transformed data is processed to compute average speeds for various segments/route in both direction i.e from Kawangware terminus to Railways terminus and from Railways terminus to Kawangware terminus directions. This process continues as long as data is being received from the GPS gadgets.

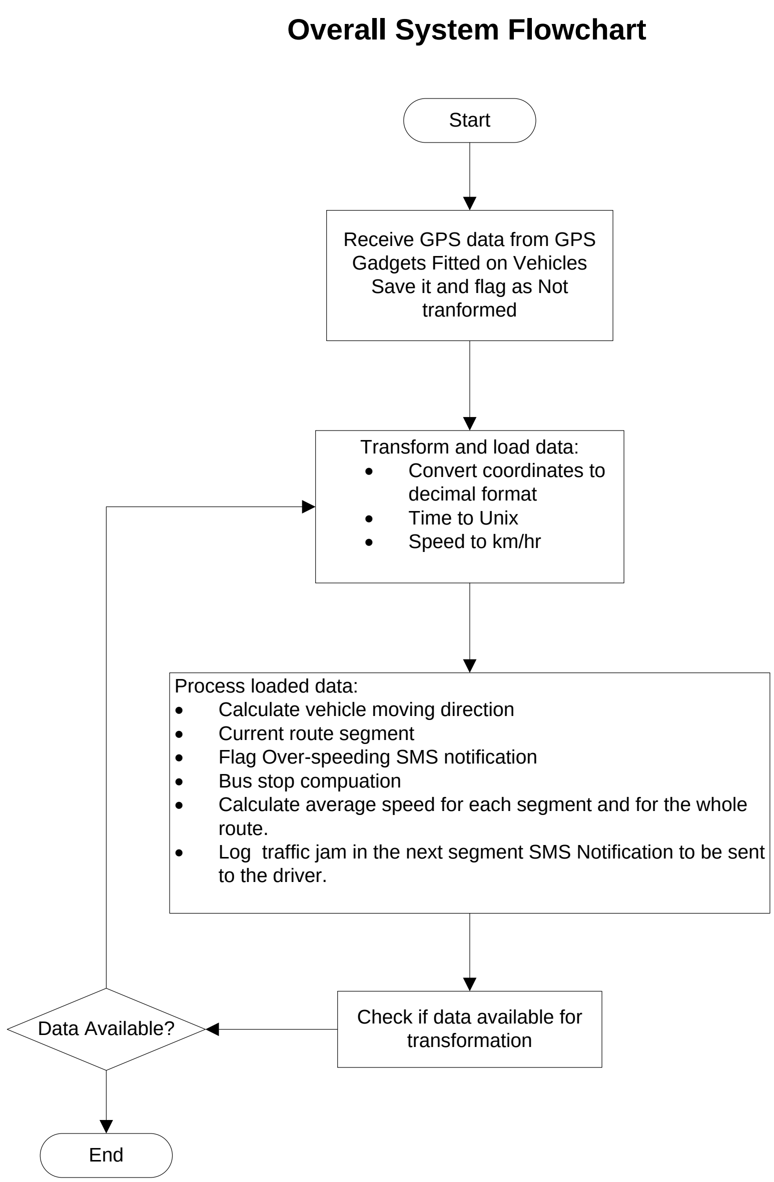


Figure 8: Overall System flowchart (Source: Author)

4.2.2.2 GPRS Data Client Receiver Flow Chart

The flowchart in figure 9, describes the communication between the GPS gadgets to the server over GPRS. The server must first be assigned a static public IP tied to TCP port which is in turn configured on the gadgets. The serves starts receiving upon running the crontab which executes the script that listens to new connections from the gadgets. If the connection is successful the sent raw data is saved into the database.

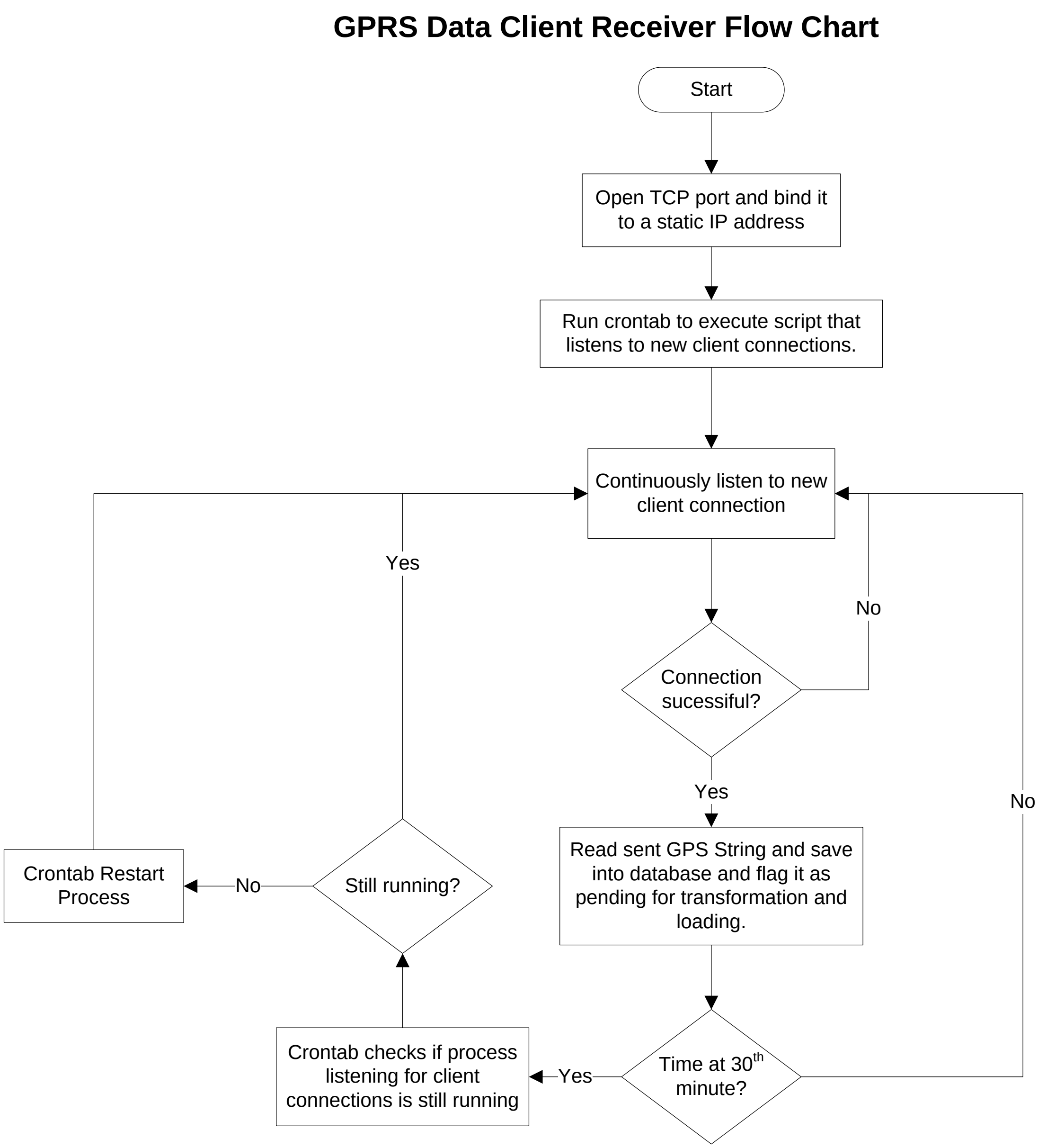


Figure 9: GPRS Data Client Receiver (Source: Author)

4.2.2.3 Transform and load data flowchart

The transform and load data flowchart, describe process used to transform raw string data into various parameters as shown in the figure 10. The transform data, the validity of the string is first checked based on the string length and the validity parameter. If the string is valid, it is then split and transformed into various parameters and format. The parameters transformed include coordinates from degrees minute second to decimal format, timestamp to unix and speed from knots to km/hr. The above process is repeated as long as their exists untransformed raw data.

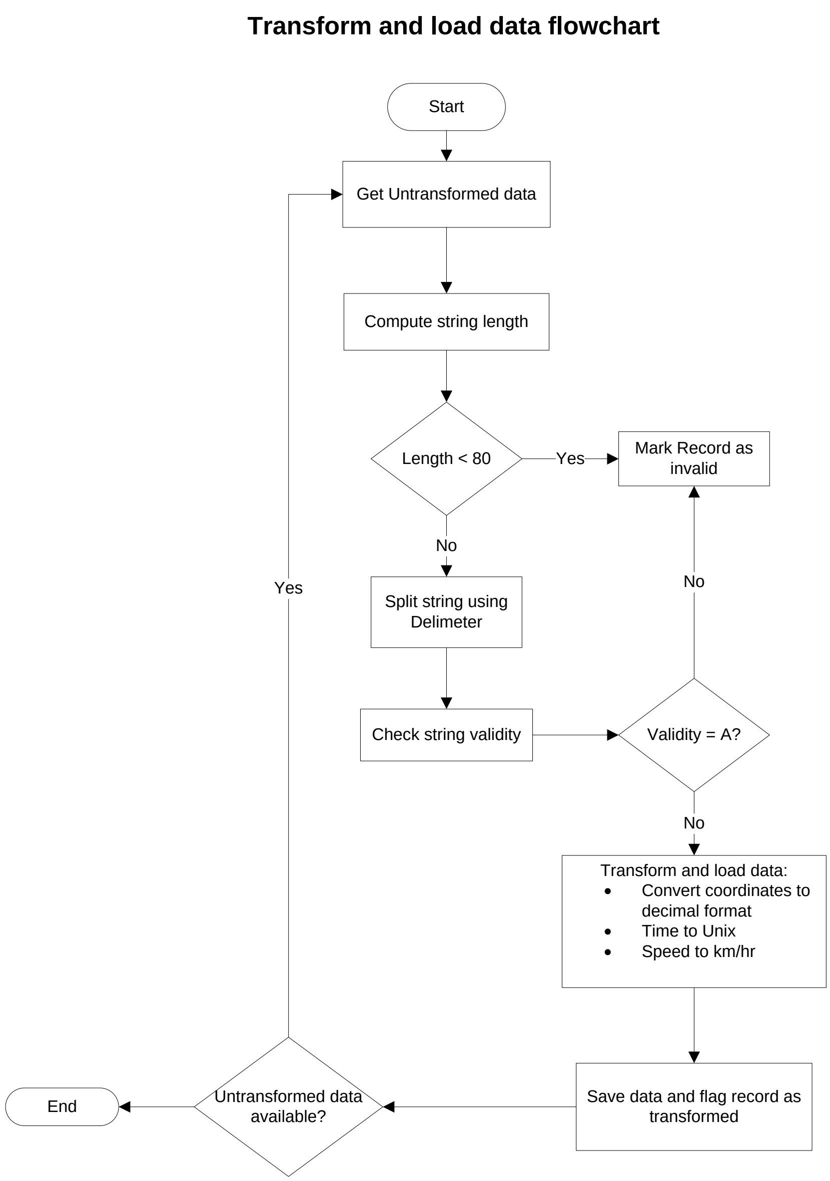


Figure 10: Transform and Load Data Flowchart (Source: Author)

4.2.2.4 Vehicle Current Moving Direction Computation Flowchart

Figure 11 the steps followed in computing the direction in which the vehicle is moving. The direction computation is critical since we can tell whether the traffic is towards town or towards Kawangare. To compute direction, the current and the previous vehicle positions are obtained then Rhumb line function is applied to get the bearing, which is in turn used to calculate the direction. If the direction is invalid then the current vehicle location is marked as invalid.

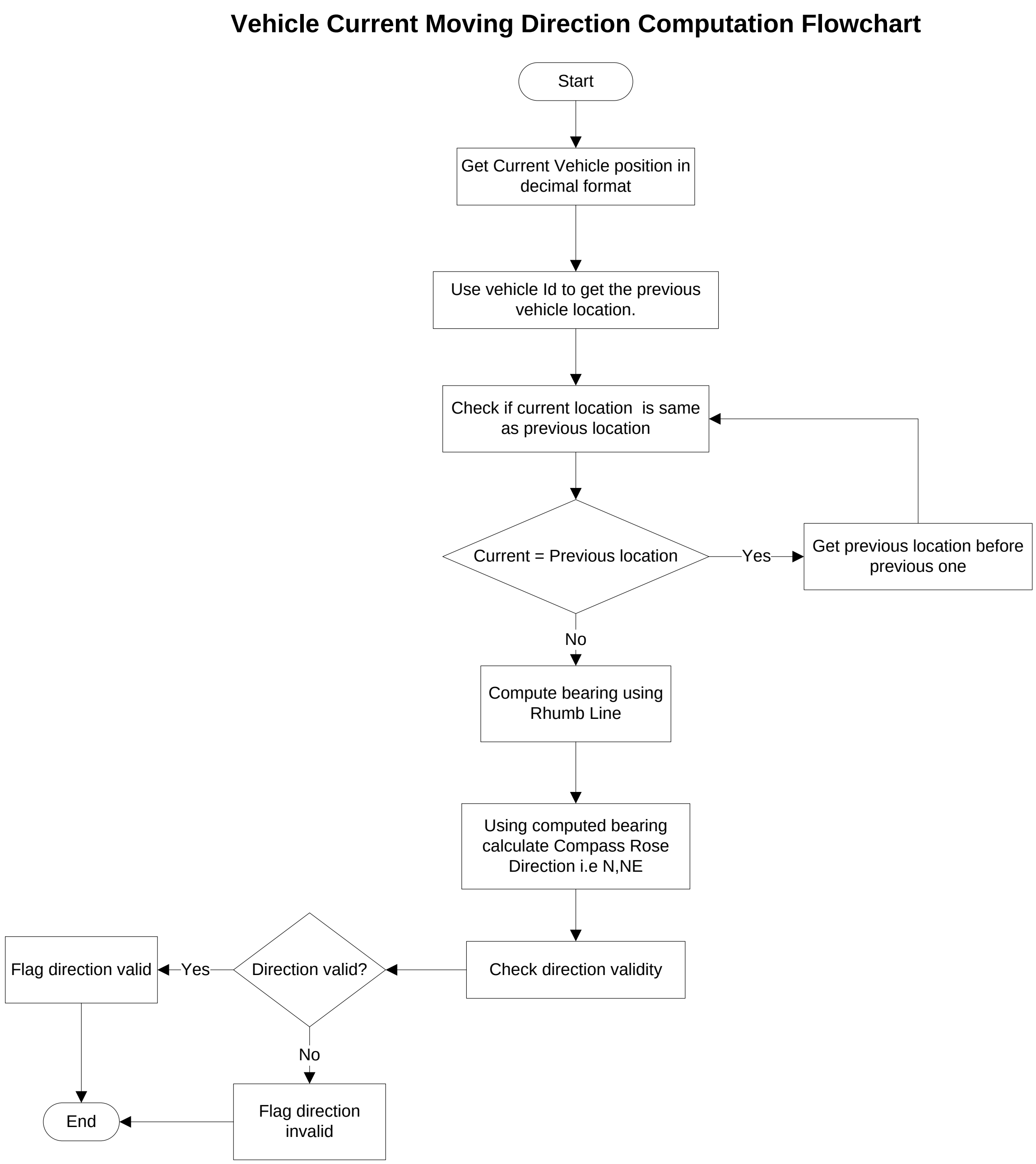


Figure 11 Vehicle Current Moving Direction Computation Flowchart (Source: Author)

4.2.2.5 Vehicle Current Segment Computation Flowchart

Route 46 Kawangware was broken down into not more 1.5 KM segment length to better quantify the performance of congestion management measures. The flowchart shown in figure 12 was used to compute current vehicle position segment. The steps involve getting the current vehicle position, and then the nearest distance between current vehicle position and the route shape data is computed using Spherical Law of Consines Formaula. If the distance obtained is less than the defined radius, then vehicle position is updated to reflect the obtained segment otherwise position is marked invalid.

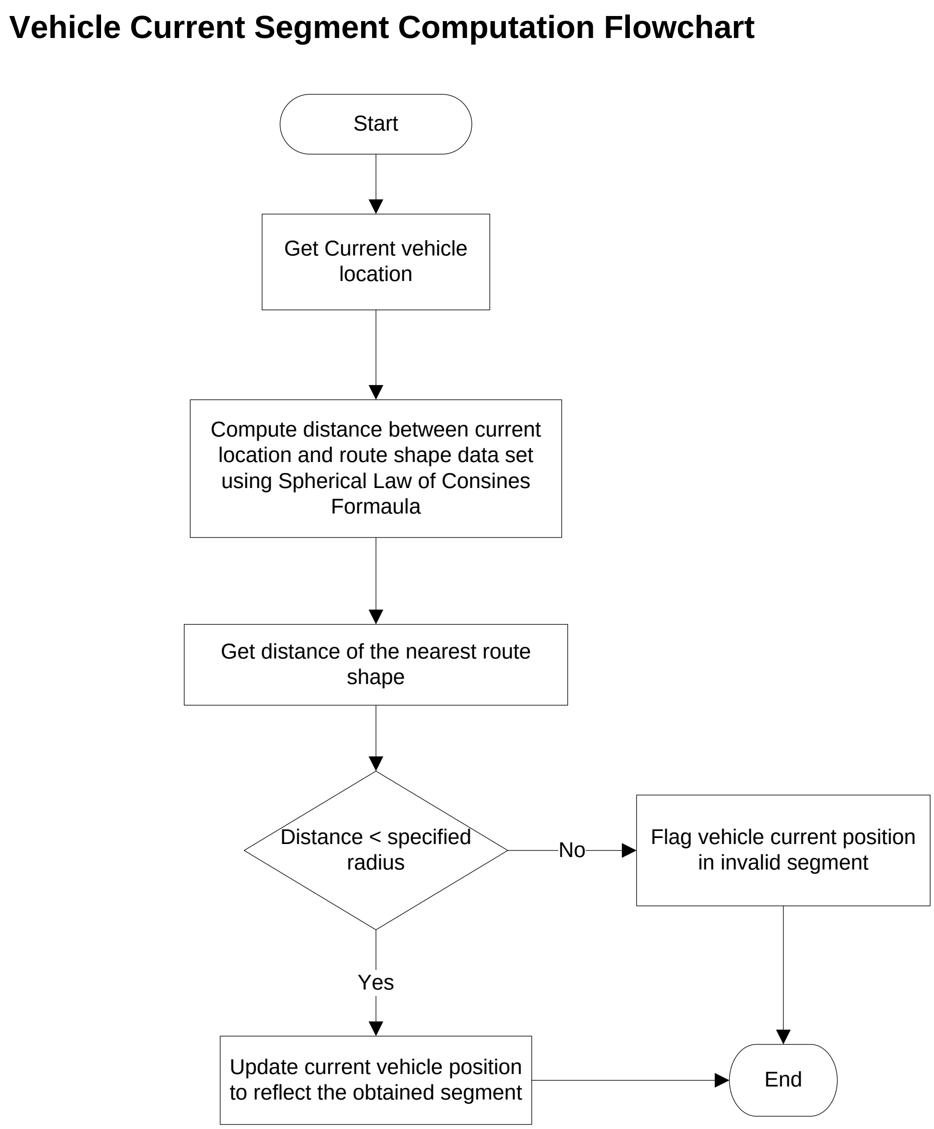


Figure 12: Vehicle Current Segment Computation Flowchart (Source: Author)

4.2.2.6 Segment/Route Average Speed Computation Flowchart

Segment/Route Average Speed Computation Flowchart was the main algorithm of this system; it provided the steps for computing travel time in each segment and the whole route at any given time using very few GPS devices which could not cover all the segments of the route at any given time, meaning that where the readings for a particular segment were missing the algorithm provided the mechanism for dealing with missing data as described in figure 13.

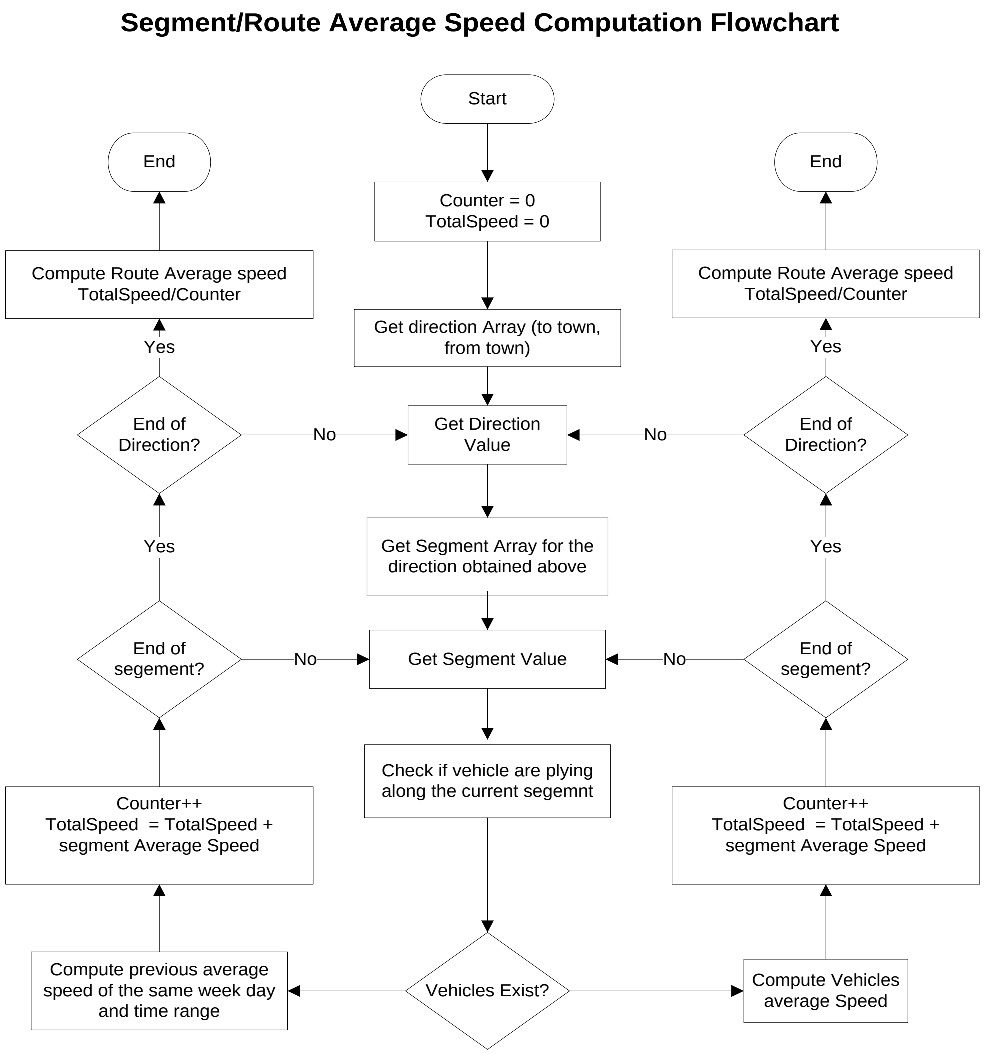


Figure 13: Segment/Route Average Speed Computation Flowchart (Source: Author)

4.2.2.7 Over-speeding SMS notification Flowchart

Over speeding SMS notification flowcharts shows the process used to determine when a vehicle is moving at speed higher than the set speed limit and then SMS notification is send to the traffic police and vehicle owner for necessary action as shown in figure 14.

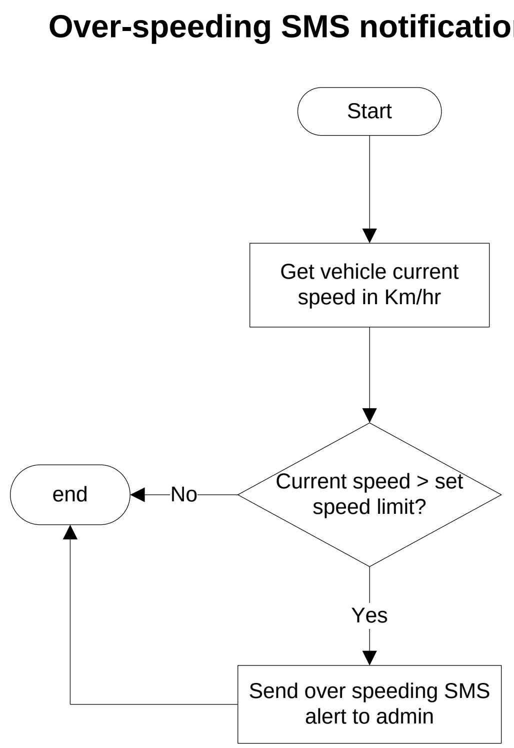


Figure 14: Over-Speeding SMS alert Flowchart (Source: Author)

4.2.2.8 Send Traffic Jam SMS Notification to the Driver Flowchart

The flowchart in figure 15 is used to send SMS notification to drivers when the traffic congestion in the next segment is below defined threshold. The formation is important as driver can make informed decision on the alternative route to use to avoid congested route.

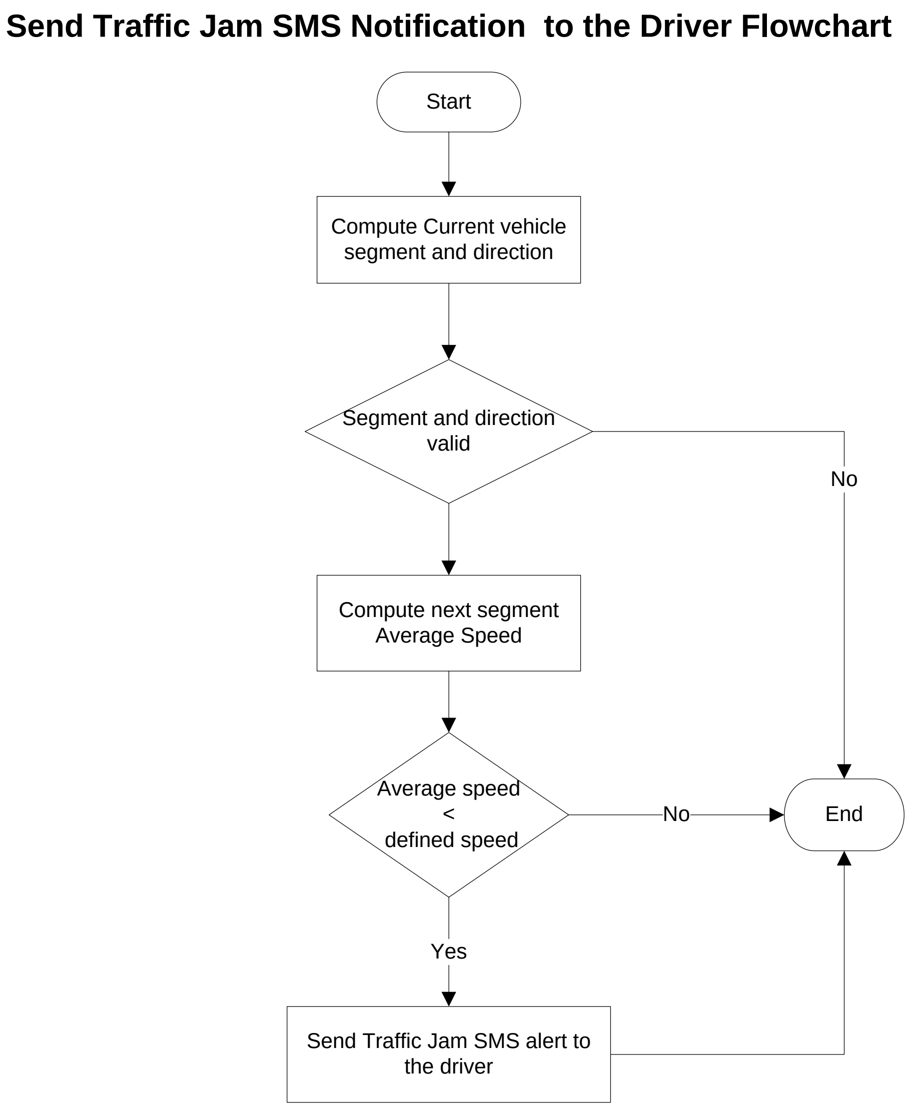


Figure 15: Send Traffic Jam Notification to the Driver Flowchart (Source: Author)

4.2.3 Hot Spot Detector Pseudo Code

The pseudo code below was used to determine the segments of route 46 that experiences high congestion level at a given week day and time of the day.

Step 1: Start
Step 2: Select Route
Step 3: Call SEGMENT(Route[m])
Step 4: For segment[i]= 1 to n step up by 1
Step 5: set cur_avg_speed = AVG_SPEED_CALC(segment[i])
Step 6: set normal_speed = avg_speed[segment[i]]
Step 7: if cur_avg_speed ≥ normal_speed
Step 8: Then goto step 4
Step 9: set front_avg_speed = AVG_SPEED_CALC(segment[i+1])
Step 10: if front_avg_speed ≤ avg_speed[segment[i+1]] + x
Step 11: then goto step 4
Step 12: if no_vehicle[segment[i+1]] < no_vehicle[segment[i]] - y
Step 13: then set hot_spot[segment[i]] = True
Step 14: End

Segment (Route[M])
Step 1: Start
Step 2: divide the Road[m] into ‗ n' equal segments
Step 3: for segment[i] = 1 to n step up by 1
Step 4: initialize avg_speed[segment[i]]
Step 5: Return

Avg_Speed_Calc(segmengt[x])
Step 1: Start
Step 2: set N= cur_no_vehicle[segment[x]] // collected from GPS data
Step 3: set cur_avg_speed = 0
Step 4: for each vehicle in segment[x] = 1 to N step up by 1
Step 5: set cur_avg_speed = cur_avg_speed + speed_vehicle[i]
Step 6: Return cur_avg_speed = cur_avg_speed/N

4.2.4 Database Design

Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a data definition language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

Figure 16 below depicts an Entity Relation diagram (ERD) for Traffic Monitor System. Data receive from GPS device fitted on vehicles is initially save in raw_data table. Raw data from this table is then transformed and loaded into master table, vehicle_position. Route and Segment average speeds are stored in route_travel_time and segment_trave_time tables respectively. Vehicle, driver, owner information is stored in vehicle, driver and vehicle_owner tables respectively. Congestion and over speeding incidences are captured segment_congestion_log and overspeeding log. All system notification messages are stored in message table.

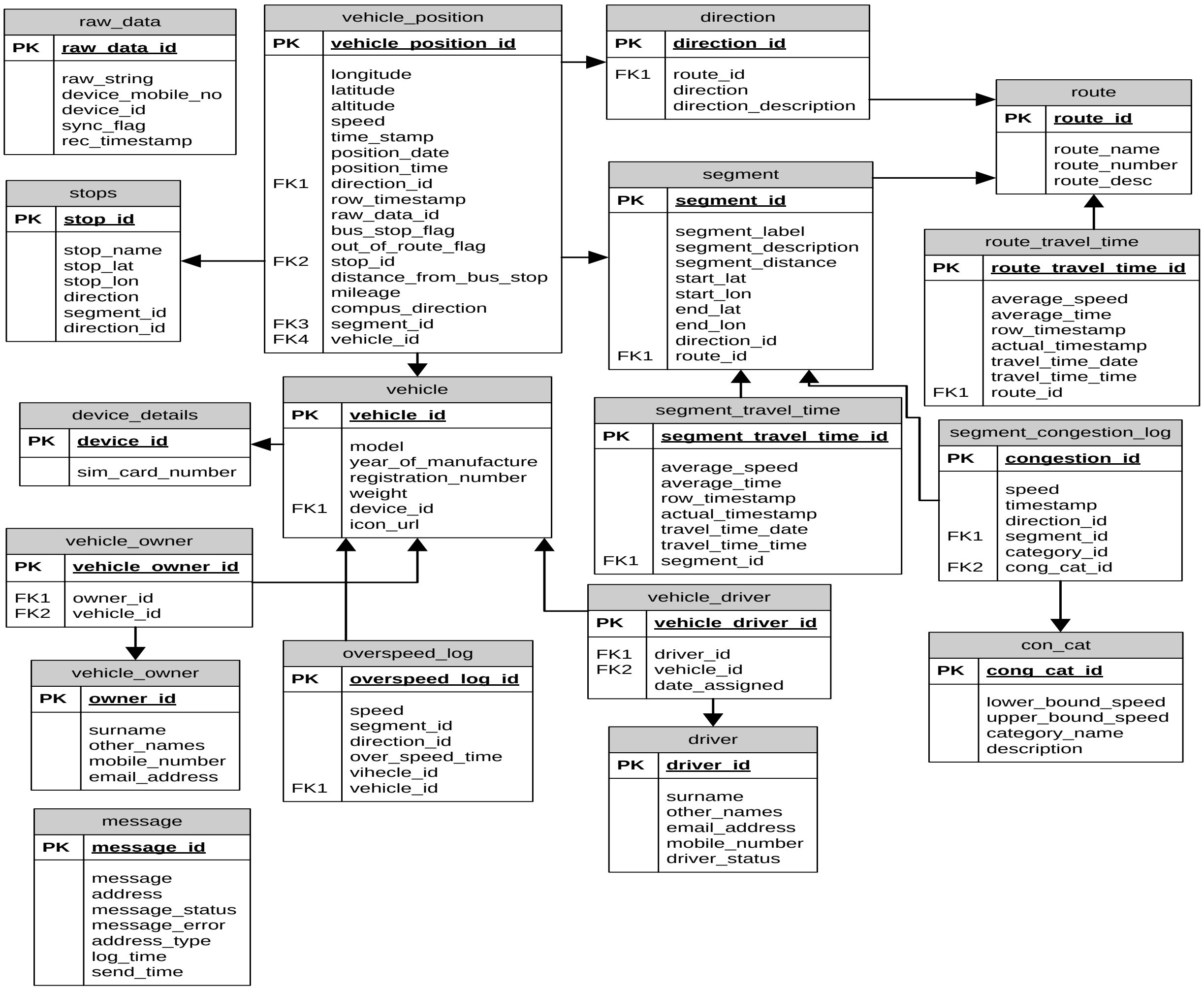


Figure 16: Traffic Monitor Schema (Source: Author)

Chapter 5

5 Results and Discussion

5.1 GPS Error Analysis and Elimination

The ideal floating car data should be continuous in terms of time intervals between each pair of continuous GPS points and the coordinates of each GPS point should be accurate compared with the actual position of the taxicab. However, there exist GPS errors caused by either blockage of the GPS signal or hardware/software bugs during the data collection process. To eliminate such errors, the following measures were taken:

1. The GPS points deviated far away from the bounding area of the study (Route 46 Kawangware) were filtered out. Such error points are outliers and there were 50,658 of them (0.078% of all) in total.

2. All AVL strings with validity parameter not equal to ‘A’ were flagged out of consideration. They were 895 records which constituted 0.00132% of the whole data set.

According to the above analysis, a standard program was developed to eliminate such errors during data loading and transformation period. For instance, if the distance between GPS point obtained and route shape coordinates is greater than defined radius the record was flagged as an outlier, this distance was calculated based on Spherical Law of Cosines. The validity of each string was checked and flagged accordingly i.e valid or invalid. In doing this, the errors in data collected were efficiently and effectively reduced.

5.2 Traffic Analysis Results

5.2.1 Real time Route Speedometer

The figure below depicts real time average speed and travel time for route 46 going towards town.

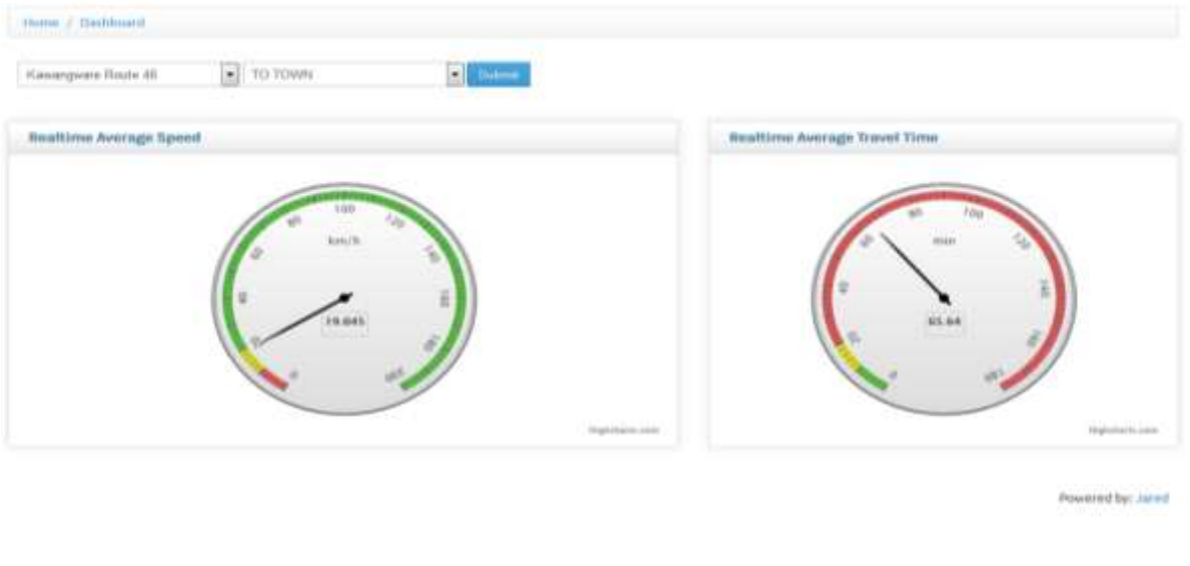


Figure 17: Real Time Route Speedometer

5.2.2 Real time Segement Speedomter

The figure below shows real time average speed and travel time for each segment. Thic can be used to monitor congestion levels to make informed decision on the alternative route to use.

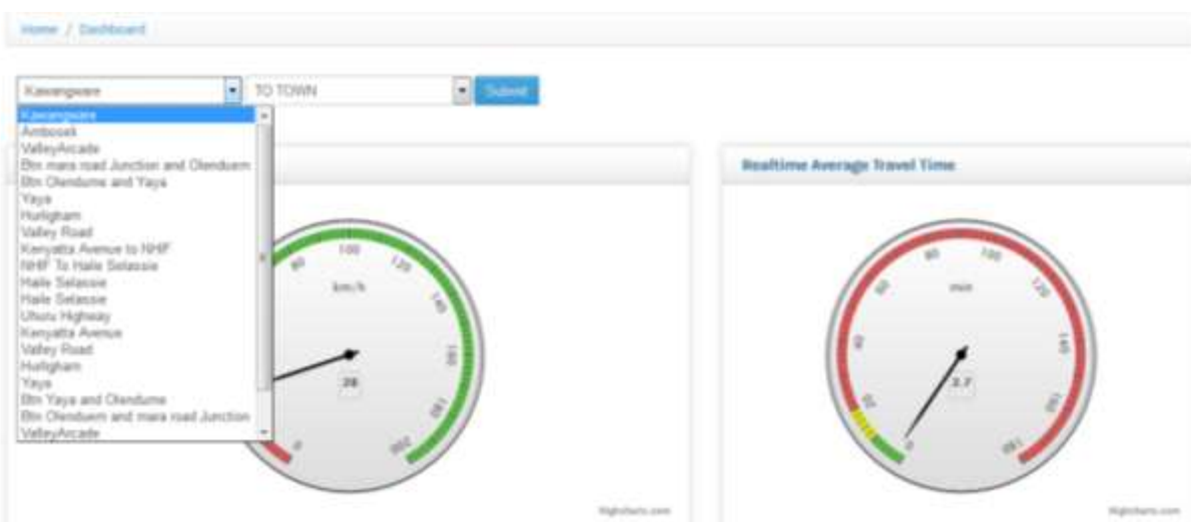


Figure 18: Real Time Segment Speedometer

5.2.3 Live monitoring of vehicle current location

The figure below is an interface using Google Map to Monitor in real time the current location, time and speed of particular vehicle.

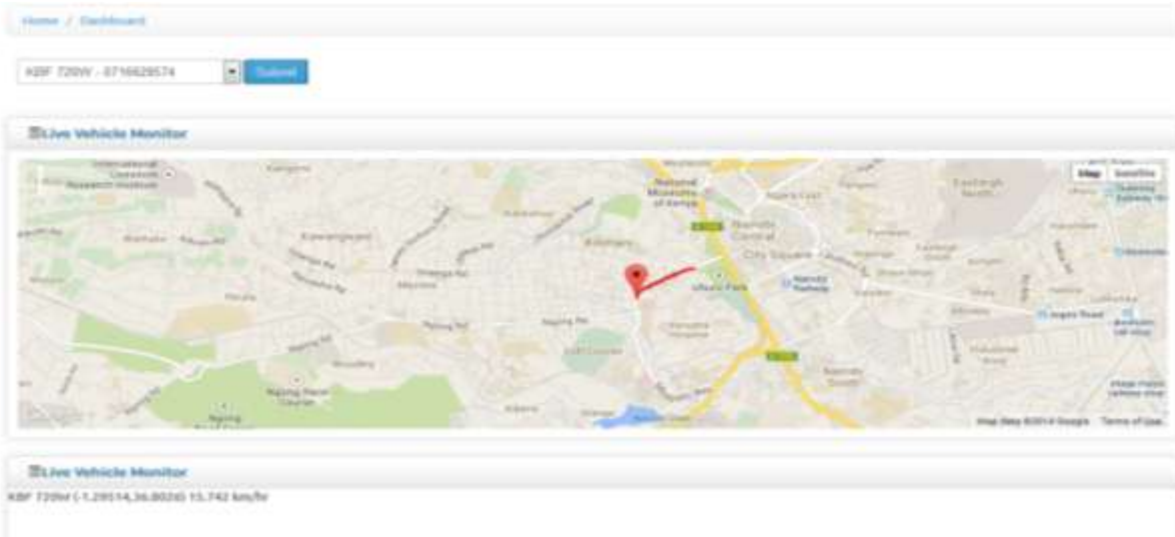


Figure 19:5.1.1.3: Live monitoring of vehicle current location

5.2.4 Route Hourly Average Speed per Week Day

The figure below depict general traffic trend on the whole route per hour and per weekday within a given time range. This can used to predict the traffic situation of a particular week day and on given hour.

Figure 20: Route Hourly Average Speed per Week Day

5.2.5 Congestion levels analysis

5.2.5.1 Congestion levels per Segment per Weekday

The average speed for each route segment, per each day of the week within date and time range was calculated as shown in the figure below. The mean speed represents the average value of all GPS gadget speed points in each segment and in each day, where the ones on the Weekends are higher than the rest of the days. Tuesday reflects lowest average speed but, average speeds of weekdays are generally the same. There is also a downward trend between Monday and Friday. We inference that it reflects the rhythm of city life: traffic congestion on Fridays is typically higher than that on Mondays. The average speeds on 'Ave' row and column in figure 21 are magnified in figure 22, 23 respectively.

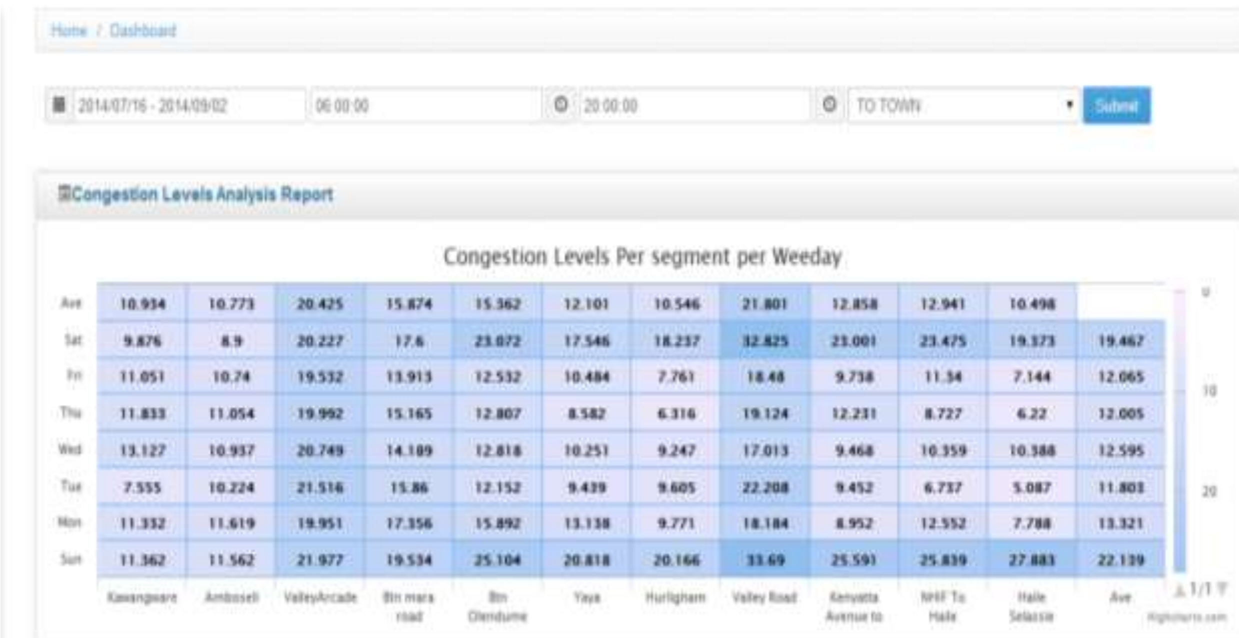


Figure 21: Congestion levels analysis per weekday and per segment

Figure 22: Week Day Average Speeds Bar Chart

Figure 23: Segment Average Speed Bar Chart

5.2.5.2 Congestion Hot Spots

Congestion hot spot pseudo code as indicated in section 4.2.3 of the document was used to identify congestion hot spots along route 46. From the analysis of the GPS data of different days of the week, congestion hot spots locations were. In addition the duration in which these hot spots have impact on the traffic pattern, on different days of week was estimated.

Each identified hot spot is in the form of (location, duration, affected distance). Here is the list of different hot spot on different days as the result of the above said pseudo code.

2014/10/29 - 2014/11/04 | Kawangware Route 46 | TO TOWN | Submit

Congestion Hotspot

Date	Hour	Segment	Distance(KM)	Travel Time (Min)
2014-07-24	11	Olendume and Yaya	1.26	40.00
2014-07-28	08	Olendume and Yaya	1.26	33.60
2014-07-30	10	Olendume and Yaya	1.26	33.90
2014-07-24	09	Hurligham	0.89	25.43
2014-07-25	08	Hurligham	0.89	24.61
2014-07-28	14	Valley Road	1.03	116.60
2014-07-24	10	Kenyatta Avenue to NHIF	0.6	12.04
2014-07-28	08	Kenyatta Avenue to NHIF	0.6	28.35
2014-07-30	09	Halie Selassie	0.87	16.78

Figure 24: Congestion Hot Spots

5.2.6 Traffic Pattern

Average speed for each day of the week and for each hour starting from 06 am to 20 pm was calculated and plotted on line graph as shown in the figures 25,26,27 below. The graphs show that each day of the week has unique traffic pattern from the rest of the days, no correlation between week days, each week day for different weeks and each day. This could be attributed to following reasons:

1. Matatu departing and arrival time at each bus stop is not defined.
2. Matatu Drivers in Kenya do not adhere to designated bus stops, they stop to pick or drop passengers at anywhere any time.
3. The set traffic rules are not fully observed.
4. Traffic incidences for example accidents vary from one day to another.

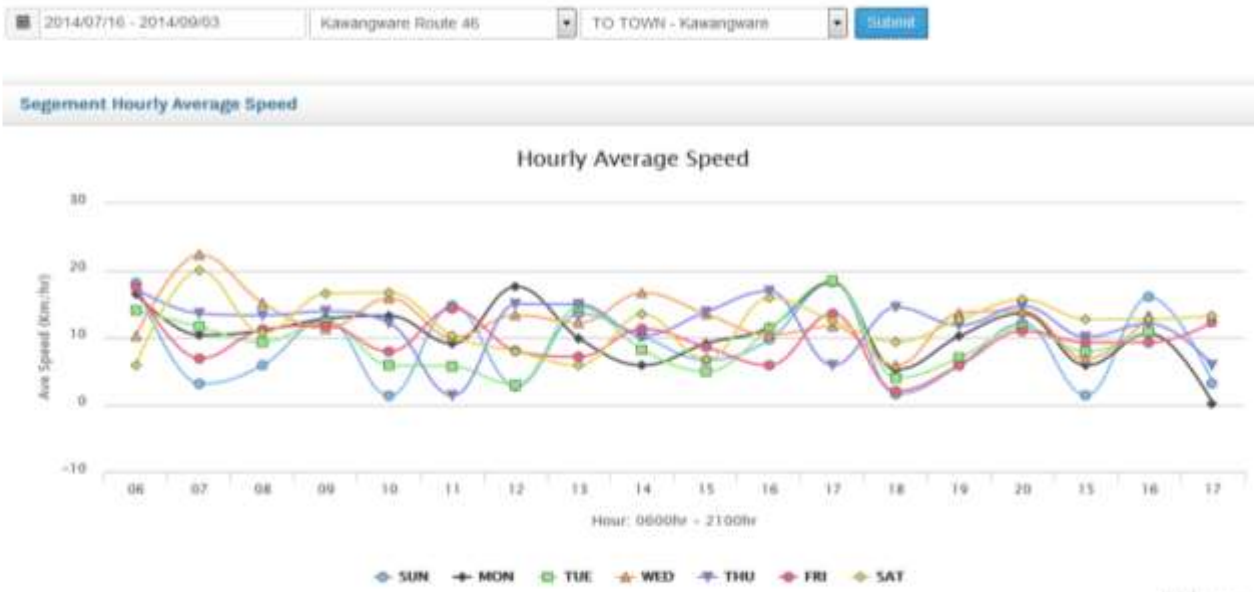


Figure 25: Hourly Average Speed for each day of week

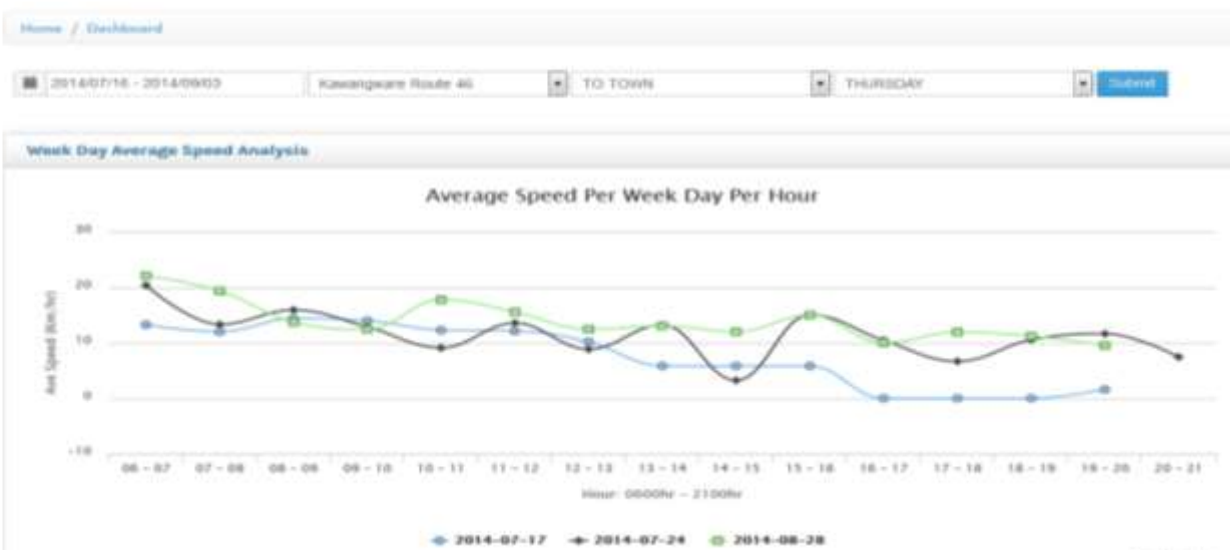


Figure 26: Hourly Average Speed for a given day of week

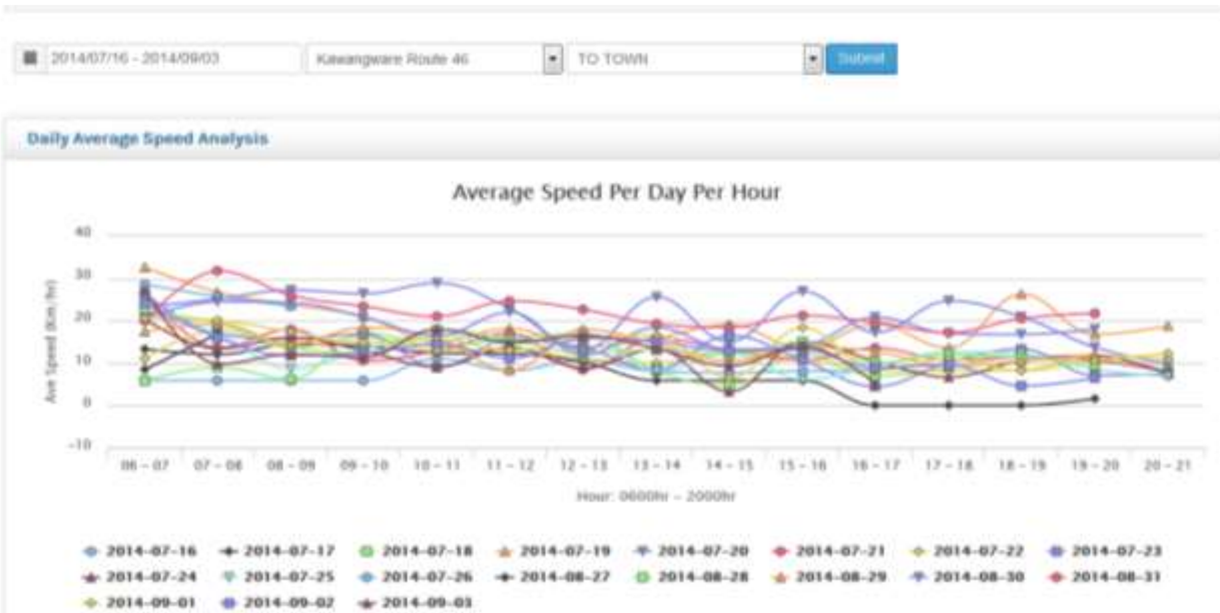
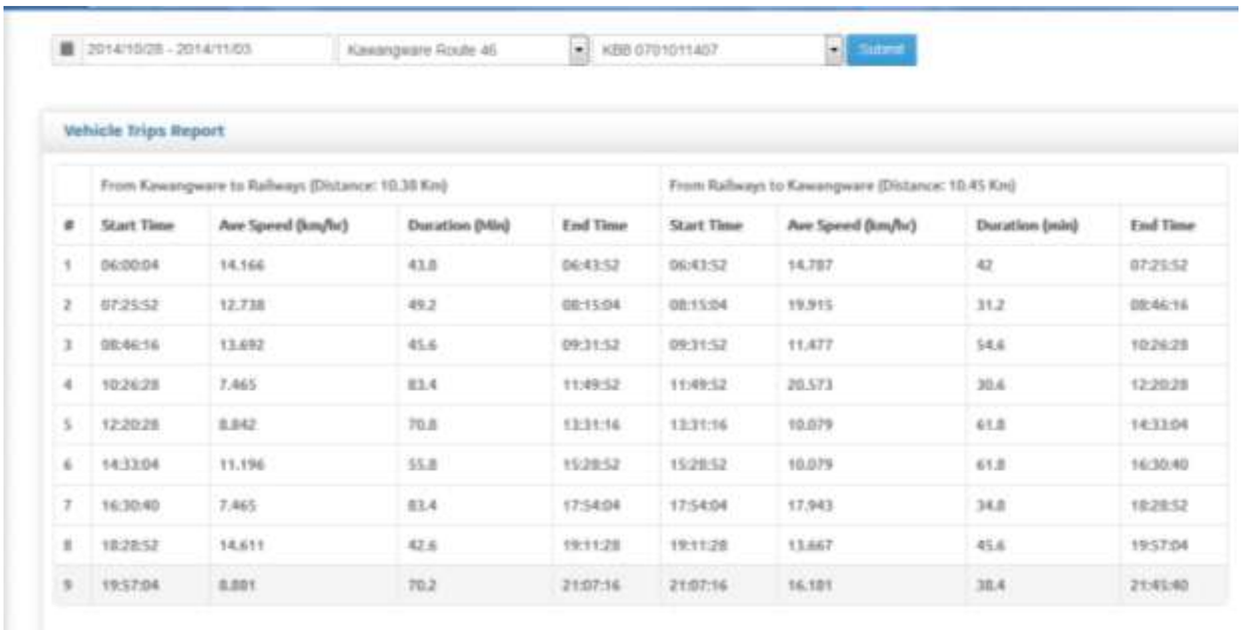


Figure 27: Hourly Average Speed for each day

5.3 Trip travel time

Using GPS location of the vehicle, designated route terminus bus stops defined in the System and then applying GPS position segment mapping algorithm illustrate in System Design section figure 28, the system generates number of trips travel time per vehicle per day report as shown in figure 29. The actual distance cover from Kawangware terminus to Railways is 10.38 km and from Railways to Kawangware is 10.45 km.



	From Kawangware to Railways (Distance: 10.38 Km)				From Railways to Kawangware (Distance: 10.45 Km)			
#	Start Time	Ave Speed (km/hr)	Duration (Min)	End Time	Start Time	Ave Speed (km/hr)	Duration (min)	End Time
1	06:00:04	14.166	43.8	06:43:52	06:43:52	14.787	42	07:25:52
2	07:25:52	12.738	49.2	08:15:04	08:15:04	19.915	31.2	08:46:16
3	08:46:16	13.692	45.6	09:31:52	09:31:52	11.477	54.6	10:26:28
4	10:26:28	7.465	83.4	11:49:52	11:49:52	20.573	30.6	12:20:28
5	12:20:28	8.842	70.8	13:31:16	13:31:16	10.079	61.8	14:33:04
6	14:33:04	11.196	55.8	15:28:52	15:28:52	10.079	61.8	16:30:40
7	16:30:40	7.465	83.4	17:54:04	17:54:04	17.943	34.8	18:28:52
8	18:28:52	14.611	42.6	19:11:28	19:11:28	13.667	45.6	19:57:04
9	19:57:04	8.881	70.2	21:07:16	21:07:16	16.181	38.4	21:45:40

Figure 29 Trip travel time

Chapter 6

6 Conclusions and Recommendation of Further Work

6.1 Achievements

6.1.1 Achievements of research Objectives

The research had four objectives as detailed in section 1.3 which were achieved as follows:

1. **Build real time map-matching GPS data system for traffic monitoring**

   A web based application was developed using PHP and MYSQL platforms. The application had two main modules which include the following:

   - **Data collection module –** The module was made in such a way that a server programs creates a specific program creates a specific type of socket that is used to listen for clients requests (AVL); in case of a successful connection request, the application creates a new socket through which it will receive data from AVL using input stream and inserts into MSQL database.
   - **Data transformation and loading module:** This module was developed based on the flowcharts outlined in section 4.2.2 figures 10, 11, 12.

2. **Determine vehicle average movement speeds on different road segments**

   Average speed for each segment of the route in both directions was computed based on flowcharts outlined in section 4.2.2 figure 13. The speeds were then displayed using gauge graphs and google maps in real time.

3. **Estimate travel time from one point to another**

   Using computed average speeds for each segment, travel time for each segment and whole route was calculated using the formula: time = segment distance (km)/Average Speed (km/hr). Segment distance was calculated using Spherical Law of Cosines.

4. **Analyze traffic flow patterns based on temporal and spatial correlations to identify hot spot areas**

   Transformed and loaded data was analyzed where by various graphs were generated and hot spots location indentified as discussed in section chapter five.

6.2 Research contribution

The main research contribution was the development of an algorithm for calculating segment average speed using four GPS gadgets. The algorithm was designed in such a way that it takes care of missing

GPS data at any given segment in both directions. How to deal with missing data in this research was very important for the success of this project as the number GPS gadgets used could not cover the whole route at any given time.

6.3 Conclusions

The application of AVL in traffic analysis is proving to be the most effective solution compared to other existing traffic management methods like safety cameras, human inspection and speed governors. Mapping of situational road traffic speed at any given time brings out the desired geographic patterns and relationships which are fundamental decision making tools for traffic management. The current complicated traffic networks, traffic speed and the huge number of the traffic participants, requires digital and automatic methods of data capture as the only best solutions to traffic control. The application of AVL technology in traffic data collection gives a detailed study of traffic conditions with an additional provision of better historical repository for road traffic data for future analysis.

The research resulted in the development of a Real time Traffic Monitoring System for automation of traffic management and analysis. The System was built based on several algorithms which were developed in the course of the study i.e. AVL Data Client Receiver, Transform and load data, Vehicle current moving direction computation, Vehicle current segment computation, Segment/route average speed computation, Over-speeding and traffic Jam SMS notification to driver, all of which are essential towards the analysis of traffic conditions on any given road segment and region. Automatic SMS alerts on over speeding drivers also provided better and clearer evidence to Traffic Police as opposed to the speed guns.

The developed System provides accurate data concerning route average speed which is essential in measuring congestion levels, and measurement in turn is a prerequisite to reducing congestion, whether by increasing capacity or through demand management. In particular, provision of accurate travel time information to the public which is important as they can make informed decision in terms of which alternatives routes to use and what is most appropriate day of the week and time to travel.

6.4 Recommendation of Further Work

Mathematical modeling with the current AVL data collected. Enhancing the system towards driver behavior analysis and reporting, with specific real-time warnings, for instance the ability to send a real-time warning alert to the reckless and over speeding drivers.

7 References

Abdelfattah, A. M. & Khan, A. M., 1998. Models for Predicting Bus Delays, In Transportation Research Record. *National Research Council, Washington, D.C.*

Casey, R. F. & Labell, N. L., 1998. Advanced Public Transportation Systems: The State of the Art Update 98.. *Publication FTA-MA-26-7007-98-1. Federal Transit Administration U.S. Department of Transportation.*

Casey, R. F. & Labell, N. L., 1998. Advanced Public Transportation Systems: The State of the Art Update 98.. *Publication FTA-MA-26-7007-98-1. Federal Transit Administration U.S. Department of Transportation.*

Casey, R. F. et al., 2000. Advanced Public Transportation Systems: The State of the Art Update 2000. *Publication FTA-MA-26- 7007-00-1, Federal Transit Administration, U.S. Department of Transportation.*

Casey, R. F., 2002. Advanced Public Transportation System Deployment in the United States. *Volpe Center of the U.S. Department of Transportation for the Federal Transit Administration of the U.S. Department of Transportation.*

Center, U. C. G. N., 2014. *GPS General Information.* [Online]
Available at: http://www.navcen.uscg.gov
[Accessed 10 March 2014].

Chien, S. I. & Kuchipudi, C. M., 2002. Dynamic Time Prediction with Real-time and Historical Data. *National Research Council, Washington D.C.*

Chien, S. I., Ding, Y. & Wei, C., 2002. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. *Journal of Transportation Engineering.*

D'Angelo, Al-Deek, H. M. & Wang, C. M., 1999. Travel Time Prediction for Freeway Corridors. *National Research Council, Washington D.C.* (Hartgen & Fields, 2009)

Dana, P. H., n.d. *Global Positioning System Overview.* [Online]
Available at: http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html
[Accessed 10 March 2014].

Dineen, M. & Cahill, V., 2001. Towards an Open Architecture for Real-Time Traffic Information Management, Dublin, Ireland: Department of Computer Science, Trinity College.

Gyimesi, K., Vincent, C. & Lamba, N., 2011. *Global Commuter Pain Survey,* s.l.: s.n.

Hartgen, D. & Fields, G., 2009. *Gridlock and Growth: The Effect of Traffic Congestion on Regional Economic Performance,* reason.org.

Johnson, C. M. & Thomas, L. E., 2000. Automatic Vehicle Location Successful Transit Applications: A Cross-Cutting Study. *Federal Transit Administration, U.S. Department of Transportation.*

Joseph, O., Francis, A. & B.E.K., P., n.d. Urban Traffic Speed Management: The Use of GPS/GIS.

Karl Petty, Hisham Noeimi, Kumud Sanwal, Rydzewski, Alexander Skabardonis and Pravin Varaiya. The Freeway Service Patrol Evaluation Project: Database, Support Programs, and Accessibility.

Labell, E. J. C. J. A. L. N. & Casey, R. F., n.d.

Mo, J., Sheng, Q. & Zeadally, S., 2009. RFID Infrastructure Design: A Case Study of Two Australian RFID Projects. *IEEE Internet Computing.*

Nermin Mudzelet, 2007, "Road network travel-time calculation using GPS data", Aalborg . 9220, Denmark

Okunieff, P. E., 1997. AVL Systems for Bus Transit. *Transportation Research Board National Research Council, Washington D.C.*

Park, D. & Rilett, L. R., 1999. Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network. Volume XIV.

Quiroga, C. A. & Bullock, D., 199. Travel Time Information Using Global Positioning System and Dynamic Segmentation Techniques. *Journal of Transportation Research Board, TRB, National Research Council, Washington, DC,* pp. 48-57.

Quiroga, C.A. and D. Bullock, 1999. "Travel time information using GPS and dynamic

Salon, Deborah and Eric Aligula (2012) Urban Travel in Nairobi, Kenya: Analysis, Insights, and Opportunities. *Journal of Transport Geography* 22 (May 2012), 65 - 76

Samadi, S., Rad, A., Kazemi, F. & Jafarian, H., 2012. Performance Evaluation of Intelligent Adaptive Traffic Control Systems: A Case Study. Journal of Transportation Technologies

Segmentation techniques", 78 th Annual Meeting, Transportation Resesearch Board, Washington, DC.

Shoaib Kmaran, "A Multilevel traffic IncidentsDetection Approach: Using real time GPS data", Proceedings of the 2007 IEEE Intelligent vehiclesSympossium, Istanbul, Turkey, June 13-15, 2007.

Smith, B. L. & Demetsky, M. J., 1995. Short-Term Traffic Flow Prediction: Neural. *National Research Council, Washington, D.C.*

Transit, B. R., 1996. Advanced Public Transportation Systems Benefits. *Federal Transit Administration, U.S. Department of Transportation.*

Turnbull, K. F., 1993. Evaluation of Automatic Vehicle Location Systems in Public Transit. *Research Report 3006-1F, Texas Transportation Institute, College Station.*

Wall, Z. & Dailey, D. J., 1999. An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location Data. *National Research Council, Washington D.C.*

Zhao, Y. & Tian, Z., 2012. An Overview of the Usage of Adaptive Signal Control System. Applied Mechanics and Materials Volumes 178-181