UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

A GPS SERVER WITH ROAD MAPPING AND DATA MINING FOR TRAFFIC ANALYSIS.

BY

OBUHUMA JAMES IMENDE

P58/61592/2010

SUPERVISOR

Mr. Christopher Moturi

July, 2012.

A Research Project submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science of the University of Nairobi.

Declaration

The Research 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 reference.

Signature: ______________ **Date:** 20/08/2012.

Obuhuma James Imende

P58/61592/2010

The Research 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:** 29 Aug 2012

Mr. Christopher Moturi

The Deputy Director

School of Computing and Informatics

Dedication

I dedicate this Master of Science Research Project to my beloved parents Mr. and Mrs. Obuhuma. There is no doubt in my mind that without their inspiration and counsel, I could not have completed this process.

Acknowledgement

My sincere gratitude goes to my Supervisor Mr. Christopher Moturi for his continued intellectual guidance, motivation, suggestions and support at all levels of my research study. Without him, this project would not have been possible. I would also wish to extend my heartfelt appreciation to my evaluation panel members; Prof. P. Waiganjo, Mr. J. Ogutu, Mr. D. Orwa and Mr. L. Muchemi for their time, input and giving me the opportunity to work on the project.

Special thanks to my parents Mr. and Mrs. Obuhuma who inspired and prayed for me from the start to the end of the course.

Above all, I owe this achievement to the Almighty God for the strength, gift of life and for enabling me pay my fees.

May God bless you all!

Abstract

Traffic control and management currently requires high-tech computerized solutions as opposed to the manual methods that commonly involve the use of traffic policemen, traffic lights and safety cameras. Collection and analysis of road traffic data is a key requirement towards establishment of traffic conditions on any given road segments. The research explores a slightly newer technology, the GPS technology incorporated with road mapping focused at traffic data collection and analysis of traffic conditions.

A GPS data receiver application and traffic analysis system was developed that collects GPS traffic data and provides the ability to monitor and analyse traffic scenarios on the roads for instance the speed of traffic for any given road segment. The system also provides planners on the road usage patterns for analysis and decision making. All these aspects can be analysed both in real-time and historically basing on the fact that historical data is captured and stored for future use. The additional capability of automatic email alerts on over speeding vehicles serves as a better informatory tool to the Traffic Police Department.

The findings show that there is a great need for real-time traffic information analysis due to the tremendous variability in traffic scenarios in major cities like Nairobi in Kenya. The system has been used to show changes in position, speed and directions of vehicles travelling on the Kenyan roads with the speed of traffic and road usage pattern algorithms developed and effectively put in place. The established centralized GPS Server database infrastructure gives room for any kinds of analysis requiring the stored GPS traffic data in a distributed client-server environment.

Using the geographic components in collected GPS data and visualizing by mapping provides a clearer view of the traffic conditions for any given region. Challenges facing the existing systems could be mitigated through the adoption of the GPS based system.

Table of Contents

List of Figures

List of Tables

List of Charts

Abbreviations

AVL:	Automatic Vehicle Location.
2G:	Second Generation.
3G:	Third Generation.
CBD:	Central Business District.
CPRM:	Continuous Route Pattern Mining.
FCD:	Floating Car Data.
FH:	FCD Hourly Average.
FW:	FCD Weighted Average.
GPRS:	General Packet Radio Services.
GPS:	Global Positioning Systems.
GSM:	Global System for Mobile Communications.
HTTP:	Hypertext transfer Protocol.
IDE:	Integrated Development Environment.
IMEI:	International Mobile Equipment identity.
IP:	Internet Protocol.
NMEA:	National Marine Electronics Association.
PHP:	Hypertext Preprocessor.
PPS:	Precise Positioning Service.
RDBMS:	Relational Database Management System.
SIM:	Subscriber Identity Module.
SMS:	Short Message Service.
SPS:	Standard Positioning Service.
SQL:	Structured Query Language.
TCP:	Transmission Control Protocol.
U.S.:	United States.
UDP:	User Datagram Protocol.
WGS:	World Geodetic System.

Definition of Terms

Data Mining: Refers to the process of discovering new patterns from large data sets.

GPRS: Refers to a packet oriented mobile data service on the 2G and 3G cellular communication system's global system for mobile communication (GSM).

GPRMC: Refers to a NMEA sentence type for position and time.

GPS: Refers to a U.S. Government-operated network of earth-orbiting satellites (space vehicles) and ground control stations. This network provides time and position information to receiving stations and devices around the globe. There are approximately 24 active satellites participating in this network at any point in time.

GSM: Refers to a globally accepted standard for digital cellular communication. Developed as a replacement to the first generation analog cellular networks, the GSM standard originally described a digital, circuit switched network optimized for full duplex voice telephony.

IMEI: Refers to a unique 15 or 17 digit code used to identify devices. With respect to GPS tracking systems, it uniquely identifies a specific GPS device from the other and it help during authentication of messages send by the devices to the GPS server.

NMEA: Refers to a non-profit association of manufacturers, distributors, dealers, educational institutions, and others interested in peripheral marine electronics occupations. The NMEA 0183 standard defines an electrical interface and data protocol for communications between marine instrumentation.

POST: Refers to a request method supported by the HTTP protocol used by the World Wide Web. The POST request method is used when the client needs to send to the server as part of the request.

SIM: Refers to an integrated circuit that securely stores the International Mobile Subscriber Identity (IMSI) and the related key used to identify and authenticate subscribers on mobile telephony devices such as mobile phones and computers.

SMS: Refers to a text messaging service component of phone, web or mobile communication systems, using standardized communication protocols that allow the exchange of short text messages between fixed line or mobile phone devices.

Socket: Refers to an endpoint of a two-way communication link between two programs running on the network. A socket is bound to a port number so that the TCP layer can identify the application that data is destined to be sent.

SQL: Refers to a programming language designed for managing data in relational database management systems (RDBMS).

Tachograph: Refers to a device fitted to a vehicle that automatically records its speed and distance, together with the driver's activity selected from a choice of modes.

TCP: Refers to a connection-oriented protocol that provides a reliable flow of data between two computers.

UDP: Refers to one of the core members of the Internet Protocol Suite. Clients can send messages, referred to as datagrams, to the server and/or other host on an IP network without requiring prior communications to set up special transmission channels or paths.

WGS: Refers to a standard for use in cartography, geodesy and navigation. It comprises a standard coordinate frame for the Earth, a standard spheroidal reference surface for raw altitude data, and a gravitational equipotential surface that defines the nominal sea where the WGS 84 is the latest revision that s the reference coordinate system used by GPS.

CHAPTER 1
INTRODUCTION

1.0 Background

In our everyday operations, there is need to incorporate Global Positioning Systems for capture of spatial and temporal data. There is growing need for practical methods involving data mining and road mapping of raw GPS data from GPS trackers for various purposes and scope of operation.

In recent decades, activity-based analysis using GPS equipment as data collectors has been a major issue. Most of these kinds of researches focus on data from wearable GPS recorders for persons because of easy detailed activity logging and interactive validation with users (Kochan et al., 2006). As data needs have increased, more sophisticated methods of data collection have been developed, represented at first by the shift from travel to activity diaries and continuing on to the development of GPS enabled activity surveying (Auld et al., 2008).

Traffic analysis is part of key aspects in developed nations that require better and efficient monitoring. In an effort to control and manage traffic on the Kenyan roads, the Government relies heavily on Traffic Policemen who are strategically positioned on major roads, by-passes and junctions. Traffic lights on major roundabouts for roads getting to the Nairobi Central Business District are often replaced by these traffic policemen due to lack of intelligence in traffic lighting systems.

In Kenya unlike in other industrialized countries, there is totally no centralized database that can be used by the Government and any other interested parties to statistically analyse operations of vehicles on the roads at the capital city and its environs. Despite the fact that several companies exist in Kenya offering vehicle satellite tracking solutions using the GPS technology, none of these companies can willingly avail digital data in the current open data world for reasons commonly attributed to security, privacy and others best known to them. Furthermore, none of these companies is tracking public service vehicles and/or doing the business for statistical analysis relating to traffic and road usage analysis hence at no point does the Government hold digital historical position, route and operational data for the entire road users.

An open data world Global Positioning System (GPS) server offers the best solution to these ever existing problems hence, it is a high time that the Government needs to setup rules that will govern digital real-time data capture and storage to help in traffic and road usage analysis. With these kind of server, the Government through the Ministry of Transport will be required to setup a policy for all vehicle owners to fit their vehicles with tracking devices and all be configured to log their periodic position data to a centralized GPS server probably owned by the Government or privatized. Moreover, basing on the fact that these will be an open GPS data bank, different applications can be optionally developed to query the data for different viable purposes with the authority from the Government.

Conclusively, the research study led to the development of a GPS TCP Server that listens to GPS trackers' data and routes it to a centralised database. In addition, a client-side application that retrieves and displays the raw GPS data in a user-friendly and human readable format has also been developed. The project targeted at streamlining the transport industry by analyzing the operation patterns on the roads and the general road usage patterns and speed of traffic analysis with email alerts on over speeding. Owusu et al. (2006), concludes that, vehicular traffic speeds in the urban environment can effectively be managed by the application of GPS and GIS.

1.1 Statement of the Problem

Traffic Control and Management in Kenya has been and still is a major concern to the Government, road users and the Traffic Department of the Kenya Police. In the recent years, Kenya has been experiencing extremely heavy traffic especially towards the Nairobi Central Business District (CBD). This is as a result of the increase in the number of vehicles on the roads especially privately owned cars and Public Service Vehicles (PSV) hence heavy load than the roads can accommodate. Furthermore, the manual systems of traffic control and analysis together with lack of intelligent traffic lighting systems on major roundabouts also contributes to congestions and traffic jams all over the Nairobi region. In the current digital era, planning for any developments, renovations and/or improvements on the roads is a great challenge due to lack of electronically collected data for justification.

It is as a result of the above stated problem that the research project focused on the use of GPS technology incorporated with a road mapping concept to help in the analysis of traffic on Kenyan roads for instance speed of traffic and road usage pattern analysis. The project was also geared towards the establishment of a centralized GPS data bank through the development of a GPS Server.

1.2 Objective of the Project

The objectives of the research are as follows:-

1. To research and explore on the developments in GPS technology and GPS data mining.
2. To develop a GPS TCP server application that listens to GPS trackers' data, transforms and routes it to a centralised database.
3. To develop client-side applications for analysis of traffic speed and road usage patterns by incorporating a road mapping concept. Speed of traffic and road usage pattern algorithms will be developed in this respect.

1.3 Research Questions

The use of GPS systems in Kenya for traffic analysis posses significant questions which formed part of the research study's scope:-

1. What data formats do GPS trackers use and how can it be interpreted, transformed then routed and stored in a database?
2. What attributes of raw GPS data are useful for traffic analysis and how can they be transformed to user-friendly format when subjected to road mapping and data mining for analysis?
3. What formula can be used to establish vehicles in a given mapped road segment in a move towards the determination of speed of traffic for any given road?
4. What GPS data mining algorithms can be generated for road traffic and road usage pattern analysis on the Kenyan roads using massive GPS data in a centralized database?
5. Can a centralized GPS data bank with road mapping and GPS data mining substitute the tachograph option proposed in November, 2011 by the Kenyan Ministry of Transport to be used for traffic management for Public Service Vehicles after the collapse of the speed governor rule? Can this add value towards the achievement of the Kenyan Vision 2030?

1.4 Justification of the Study

GPS technology is one of the areas that are experiencing growth worldwide with quite a number of applications being researched on and implemented. Traffic Control and Management is one of the most affected aspects in Kenya with high congestions and traffic jams experienced on the roads getting into the CBD of Nairobi. This leads to wastage of useful time and resources and fatigue to road users. The study aimed at addressing these issues through incorporation of GPS technology in Traffic analysis. The improvements in computing, and the ability to collect volumes of GPS data has provided the mechanism by which researchers can begin harnessing this data. For this to be of ultimate value to researchers, there is need to build an object model of transportation. From there one needs to devise a database schema that will allow for data mining and analysis.

The establishment of a centralized GPS data bank provides an opportunity for the development of different applications that can query the data for different viable purposes with the authority from the Government who will own the data hence increasing efficiency in the Public Service Vehicle sector, Ministry of Transport, Ministry of Roads and Public Works, Traffic Police Department, road users among others. These applications may include:-

a) Vehicle position inquiry – an SMS system module can be developed to be used by travelers waiting for specific vehicles at any given bus stop or bus terminus.
b) Traffic Control and road management System – different kinds of systems can be developed to utilise the raw GPS data aimed at further traffic management, planning on renovations of any specific road by analyzing historical data on the times and days that a specific road, area and/or region experiences high or low traffic among other purposes.
c) Data mining models – different data mining models can be developed for analysis of the raw data for example predictive data mining models for different vehicle operation pattern. Analysis on the operation times of public service vehicles with respect to peak and off-peak hours and/or the earliest and latest operation hours of vehicles among other possible user requirements.

Mapping with analysis of situational road traffic speed at any given time brings out the desired geographic patterns and relationships which are fundamental decision making tools for the management of the urban traffic system by the Urban planner. This formed one of the key project's deliverables.

1.5 The Scope of the Study

The study concentrated more on major roads getting to the Nairobi Central Business District. Related literature on GPS for traffic control and management was reviewed. Furthermore, GPS data mining and road mapping literature was also reviewed.

A system development of a GPS server with client-side applications for traffic speed and road usage patterns to demonstrate the concept of use of GPS for traffic analysis formed part of the project's deliverables.

It is understandable that GPS systems rely on specific available GSM network hence, in cases of low or poor GSM network; GPS trackers will not be able to send data to the GPS server. This forms part of the assumptions and limitations outlined in section 1.6.

Assumptions

following assumptions were considered:-

nternet connection to the GPS server shall be stable with minimal and/or no downtimes.

A law and/or policy will be enforced by the Government prompting all vehicle owners to fit their vehicles operating on Kenyan roads with GPS tracking devices.

The GSM network shall be stable and always available. A GPS device used in a low GSM network region will fail to send the position updates hence the affected vehicle may be excluded from the analysis basing on the time bracket.

Real-time tracking in this case is meant to be some few minutes behind the normal global time. This bases on the fact that tracking devices are configured to be sending data at a specified time interval with most device manufacturers recommending 5 minutes periodical intervals.

GPS GPRS messages are relayed almost immediately and there is no big lag.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

This chapter presents the literature review relating to previous work and developments in GPS technology with respect to tracking systems. In addition, the section also analyses related work in GPS data mining and GPS technology for traffic control and management. The chapter further presents the conceptual framework of the research project's system and the communication framework.

2.1 GPS Technology in Tracking Systems

Initially vehicle tracking systems developed for fleet management were passive tracking system. In passive tracking system, a hardware device installed in the vehicle store GPS location, speed, heading and a trigger event such as key on/off, door open/closed (Mukesh and Muruganandham, 2010). When the vehicle returns to a specific location, the device is removed and data downloaded to a computer. They further state that due to real time data requirement, active systems were developed that transmit the vehicle's data in real time via cellular or satellite networks to a remote computer or data centre. According to (Mukesh and Muruganandham, 2010), many vehicle systems that are in use nowadays are some form of Automatic Vehicle Location (AVL), which is a concept for determining the geographic location of a vehicle and transmitting this information to a remotely located server. The location is determined using GPS and transmission mechanism could be a satellite, terrestrial radio or cellular connection from the vehicle to a radio receiver, satellite or nearby cell tower. Mukesh and Muruganandham (2010) concluded that for vehicle tracking in real time, in-vehicle unit and a tracking server is used. The information is transmitted to a tracking server using GSM/GPRS modem on GSM network by using SMS or using direct TCP/IP connection with tracking server through GPRS. The tracking server also has GSM/GPRS modem that receives vehicle location information via GSM network and stores this information in a database.

Hasan et al. (2009) presents a system that allows a user to view the present and the past positions recorded of a target object on Google Map through the internet. The system reads the current position of the object using GPS, the data is sent via GPRS service from the GSM network towards a web server using the POST method of the HTTP protocol. The object's position data is then stored in the database for live and past tracking. A web application is developed using PHP, JavaScript, Ajax and MySQL with the Google Map embedded. Hasan et al. (2009) used the GPRS service which made their system a low cost tracking solution for localizing an object's position and status. To find the nearest location, they used the Spherical Law of Cosines. The formula is used generally for computing great-circle distances between two pairs of coordinates on a sphere and they presented it as follows:

d=R*acos(cos(lat1).cos(lat2).cos(lng2−lng1)+sin(lat1).sin(lat2))

Where, d is the distance between two coordinates (lat1,lng2) and (lat2, lng2) and R is the radius of the earth.

Conclusively, the existing tracking systems are very useful for car theft situations (alarm alert, engine starting, localizing), for drivers being watched and monitored by vehicle owners (speed limit exceeding, leaving a specific area, fuel thefts), as well as for human, animal and pet tracking.

2.2 GPS technology in Data Mining

The feasibility of using computer-aided data collection in conjunction with GPS data tracking was first demonstrated by the Lexington study (Batelle Transport Division, 1997). This study showed that highly accurate trip times and activity locations could be obtained from GPS logging and combined with user input to generate travel patterns. The study, however, was limited to vehicle tracking.

The use of GPS data began with a series of demonstration studies designed to prove the ability to use GPS for identifying activity-travel patterns, and has branched out to several more advanced applications in travel surveying. Currently, most GPS surveys are conducted to provide trip rate corrections to traditional activity diary surveys. However, work is being done on using GPS to monitor changes in overall travel patterns, develop passive activity-travel diaries, and to generate interactive prompted recall activity-travel surveys (Auld et al., 2008).

In a paper, (Agamennoni et al.) presents a fast and robust algorithm for extracting significant places from a set of raw GPS data points. The paper further state that, determining such places provides valuable context information in a variety of applications such as map building, vehicle tracking and user assistance. In their case, they were interested in obtaining context information as a preliminary step towards improving mining safety. They worked with two types of places, namely low-speed areas and loading sites, and showed how to characterize them in mathematical terms via a scoring function.

According to (Marketos, 2009), trying to understand, manage and predict the traffic phenomenon in a city is both interesting and useful. For instance, city authorities, by studying the traffic flow, would be able to improve traffic conditions, to react effectively in case of some traffic problems and to arrange the construction of new roads, the extension of existing ones, and the placement of traffic lights. Marketos (2009) provided a brief outline of the framework (shown in figure 2.1) they propose for efficient and effective Mobility Data Warehousing and Mining. They proposed a trajectory reconstruction algorithm that employs the idea of a filter based on appropriate parameters. The input of the algorithm includes raw data points (i.e., time-stamped positions) along with object-id, and a list containing the partial trajectories processed so far by the trajectory reconstruction module; these partial trajectories are composed by several of the most recent trajectory points, depending on the values of these parameters.

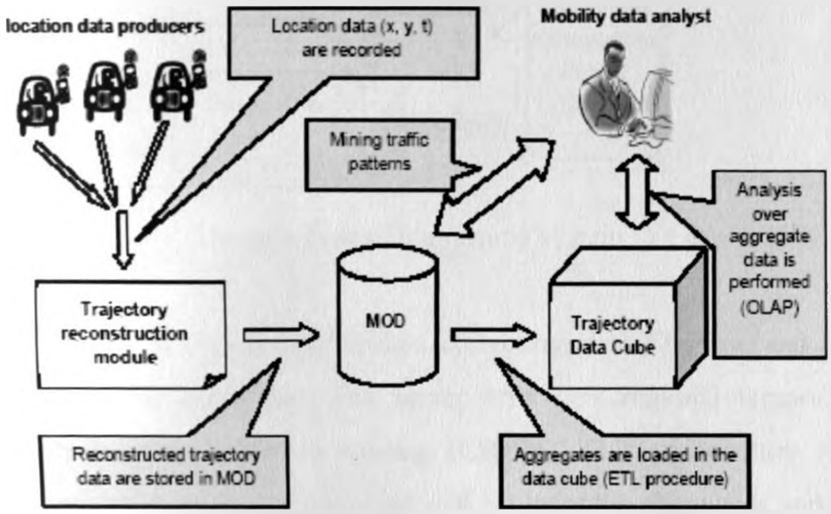


Figure 2.1: The Proposed framework for Mobility Data Warehousing and Mining (Marketos, 2009).

Ye et al. (2009) presents a mining system that was developed to find the continuous route patterns of personal past trips. The system includes two separated parts: the client part and the server part. Figure 2.2 shows the data flow of the mining system. The mining system employs the adaptive GPS data recording and five data filters to guarantee the clean trips data. The mining system uses client/server architecture to protect personal privacy and to reduce the computational load. The server conducts the main mining procedure but with insufficient information to recover real personal routes. In order to improve the scalability of sequential pattern mining, a novel pattern mining algorithm, Continuous Route Pattern Mining (CRPM), is proposed. This algorithm can tolerate the different disturbances in real routes and extract the frequent patterns. Experimental results based on nine persons' trips show that CRPM can extract more than two times longer route patterns than the traditional route pattern mining algorithms (Ye et al., 2009).

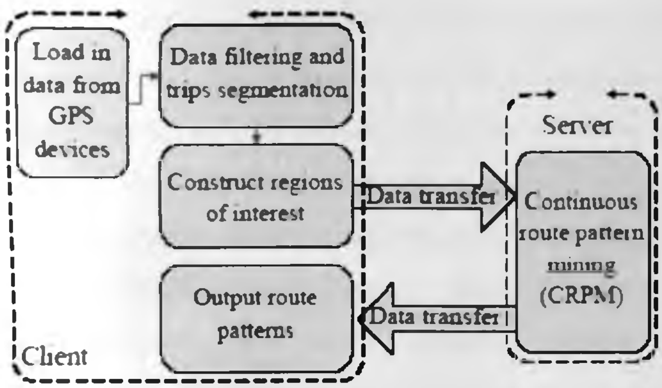


Figure 2.2: The data flow of the mining system (Ye et al., 2009).

As shown in Figure 2.2, the data collecting, data filtering, and construction of interested regions are done at the client part. The server only gets regional temporal sequences and is responsible for continuous route pattern mining (CRPM). This architecture makes the server take most of the computation while keeping users' real routes at the client part, and thus keeps the privacy of users (Ye et al., 2009).

Jan et al. (2000) concluded that GPS is a viable tool to study travelers' route choice patterns. GPS data collection methods can reveal important travel behavior information that was impossible to discern with earlier conventional survey methods. They found that travelers habitually followed the same path for the same trip. However, path deviation increases as origins and destinations become farther apart.

2.3 GPS technology in Traffic Control and Management

Whereas excessive speeds affect the severity of the road traffic accidents (Owusu et al., 2006), crawling speeds in the urban environment is also indicative of congestions. One of the key elements in speed management planning is the functional classification of roads by speed. Owusu et al. (2006), gives an example that 30km/h has been designated for residential areas and 60km/h and above for major arterial roads. They continue to state that nowadays, efficient vehicle monitoring can be achieved by integrating Global Positioning System (GPS) derived traffic data such as vehicle speed and direction of traffic flow into a Geographical Information System (GIS) environment. Owusu et al. (2006) says that, this GPS-GIS integrated system provides real-time meaningful location and status of the vehicles in the network. The system has been used to show the second-to-second position changes in speed and directions of vehicles traveling in Kumasi, the second largest city in Ghana. The system has been used to provide a clearer picture of the traffic-state of every route in the network. It has clearly indicated the road sections where speeds are unacceptable and driver behaviour is affected giving transport planners the option to choose the desired speed management techniques to improve the traffic system.

According to (Yoon et al., 2007), traffic problems such as a traffic jam or delay are some of the most critical issues in our life. If correct traffic information can be given to drivers in advance, these problems may be significantly alleviated. However, collecting traffic information in the traditional way requires enormous costs for sensor installation and maintenance on the roads. As GPS gains in popularity and mobile data communication systems are available, a new solution is possible. In their paper (Yoon et al., 2007), proposes a simple yet very effective method that can capture traffic states in complex urban areas. Yoon et al. (2007) further states that, unlike existing systems which simply report vehicle speeds on the roads, their system can characterize unique traffic patterns on each road and identify traffic states in a road-specific manner. For evaluation, (Yoon et al., 2007) applied their system to two different GPS trace data sets collected in the Ann Arbor area in Michigan. The results showed that accuracy of higher than 90% can be achieved if ten or more traversal traces are collected on each road. Moreover, traffic patterns turned out to be fairly consistent over time, which allowed them to use a larger history in classifying current traffic conditions.

Thianniwet et al. (2009), proposed a technique to identify road traffic congestion levels from velocity of mobile sensors with high accuracy and consistent with motorists' judgments. The data collection utilized a GPS device, a webcam, and an opinion survey. Human perceptions were used to rate the traffic congestion levels into three levels: light, heavy, and jam. Then the ratings and velocity were fed into a decision tree learning model (J48). Thianniwet et al. (2009) successfully extracted vehicle movement patterns to feed into the learning model using a sliding windows technique. The model achieved accuracy as high as 91.29%. They further state that, by implementing the model on the existing traffic report systems, the reports will cover on comprehensive areas. The proposed method can be applied to any parts of the world.

Biem (2010) describes some of their recent work in supporting real-time Traffic Information Management using a stream computing approach. They used GPS data from some taxis and trucks to highlight some of their findings on traffic variability in the city of Stockholm. Biem (2010) also shows how they have used IBM's System S stream processing platform for the purpose of real-time traffic information management. They have developed applications on this platform that process real-time GPS data, generate different kinds of real-time traffic statistics, and perform customized analysis in response to user queries. Examples of customized analyses include continuously updated speed and traffic flow measurements for all the different streets in a city, traffic volume measurements by region, estimates of travel times between different points of the city, stochastic shortest-path routes based on current traffic conditions, etc.

The system can handle large volumes of incoming GPS data. For instance, on a cluster of four x86 blade servers, it can process over 120,000 incoming GPS points per second, combine it with a map containing over 600,000 links, continuously generate different kinds of traffic statistics and answer user queries (Biem, 2010).

In order to benefit from telematics based data collection, (Ehmke and Mattfeld, 2010) time-dependent travel time estimates have to be integrated into time-dependent vehicle routing frameworks. Ehmke and Mattfeld (2010) discusses data collection and the conversion from raw empirical traffic data into information models, an application example compare several information models based on real traffic data regarding its benefits for time-dependent route planning. Then, the integration of information models into time-dependent vehicle routing frameworks is discussed. Ehmke and Mattfeld (2010) consider the allocation and application of travel time estimates for time-dependent vehicle routing in city logistics. Telematics based traffic data is converted into time-dependent planning data. An approach for sophisticated data analysis including data clustering and reduction of memory requirements for routing algorithms is introduced. The resulting information models are compared regarding to usefulness for time-dependent route planning in terms of a fictitious city logistics example, based on real traffic data. The time-dependent calculation of shortest paths is identified as important part of an efficient and realistic representation of the road network typology for time-dependent planning problems (Ehmke and Mattfeld, 2010). The experiments underline the superiority of time-dependent information models (FH, FW) over common static data sets. Floating Car Data (FCD) based time-dependent travel time estimates lead to more reliable routing results compared to route planning based on common digital roadmaps. The data mining approach presented by Ehmke and Mattfeld (2010) provides time-dependent travel times in a memory efficient way without a significant reduction of the itineraries' reliability and robustness.

Nowadays, (Kardashyan, 2011) due to the complicated traffic networks, traffic speed and the huge number of the traffic participants, the safety cameras and other existing traffic management methods are not good enough for controlling and managing traffic in any situation and in any location. Kardashyan (2011) describes a new traffic management solution based on the automatically individual control to any traffic user anywhere and anytime. The principle of the method is as follows: any registered vehicle periodically sends information about itself, which is being decoded and analyzed by the central traffic management unit. As a result, the central traffic management unit knows the location, speed and condition for every single registered vehicle. According to the information, the system can establish traffic management due to the traffic management algorithm. Conclusively, (Kardashyan, 2011) taking into consideration the existing mobile network and the GPS availability almost in all areas of the world the new concept seems economically profitable against the existing 'speed cameras' or human inspection methods.

Tripathi (2010) presents an algorithm for detection of hot spots of traffic through analysis of GPS data where he analyses two data clustering algorithms: the K-Means Clustering and the Fuzzy C-Means Clustering. According to (Tripathi, 2010), after the clustering process stops, a cluster center can be selected, which will display the membership grades of all data points toward the selected cluster center. Tripathi (2010) justifies the fact that they use clustering algorithm for the detection of the hot-spots, where each cluster represents the group of GPS data points having latitude and longitude as their co-ordinate and very small distance between them. To measure the distance between two points on the earth surface, (Tripathi, 2010) derived a formula for calculating geodesic distance between a pair of latitude/longitude points on the earth's surface, using the WGS-84 ellipsoidal model of the earth, which they perceive as the most accurate and widely used globally-applicable model for the earth.

2.4 The Conceptual Framework

2.4.1 How the system works

The following image depicts the general concept of the system.

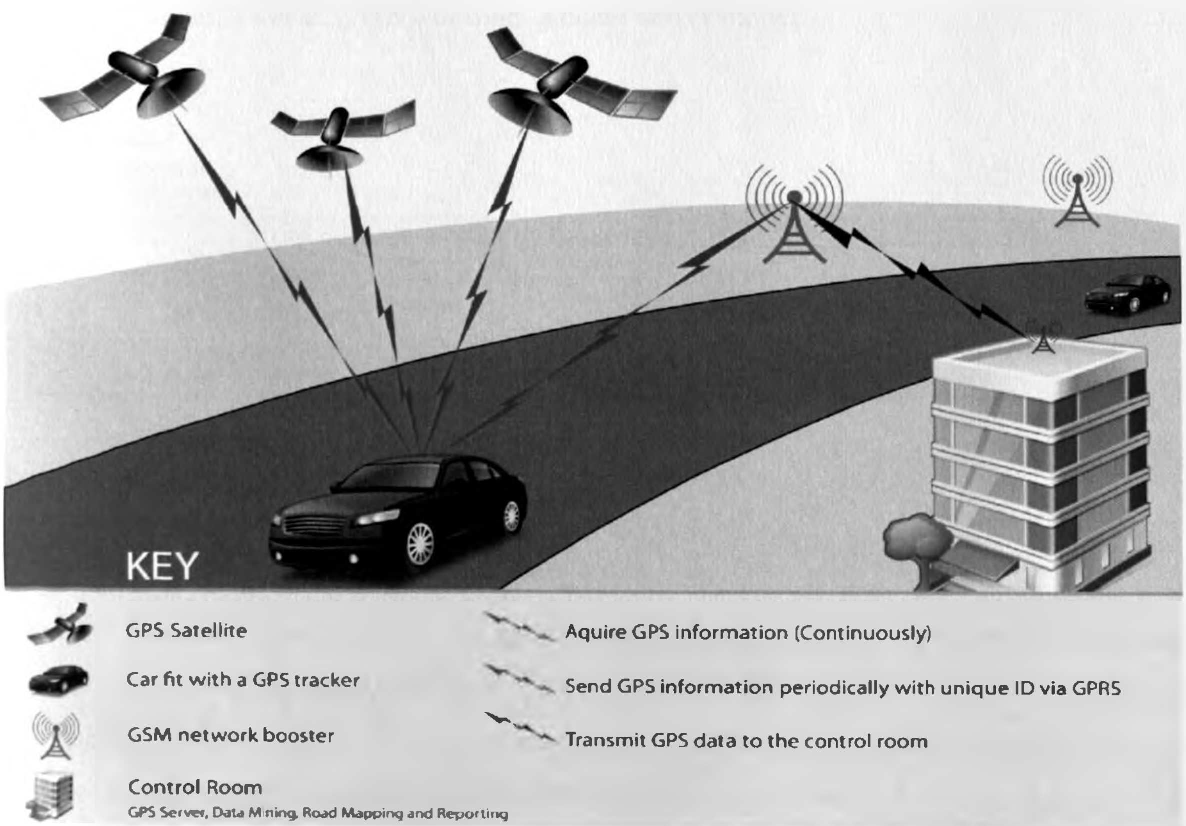


Figure 2.3: The concept of the system.

agram in figure 2.3 outlines three main phases of the GPS based system that occur continuously spectively as follows:-

iPS trackers fit in vehicles on the roads acquire position information continuously from GPS atellites.

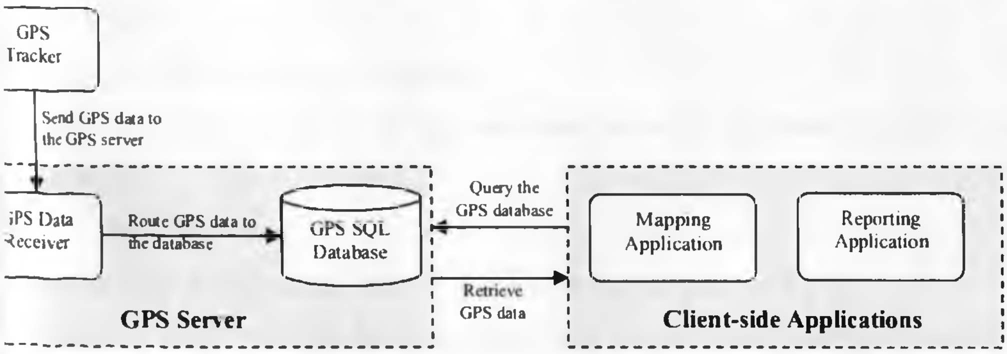
The GPS tracker sends the acquired information to the nearest GSM network booster/access point via GPRS. This occurs periodically basing on the specified time interval, IP address and port onfigured in the GPS tracker. The GSM network bases on the SIM installed in the tracker.

The GSM network booster/access point transmits the data to the Control Room's receiver. These occur via GPRS.

iPS Server at the Control Room receives, interprets and stores the GPS data. Data mining, road ng and reporting operations rely on this captured data for analysis.

The Conceptual Model

esearch study has three main sections; the GPS tracker, the GPS server and the Client-side ations. Figure 2.4 shows the conceptual model with data flow for the system. The data flow of the g system by Ye et al. (2009) shown in figure 2.2 provided the basis for the development of the ptual model. Ye et al. (2009), outlines a client-server model.



e 2.4: The conceptual model.

The GPS Tracker

iPS tracker receives position information continuously from the 24 GPS satellites. The data is send e GPS server via GPRS through TCP communication in NMEA format containing the key eters; device identity, GPS coordinates, date, time and vehicle speed among other parameters. road user should have a GPS tracker installed in his/her vehicle.

2.4.4 The GPS Server

The GPS Server has two main components:-

a) **GPS Data Receiver** – an application interface that receives the GPS data from the tracker in NMEA format, interprets, transforms and routes the data to the GPS SQL Database for storage.

b) **GPS SQL Database** – an SQL database that stores the GPS data.

As GPS trackers' data is received by the GPS server via socket communication, the NMEA format is first interpreted by slitting it into different sections and converting the coordinates into decimal degree format after which, the data is updated in the database.

Communication Framework

1. Socket Communication

A server program creates a specific type of socket that is used to listen for client requests (server socket); in case of a connection request, the program creates a new socket through which it will exchange data with the client using input and output streams. The source and destination IP addresses, and the port numbers constitute a network socket. The process can be summarized as follows:-

i) The client must contact the server by creating a client-local TCP socket, specifying IP address and port number of the server. As a result of this, the client TCP establishes a connection to the server TCP.

ii) The server process must be running.

iii) When contacted, the server TCP creates a new socket for the server process to communicate with the client.

2. Communication between the GPS Tracker and the GPS Server

Socket communication is used to facilitate connection establishment and final receipt of data from the GPS tracker to the GPS server. The GPS server must be running on a machine with a static IP address. A GPS Data Receiver application opens a specific port over which the server listens to data from the GPS tracker. This port must also be opened on the router within which network the server is setup. A GPS tracker configured with the server IP address and port transmits data over the IP layer using either UDP or TCP through a connection request to the server. The GPS Data Receiver accepts the connection request, receives the packets, validates it against authenticated devices using a specific unique unit identifier and then stores in a database of the corresponding device as per a specific unique database identifier.

2.4.5 Client-side Applications

Client-side applications are systems created to retrieve and utilize the data in the database. The project incorporated data mining algorithm(s) for traffic control and management. Road mapping where specific roads are mapped basing on a given set of coordinates for traffic flow analysis on a given road was also factored in with a speed of traffic analysis' algorithm generated.

The client-side application encompasses; data mining, road mapping and reporting. To facilitate these three major aspects. the client queries the GPS SQL database and retrieves GPS data as per the request. This depends on the desired outcome hence different algorithms were developed in this respect. A Road Mapping table was setup to hold road map data with four key fields that hold the following data; the unique ID (Address Code), the address (name), the coordinates of a central spot and the direction values of the mapped area.

CHAPTER 3
METHODOLOGY

3.0 Introduction

This chapter presents the methodology that was used in system analysis, design, implementation and testing. It outlines sources of data and data analysis methods, the data mining algorithm that were used and the design.

3.1 Data Collection

3.1.1 Sources of Data

The study utilized both primary and secondary data in generating additional facts on the subject. A set of GPS historical data to demonstrate data mining was sourced from one of the existing vehicle tracking companies in Kenya. Other sources of data included books, published papers, journals and the internet among other relevant materials.

3.1.2 Data collection Tools, Procedures and Methods

A set of GPS data was sourced from an existing tracking company in electronic format. This was an MS-SQL database file collected using a USB flash disk. The key data attributes required were; coordinates (latitude and longitude), timestamp, speed and optionally the direction.

After the development of the GPS Server, a series of experiments were carried out at various stages of the project to get more data for demonstration of real-time data analysis.

Why use the above procedure and method of data collection?

GPS tracking systems operate in such a way that GPS trackers send position information periodically to the server where it is directly stored in a database. Basing on this, any analysis and/or utilisation of the data requires the analyst to have the database file in electronic format.

3.2 Data Analysis Methods

The study based its findings through quantitative research methods for flexibility and efficiency. The different aspects that were subject to analysis incorporated algorithms for evaluation. Quantitative research which is the process of presenting and interpreting numerical data was employed in the following two phases:-

1. **Descriptive statistics** – The Speed of Traffic algorithm that was developed included measures of central tendency i.e. speeds at a given road segment (average speed i.e. the mean value).

To find a group of vehicles in a given road segment, the Spherical Law of Cosines was used to compute the nearest locations with the K-nearest neighbour algorithm factored in.

2. **Inferential statistics** – The outcomes of the descriptive statistical tests helped in making deductions from the given set of GPS data collected, to test hypotheses set and relating findings with respect to traffic analysis. Graphical representation was also one of the methods for representing and analysing the outcomes with the Google API employed heavily in generation of speed charts and road usage patterns.

3.3 Data Mining Algorithm

The K-Nearest Neighbor Algorithm was employed towards the determination of the congestions and usage patterns on the roads. The K-nearest neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-nearest neighbor category. The purpose of this algorithm was to classify new vehicles based on the position of the initial point. Its operation was based on minimum distance from the query instance to the GPS position data samples to determine the K-nearest neighbors. After we gather K nearest neighbors, we take simple majority of these K-nearest neighbors to be the prediction of the query instance.

The algorithm on how to compute the K-nearest neighbors is as follows:

1. Determine the parameter K = number of nearest neighbors beforehand. This value is all up to you.
2. Calculate the distance between the query-instance and all the training samples. You can use any distance algorithm.
3. Sort the distances for all the training samples and determine the nearest neighbor based on the K-th minimum distance.
4. Since this is supervised learning, get all the Categories of your training data for the sorted value which fall under K.
5. Use the majority of nearest neighbors as the prediction value.

The algorithm was subjected to the GPS position data with the Spherical Law of Cosines formula (see section 3.4) used to compute the distance between two pairs of coordinated.

Depending on the GPS data points that are detected to be close together, the K-Means Clustering algorithm was applied at certain points to establish clusters of vehicles with a given k centroid. This was useful in determination of the road usage patterns.

Why use algorithms?

Algorithms give clear step by step procedures for achieving the targeted outcomes (traffic and road usage patterns) for the given inputs (GPS data attributes).

The choice and design of methods was constantly modified basing on the progress of the development phases of the research study.

3.4 Computing the Nearest Location

The **Spherical Law of Cosines** was used towards the determination of the nearest location basing on the distance required. To achieve this, two different pairs of coordinate data in different tables were subjected to the formulas through SQL statements.

The Sperical Law of Cosines:

$$\text{distance} = \text{acos}(\sin(\text{latitude}_1).\sin(\text{latitude}_2)+\cos(\text{latitude}_1).\cos(\text{latitude}_2).\cos(\text{longitude}_2-\text{longitude}_1)).\text{Radius}$$

Note that:

1. **Radius** – is the earth's radius (mean radius = 6,371km).
2. The angles to be passed to the trigonometric functions need to be in radians.

Why Spherical Law of Cosines?

The Spherical Law of Cosines incorporates six trigonometric functions in its computation hence results in a more accurate distance value that can be as small as 1 metre. It is a simple one line alternative to the Haversine formula. The Equirectangular Approximation formula was eliminated since accuracy was of essence towards the determination of the distance between the K-Nearest Neighbors.

3.5 Hardware and Software Requirements

1. Hardware Requirements:-
   i) Laptop.
   ii) GPS tracker(s).
2. Software Requirements:-

i) Wamp

Wamp contain PHP, Apache, phpMyAdmin and MySQL which had the following benefits towards the development phase:-

i) PHP – the development of the GPS Data Receiver for the GPS TCP server requires sockets. PHP has the capability for socket programming hence it was used at this level.
ii) Apache – acts as the server for the whole application.
iii) MySQL – the database for storage of GPS data.
iv) PhpMyAdmin – provided a user-friendly interface for the management of the MySQL database helping reduce error rates that are common when the console is used.

ii) MS-SQL Server 2008R2

Historical GPS data was sourced in form of an MSSQL Server 2005 file. This data was loaded in MSSQL Server 2008R2 for analysis.

iii) Adobe Dreamweaver CS5

Adobe Dreamweaver CS5 offered a quicker and user-friendly interface for the development of web pages.

iv) JavaScript

JavaScript was employed in the loading of maps where google maps API was called through a JavaScript code. JavaScript code is written into an HTML page. When a user requests an HTML page with JavaScript in it, the script is sent to the browser and it's up to the browser to do something with it.

v) Windows iexpress tool

The Windows iexpress tool was used to package the systems file into executable files.

3.6 Requirement Specification and Analysis

Requirements may be functional or non-functional.

3.6.1 Functional Requirements

This requirement defines functions of a given system or its components, describing an activity or process that the system must perform. The GPS System is split into two sections each with the following functional requirements:-

1. GPS Server – has two sections:-

a) The GPS Data Receiver
   - i) Open a specific TCP port on the server machine.
   - ii) Listen to GPS tracker's messages send in NMEA format over TCP.
   - iii) Interpret the NMEA format message by splitting the message into different parts each with its own meaning and transforming it into the required formats.
   - iv) Route the message to the database.

b) The Database – Used to store the GPS tracking data.

2. Client-side Applications

Client-side applications encompass two aspects each with different functionalities as outlined below:-

a) The road mapping aspect for determination of the speed of traffic functions as follows:-
   - i) Major roads getting to the Nairobi CBD are mapped and identified by specific unique names.
   - ii) Position data in form of coordinates is extracted and subjected to the mapped roads to help determine the state of the roads with respect to traffic flow.

b) The road usage pattern analysis aspect function as follows:-
   - i) First, extract the data from the source system (the GPS database).
   - ii) At the transform stage, only extract the data required for the analysis. This involves selection of required columns only.
   - iii) In the load phase, the data is loaded in form of road usage patterns using traces on the map i.e. polylines.

Automatic email alerts are send when a vehicle's speed exceeds the set speed limit.

3.6.2 Non-functional Requirements

A non-functional requirement specifies systems' properties and constraints. Non-functional requirements may be more critical than functional requirements. If these are not met, the system becomes useless.

The system meets non-functional requirements categorized as:

1. Product requirements – Requirements which specify that the delivered product must behave in a particular way for instance:-
   i) Performance – data throughput, speed and accuracy of the server processing GPS tracker's data is as fast as possible. Response time for client-side application requests for data is also made easier and faster.
   ii) Efficiency – error rates is minimized as much as possible with high speeds of system operation.
   iii) Maintainability – the system is designed in such a way that repairs and new functionalities are easily acceptable.
   iv) Reliability – the GPS server application runs with minimal and/or no system outages.
   v) Compatibility with the Windows OS while the client-side application is browser independent.
   vi) Other aspects put into consideration include: extensibility, flexibility, usability and portability.

2. Organisational and External requirements – Requirements which arise from factors external to the system and its development process and that are also a consequence of organisational policies and procedures e.g. interoperability requirements, legislative requirements, process standards used, implementation requirements etc. This based on Government policies and laws basing on the fact that the study was geared towards setting up a system that targets the Nation as a whole and that privacy may be a key requirement.

3.7 Agile Methodology

The development phase used the agile methodology which is a software development methodology typically used in rapid development and implementation. Agile methodology 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.

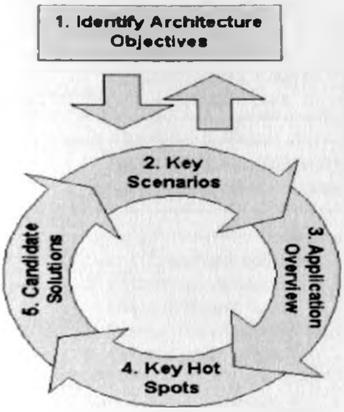


Figure 3.1: Agile Methodology Framework (Meier, 2008).

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.

3.8 Stepwise Refinement Methodology

The Stepwise Refinement methodology was employed in the design of the speed of traffic algorithm and the road usage algorithm. The methodology works as follows:-

1. Break a complex problem down into a number of simpler steps, each of which can be solved by an algorithm which is smaller and simpler than the one required to solve the overall problem.
2. Refinement of the algorithm continues in this manner until each step is sufficiently detailed where refinement means replacing existing steps/instructions with a new version that fills in more details.

Why Stepwise Refinement Methodology?

The fact that the stepwise methodology breaks the given problem into smaller and simpler steps makes it easier to construct and sketch an algorithm in details. Refinement occurs repetitively to streamline the whole process.

3.9 Design

3.9.1 The GPS Server – The Data Receiver

The GPS Data Receiver receives data that is in $GPRMC sentence format as outlined in section 4.3.1. The flow chart below summarises the functionality of the GPS Data Receiver application that is required to be running all the times.

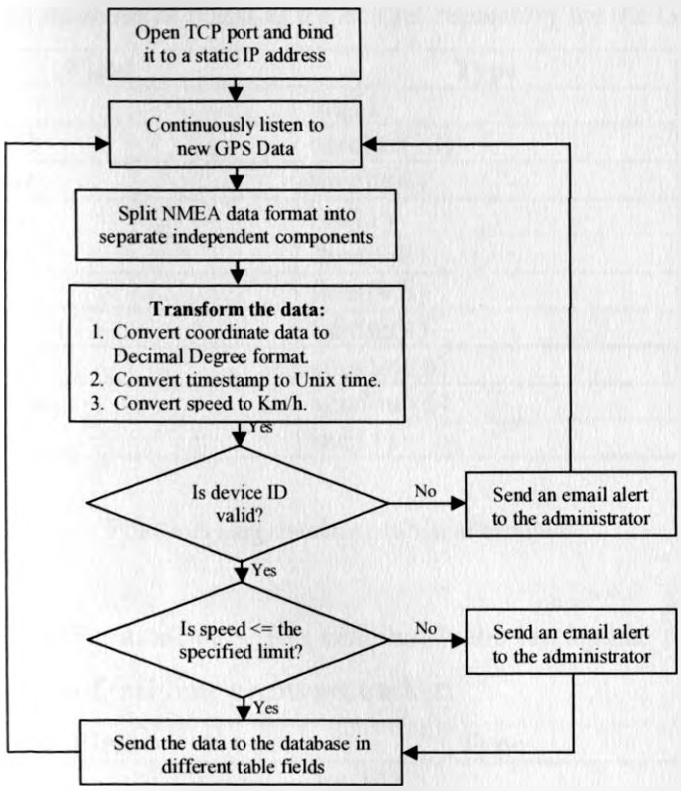


Figure 3.2: The flow chart for the GPS Data Receiver's functionality.

Data Transformation

As the data is received via an open TCP port, the GPS Data Receiver Application transforms it as follows before it is routed to the database:-

i) The data string/sentence is first split in different sections by targeting the separators with each independent attribute assigned to a variable where separators can include "&" "," or "#" depending on device manufacturers.

ii) Variables holding coordinates data are subjected to a transformation to convert the data from Decimal Minute to Decimal Degree format.

iii) The two different variables holding date and time data are combined to form a timestamp (dd-mm-yy hh:mm:ss). The timestamp is then converted to Unix time format to save on storage space.

iv) The variable holding speed data is converted from knots to Km/h where 1knot=1.852km/h.

3.9.2 Database Design

A MySQL database named "GPSData" was created with several tables and triggers as outlined below:-

A. Tables

1. TS_PositionLog – The table holds all the updates received from GPS trackers. It is the largest table in the database as it acts as the central repository for the GPS data.

Field	Type	Null	Default
dbID	int(11)	No	
AssetUID	varchar(20)	No	
GPSState	tinyint(4)	Yes	*NULL*
GPSTime	int(11)	No	
Lon	float(9,6)	Yes	*NULL*
Lat	float(9,6)	Yes	*NULL*
Speed	tinyint(4)	Yes	*NULL*
Direction	smallint(6)	Yes	*NULL*
EngineState	smallint(6)	Yes	*NULL*
Mileage	int(11)	Yes	*NULL*

Table 3.1: Position Log database table attributes

2. TS_LastPositionLog – The table holds the last update per GPS tracker hence it is vital for use in analysis of real-time events per tracker.

Field	Type	Null	Default
dbID	int(11)	No	
AssetUID	varchar(20)	No	
GPSState	tinyint(4)	Yes	*NULL*
GPSTime	int(11)	No	
Lon	float(9,6)	Yes	*NULL*
Lat	float(9,6)	Yes	*NULL*
Speed	tinyint(4)	Yes	*NULL*
Direction	smallint(6)	Yes	*NULL*
EngineState	smallint(6)	Yes	*NULL*
Mileage	int(11)	Yes	*NULL*

Table 3.2: Last Position Log database table attributes

3. TS_RoadMap – The table holds the mapped road segments data i.e. the address code, segment name and the coordinates of the given region.

Field	Type	Null	Default
AddressCode	int(11)	No	
Address	varchar(30)	No	
Lon	float(9,6)	No	
Lat	float(9,6)	No	
Direction	Smallint(4)	No	

Table 3.3: Road Map database table attributes

4. TS_User – The table holds the usernames and passwords for system users.

Field	Type	Null	Default
username	varchar(10)	No	
password	varchar(10)	No	

Table 3.4: Users database table attributes

5. TS_Asset – The table holds the details for the trackers updating on the GPS server. It defines the details for any given unique asset ID and the type of device installed in the asset among others.

Field	Type	Null	Default
dbID	int(11)	No	
AssetReg	varchar(20)	No	
AssetUID	varchar(20)	No	
AssetGSM	varchar(20)	Yes	*NULL*
AssetType	varchar(30)	Yes	*NULL*
AssetModel	varchar(30)	Yes	*NULL*
MileageInit	int(11)	Yes	*NULL*
OwnerID	varchar(10)	No	
AdminID	varchar(10)	No	

Table 3.5: Assets database table attributes

6. TS_Alerts – The table holds the speed limit value and the email address to receive the alert.

Field	Type	Null	Default
dbID	int(11)	No	
SpeedLimit	tinyint(4)	No	
EmailAddress	varchar(40)	No	

Table 3.6: Alerts database table attributes

B. MySQL Triggers

1. An Insert Trigger

The trigger inserts/creates a new record to the '**TS_LastPositionLog**' database table after the user adds a new asset in the '**TS_Asset**' database table.

2. An Update Trigger

The trigger updates existing record basing on the 'AssetUID' field in the '**TS_LastPositionLog**' database table after new position data is received and inserted into the '**TS_PositionLog**' database table.

3. A Direction Trigger

The trigger evaluates the direction of movement for the given asset. The trigger runs on every new record that is received by the GPS Receiver. It first retrieves the 'Lon' and 'Lat' data for the last update for the given 'AssetUID' and then subjects it to the new 'Lon' and 'Lat' with respect to the compass directional points. The computed direction now acts as a new data attribute that is inserted into the 'Direction' field of the '**TS_PositionLog**' database table together with the other data attributes that had been received. To save on database space, the direction data is stored in a coded format i.e. '0' represents 'Stop', '1' represents 'North', '2' represents 'NorthEast', '3' represents 'East', '4' represents 'SouthEast', '5' represents 'South', '6' represents 'SouthWest', '7' represents 'West' while '8' represents 'NorthWest'.

3.9.3 Client-Side Applications' Design

Road Mapping

A Road Mapping table interface was setup to facilitate the mapping of road segments. It functions as per the steps the following three steps:-

1. Determine the length of a road segment to map.
2. Determine the coordinates for a central point for the segment.
3. Update the data in the road map table by filling the road map form. Note that a unique segment name has to be defined for identification of the road segments.

Analyzing data with respect to the road mapped region is key to the determination of the speed of traffic. This is a form of data mining targeted at the analysis of status of roads with respect to speed of traffic for a given road segment. A speed of traffic analysis algorithm outlined in section 3.9.4 was developed in this respect.

Data mining

The stages involved can be depicted in the flow chart shown in figure 3.3. This is used to get the road usage patterns. A road usage pattern analysis algorithm outlined in section 3.9.4 was developed and implemented for better planning and analysis by the users for instance city planners planning for renovation on any road(s).

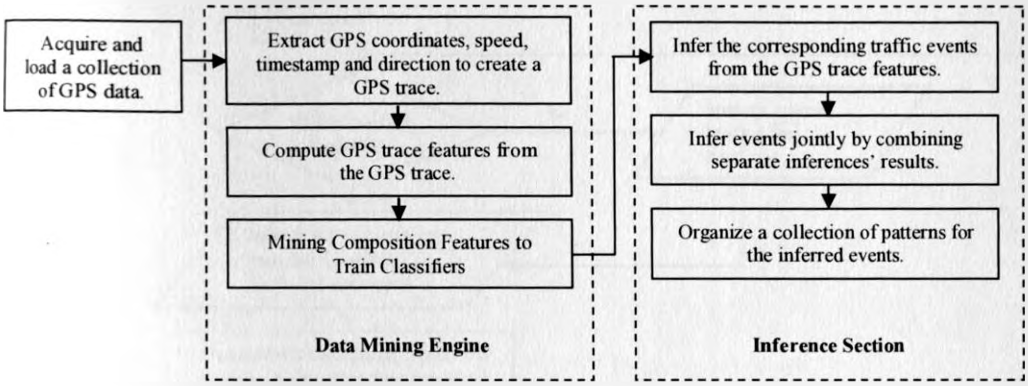


Figure 3.3: The flow chart for the data mining aspect.

The two main sections of the Data Mining aspect are:

a) **The Data Mining Engine** – This is where the required parameters (coordinates, speed, timestamp and direction) are extracted from the set of data subjected to the mining process. Trace features are established from these parameters.

b) **The Inference Section** – This is where conclusions are made basing on the expected results e.g. road usage patterns.

Road Usage Pattern Algorithm

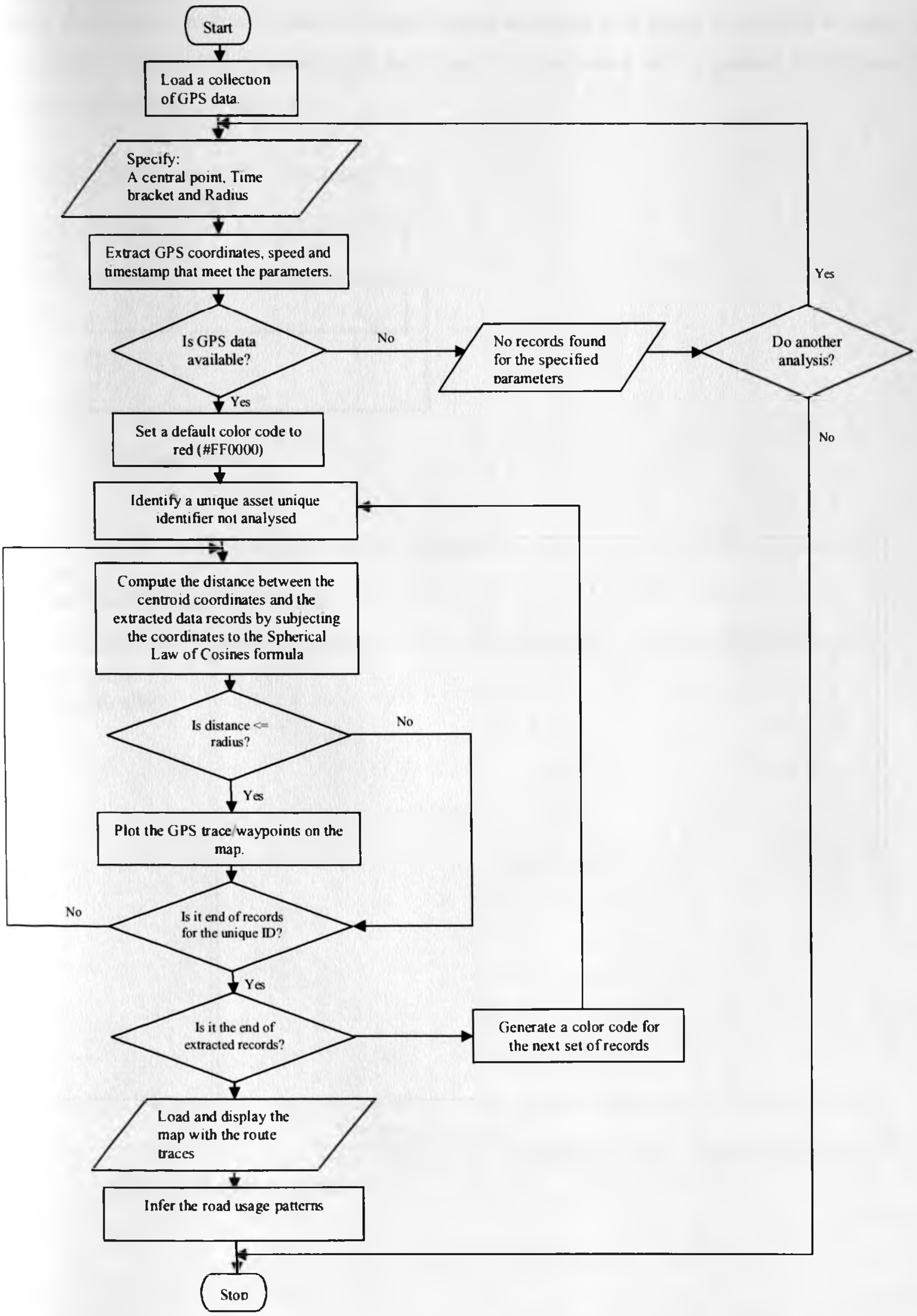


Figure 3.5: The flow chart for the road usage pattern algorithm.

3.9.5 Screen Design

The client-side application has a standard system screen designed as a frame as outlined in figure 3.7. All the interfaces have a link appearing on the menus (top and side) with a general output area. The login screen is as outlined in figure 3.6.

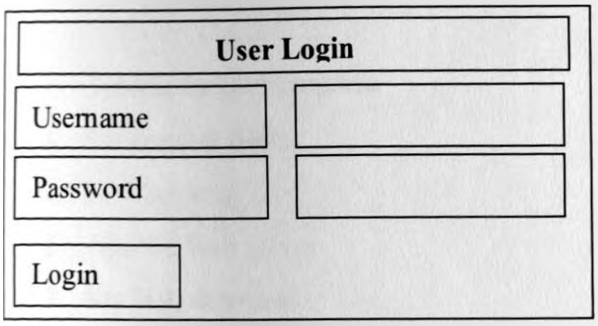


Figure 3.6: The login screen

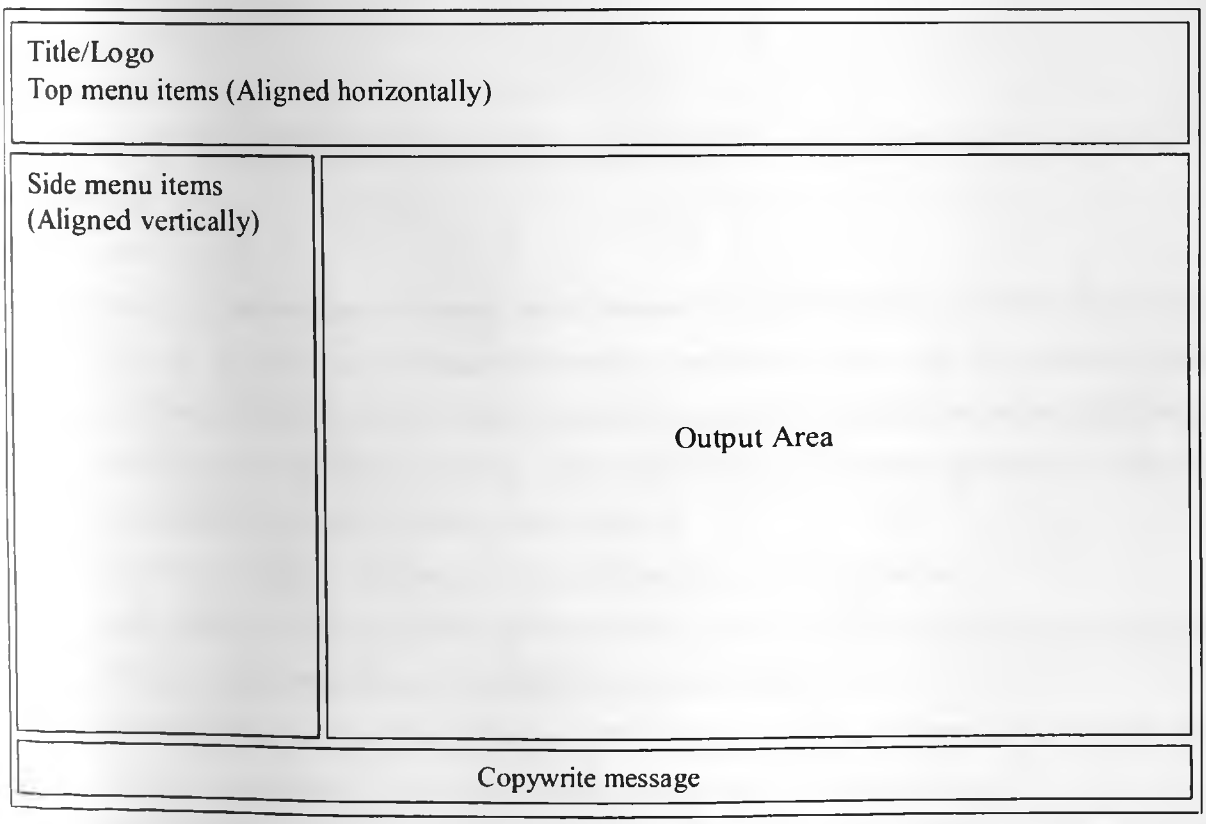


Figure 3.7: The overall system screen

CHAPTER 4
IMPLEMENTATION AND RESULTS

4.0 Introduction

This chapter presents the implementation and testing phases of the research study. It further outlines the results, interpretation and their discussion.

4.1 System Implementation

4.1.1 Server side tools

1. PHP sockets.
2. Apache Web server.
3. MySQL database

4.1.2 Client side languages

1. HTML.
2. JavaScript.

4.1.3 Installation and Configuration of the Application Prototype and other Installations

Installation

1. WAMP

The system uses the wampserver hence install Apache Server, MySQL, and PHP in one step using WAMP. WAMP is a freeware package that bundles Apache, MySQL, and PHP into one executable. Once you download WAMP, double click on the file then begin the installation process as follows:

i) Click Next to begin: After agreeing to the WAMP license, select the destination location. Leave the default location as "c:/wamp" and click Next.

ii) Leave the default Start Menu shortcut as "WampServer" and click Next.

iii) Select automatically launch WAMP5 on startup. This will allow your machine to act as a server whenever it is started. Select the check box and click Next.

iv) WAMP will summarize your selections. Click Install and when prompted to choose a folder for your "DocumentRoot." Leave the default folder as "www" and click Ok.

v) WAMP will prompt you to enter the SMTP server to be used by PHP to send emails. Leave the default value as "localhost" and click Next.

vi) WAMP will then prompt you to enter the default email address to be used by PHP to send emails. Put your email address in this field and click Next.

vii) Choose a default browser for WAMP then click Next.

viii) Congratulations, the installation process is complete, click Finish and Launch WAMP5 now.

2. **The GPS Server**

Run the GPS Server Application setup. When prompted with the path to extract the files to, direct it to "c:/wamp/php/". These will extra the following files in under the PHP folder: gps_server.php, email_alert.php, update_db.php, SMTPClass.php, SMTPconfig.php, config.php, start.vbs and start.bat. Copy the file named "start.vbs" to the startup folder. This is the file that automatically starts the GPS server application through a call to "start.bat". All these run as background processes.

3. **The Client Side Application**

Run the Client-Side Application setup. When prompted with the path to extract the files to, direct it to "c:/wamp/www/". These will extra all the necessary files for the client side application with the index file that starts up the application.

4. **The Database**

Click the WAMP icon in the system tray and Stop All Services then run the Database setup. When prompted with the path to extract the files to, direct it to "c:/wamp/mysql/". A zip file named "data.zip" will be added to the root. Unzip the file and allow it to overwrite the existing data folder with all its files. Click the WAMP icon in the system tray and Restart All Services

Configuration

1. **The router**

The GPS server listens for data via a dedicated TCP port hence a port needs to be opened on the router to allow incoming traffic. Optionally, another port could be opened to allow the router to forward requests to the web server. This will allow you to connect to the client-side applications from outside the LAN through a different port rather than the default port 80.

2. **The web server**

One more file needs to be edited to enable access from outside your LAN. Click the WAMP icon that has been added to the system tray, choose "Config files," and select "httpd.conf": This will open httpd.conf in Notepad. You will need to edit the section called "Controls who can get stuff from this server" by changing the line that says "Deny from all" to "Allow from all" the save the changes. In addition, if you intend to use a different port rather than port 80, then change the "Listen 80" to the desired port number: Click the WAMP icon in the system tray and Restart All Services.

Finally, click on the WAMP icon in the system tray and click on "Put Online" to allow access from outside the LAN.

3. The PHP Service

PHP service also requires configuration as follows: Click the WAMP icon on system tray, choose "Config files," and select "php.ini": This will open the file in Notepad. You will need to edit the section under "Dynamic Extensions" by uncommenting ";extension=php_sockets.dll" to enable PHP sockets. Save the changes then click the WAMP icon in the system tray and Restart.

4. The GPS Server and Email Alerts

Under "c:/wamp/php/", open the file named "gps_server.php" in notepad and set the address variable to that of the server machine and port variable to the port opened on the router. Save the changes. Lastly, the "SMTPconfig.php" file needs to be configured to facilitate automatic email alerts. Open the file under notepad and set the SMTP server, port, user and password using details provided by your mail provider then save the changes.

Running the Application on the local machine

Open a browser and enter "localhost" or "127.0.0.1" as the URL. If your webserver is configured to a different port rather than the default port 80, then add the port number at the end of the URL preceded by a colon for instance "localhost:8500" or "127.0.0.1:8500". This should open the index page i.e. the login page to the system. Using valid username and password, login accordingly.

Running the Application online

Open a browser and enter the public IP address for the network under which the system is installed for instance "http://xxx.xx.xx.xx". If your webserver is configured to a different port rather than the default port 80, then add the port number at the end of the public IP address preceded by a colon for instance "http:// xxx.xx.xx.xx:8500". This should open the index page i.e. the login page to the system. Using valid username and password, login accordingly.

4.2 System Testing

4.2.1 The GPS Server

The GPS server was setup on a test environment and put online for real-time testing. Two GPS trackers were configured to send data to a test IP address over TCP port 8500 within an update interval of 30 seconds. The wamp server was set online over port 8600. The Server application was setup to auto start upon system boot up hence always listening to incoming GPS data traffic over port 8500.

The GPS server was tested using four different models of GPS trackers from different manufacturers.

4.2.2 The Client Side Application and the Database

The necessary files were setup under the web server's root directory with the database file uploaded accordingly as per section 4.1.3.

4.2.3 Usability testing

The application was setup online and locally via "http://localhost:8600/" which served as the real-time online test environment. The online application's URL was circulated to several people (friends, colleagues and lecturers) for testing. All the comments received during the usability testing exercise were analysed and put into consideration basing on their viability and suitability. A test GPS tracking device was installed in a personal car that mostly operates along Magadi road (Ongata Rongai), Langata road, Upper Hill regions and the Nairobi CBD. The test registration number was set to "Demo2".

4.2.4 Usability Test Results

i) A concern on privacy and security basing on the fact that the system will require all vehicles installed with tracking devices. The concern was mitigated through the inclusion of user accounts and login sessions.

ii) The client-side interface was initially reported to be a little bit unproportional. This was mitigated through use of html frames.

iii) Real-time speed of traffic analysis was of a concern to some users where they felt that they couldn't clearly feel the interesting part of the system. This was a major challenge attributed by the fact that only one test device was used. The challenge was not easy to mitigate due to the financial implication that could have been involved to have many test devices.

iv) A general test result was a better comment on the road usage pattern that indicates the routes that a given car used. This outstandingly got the highest vote.

v) Real-time vehicle position on maps and reports with speed charts was positively commented upon.

4.3 Results

4.3.1 The GPS Server – Data Format

The GPS trackers used in the study send data as one sentence with $$ being the start and ## as the end of the sentence. Sample data sentences received by the GPS Data Receiver were as follows:-

a) $$1000000001???&A9955&B051304.000,A,0116.8618,S,03645.5808,E,0.00,202.38,280412,,,A*79|0.9|&C000001 1111&D0026:164&E10000000&Y00180000##

b) $$10737020005???&A9955&B081158.000,A,0116.8607,S,03645.5800,E,0.00,274.29,080512,,,A*78|0.9|&C00000 11111&D00269<21&E00000001&Y00000000##

Other GPS Trackers tested

Three more devices from different manufacturers were tested with the following GPS data sentences received:-

a) ·°? LnY<

b) $$ V?s?ü'? ÖU070548.000,A,0117.6454,S,03647.0322,E,0.00,,280312,,*0D|1.0|1780|0000s-

c) $$ r ÿÿ™U141708.977,A,2309.7484,N,11318.6959,E,000.0,039.1,080612,,,A*64|1.7|51.8|0000|0000,0000| 000000090ÿê

Interpretation of GPS Server – Data Format Results

GPS trackers' manufacturers often pick and 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 apply directly to GPS devices. The GPS device used in this research study uses the $GPRMC sentence that contains all the information that was required: position, speed, and unique identifier data.

The format of the $GPRMC sentence is:

$GPRMC,aaaaaa.aa,b,cccc.ccc,d,eeeee.eee,f,ggg.g,hhh.h,jjjjjj,kkk.k,l*mm

Where:

- **aaaaaa.aa** - Is the time of the fix UTC in hhmmss format.
- **b** - Is the validity of the fix ("A" = valid, "V" = invalid).
- **cccc.ccc** - Is the current latitude in ddmm.mmm format.
- **d** - Is the latitude hemisphere ("N" = northern, "S" = southern).
- **eeeee.ee** - Is the current longitude in dddmm.mmm format.
- **f** - Is the longitude hemisphere ("E" = eastern, "W" = western).
- **ggg.g** - Is the speed in knots.
- **hhh.h** - Is the true course in degrees.
- **jjjjjj** - Is the date in DDMMYY format.
- **kkk.k** - Is the magnetic variation in degrees.
- **l** - Is the direction of magnetic variation ("E" = east, "W" = west).
- **mm** - Is the checksum.

Basing on our sample data received by the GPS Data Receiver which is 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##

The following table outlines an interpretation of the above 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 '#'	Separators for the various data parameters	

Table 4.1: An interpretation of the GPS data sentence

The string "**A*79|0.9|&C0000011111&D0026:164&E10000000&Y00180000**" represents proprietary data that the device tracked along the way. This data is usually secreted by device manufacturers with ***79** representing the checksum.

The GPS Data Receiver splits the received sentence into different independent attributes then interprets and transforms where necessary after which it routes the data to the database. The interpretation and transformation is based on the flow chart and the data transformation process outlined in section 3.9.1.

The interpretation of the data string is similar to that of the following string from a different GPS device manufacturer:-

$$ r ÿÿ™U141708.977,A,2309.7484,N,11318.6959,E,000.0,039.1,080612,,,A*64|1.7|51.8|0000|0000,0000|000000090ÿé

The only difference is that the unique identifier section "**$$ r ÿÿ™**" in this data string requires decoding before it is pushed to the database. It should be noted that the unique identity part depends on a given manufacturer's technical protocol hence for this particular manufacturer, Unicode characters are used to represent the unique identifier as opposed to the normal ASCII characters as illustrated in the first sample GPS data sentence.

A similar concept of decoding was also applied to the following data string from another different device manufacturer with positive results achieved:-

$$ V?s?ü'? ÖU070548.000,A,0117.6454,S,03647.0322,E,0.00,,290911,,*0D|1.0|1780|0000s-

Challenges in interpretation of GPS data

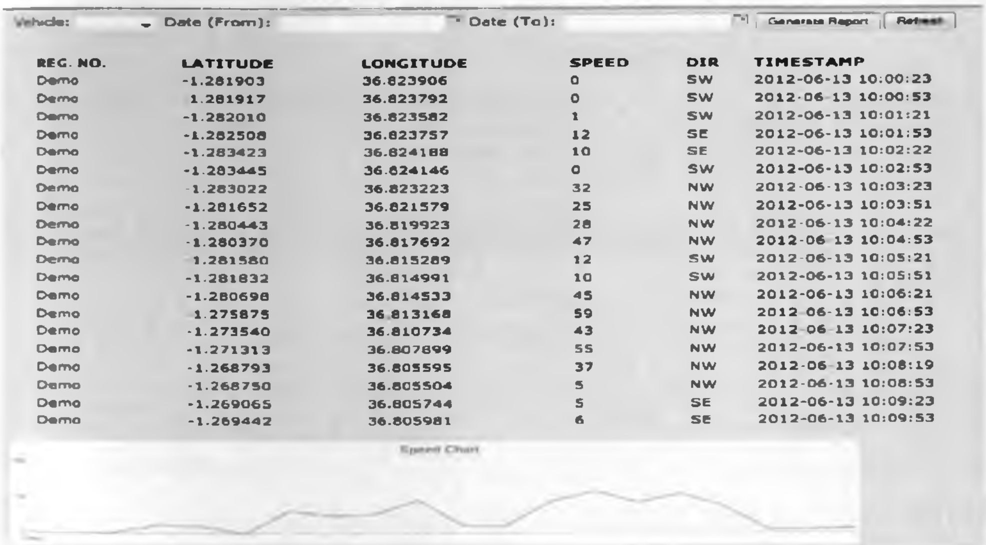
Basing on the GPS tracker's manufacturer, the data sentences depend on the protocol used. In this regard, one more device from a different manufacturer was also tested with the following GPS data sentence received:

" ·°? LnY< "

The data sentence could not be interpreted hence the device was excluded from the research study. It is advisable for GPS system developers to get the relevant technical protocol manuals from device manufacturers to clearly outline an interpretation of the components of the data sentences.

4.3.2 The Speed of Traffic Analysis

A special algorithm depicted in section 3.9.4 has been developed to facilitate determination of traffic speeds. The output is depicted in form of speed charts and a computation of average speed for any given road segment. Figure 4.1 and 4.2 shows sample speed charts generated from individual vehicles moving in different traffic levels analysed basing on the speed, time and region. Section 4.4.1 discusses a performance analysis that elaborates further on the significance of speed of traffic analysis.



REG. NO.	LATITUDE	LONGITUDE	SPEED	DIR	TIMESTAMP
Demo	-1.281903	36.823906	0	SW	2012-06-13 10:00:23
Demo	1.281917	36.823792	0	SW	2012-06-13 10:00:53
Demo	-1.282010	36.823582	1	SW	2012-06-13 10:01:21
Demo	-1.282508	36.823757	12	SE	2012-06-13 10:01:53
Demo	-1.283423	36.824188	10	SE	2012-06-13 10:02:22
Demo	-1.283445	36.824146	0	SW	2012-06-13 10:02:53
Demo	-1.283022	36.823223	32	NW	2012-06-13 10:03:23
Demo	-1.281652	36.821579	25	NW	2012-06-13 10:03:51
Demo	-1.280443	36.819923	28	NW	2012-06-13 10:04:22
Demo	-1.280370	36.817692	47	NW	2012-06-13 10:04:53
Demo	1.281580	36.815289	12	SW	2012-06-13 10:05:21
Demo	-1.281832	36.814991	10	SW	2012-06-13 10:05:51
Demo	-1.280698	36.814533	45	NW	2012-06-13 10:06:21
Demo	1.275875	36.813168	59	NW	2012-06-13 10:06:53
Demo	-1.273540	36.810734	43	NW	2012-06-13 10:07:23
Demo	-1.271313	36.807899	55	NW	2012-06-13 10:07:53
Demo	-1.268793	36.805595	37	NW	2012-06-13 10:08:19
Demo	-1.268750	36.805504	5	NW	2012-06-13 10:08:53
Demo	-1.269065	36.805744	5	SE	2012-06-13 10:09:23
Demo	-1.269442	36.805981	6	SE	2012-06-13 10:09:53

Figure 4.1: Speed chart – Nairobi CBD to Riverside Park (2012-6-13 10:00:00 to 2012-6-13 10:10:00).

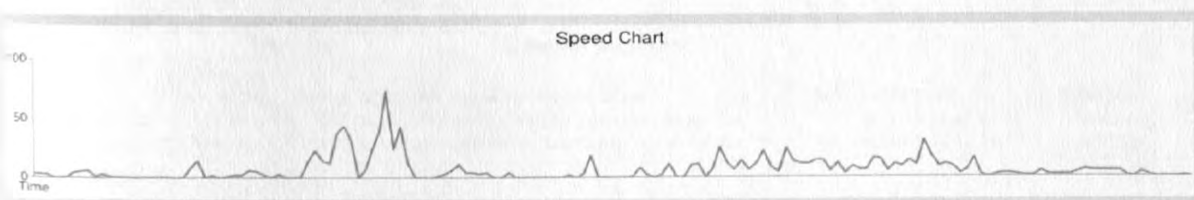


Figure 4.2: Speed chart – Nairobi CBD to Langata (2012-6-11 17:45:00 to 2012-6-11 20:00:00).

Over speeding email alerts are triggered when a given speed limit is exceeded. This depends on the speed limit set by the user. Figure 4.3 shows sample speed alert messages received in a gmail address. The messages were send from james@japnavi.com which was a test address configured for email alerts.

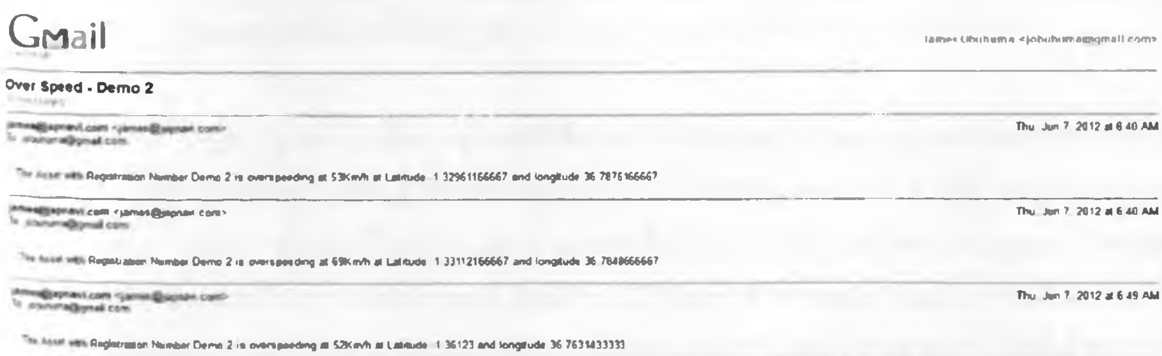


Figure 4.3: Over speeding email alert.

Using this information from many different vehicles, system users at the Control Room can predict the speed of traffic for any given road segment. Figures 4.4, 4.5 and 4.6 show traffic speed analysis for three different road segments at different times of the day for different time intervals. Vehicle registration numbers **Demo** and **Demo 2** in figures 4.4 and 4.5 represent the two tracking devices that were used during system testing while **1097287730, 1163425332** and **3256525228** in figure 4.6 represent three vehicles in the Historical data sourced for GPS data mining demonstration.

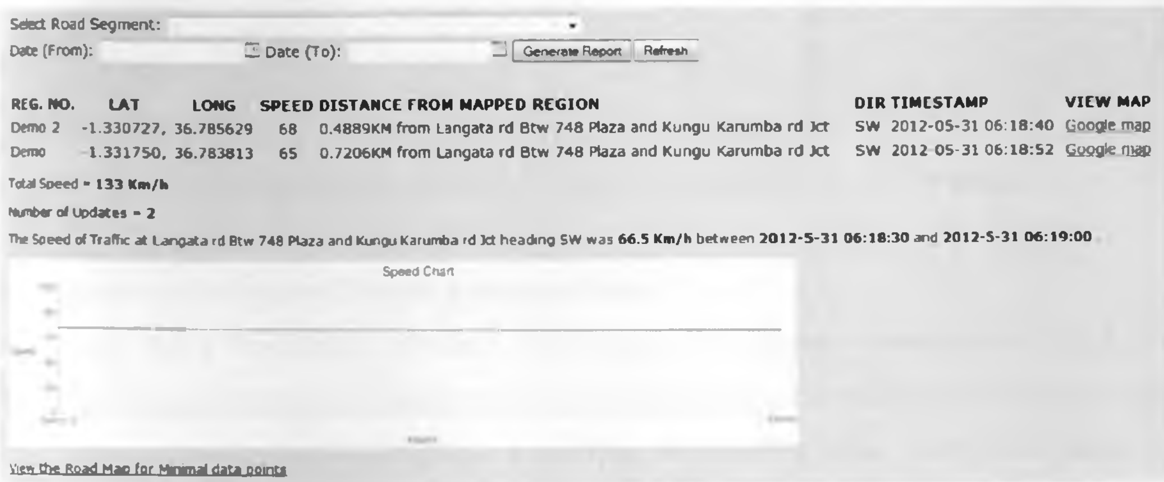


Figure 4.4: Output and Speed chart (2012-5-31 06:18:30 to 2012-5-31 06:19:00).

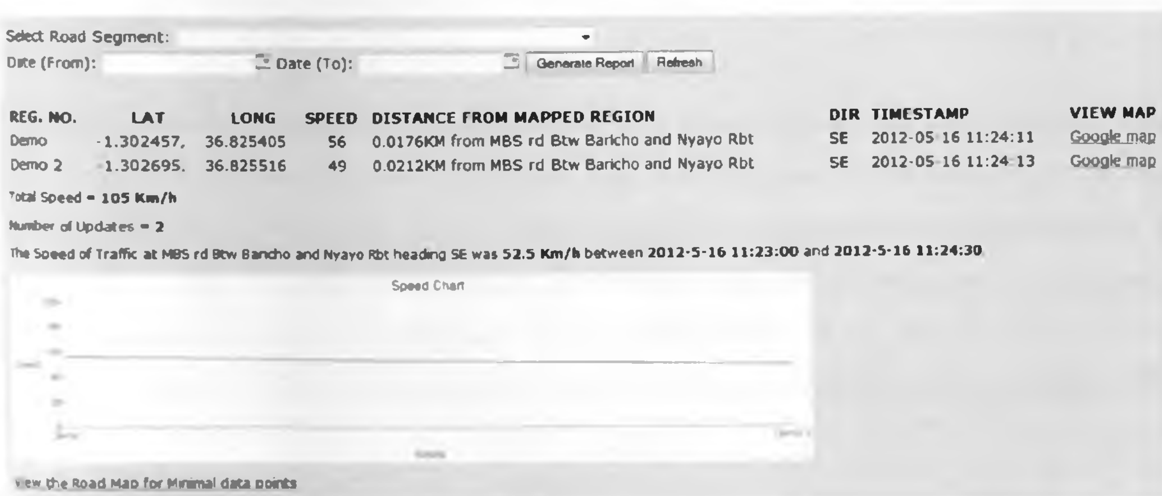
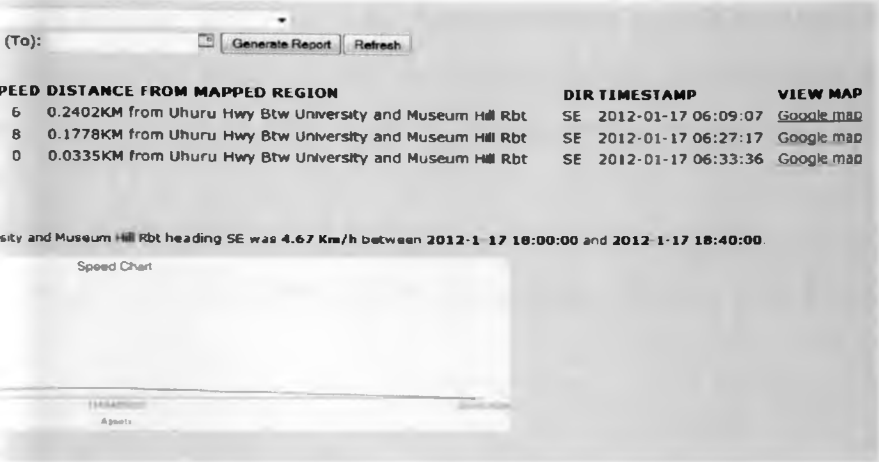


Figure 4.5: Output and Speed chart (2012-5-16 11:23:00 to 2012-5-16 11:24:30).

It can be noted from figure 4.4 that two vehicles were detected at Langata road between 748 Plaza and Kungu Karumba junction heading SW where the speed of traffic was 66.5 Km/h. Only two updates were detected with a total speed of 133 Km/h. Figure 4.5 shows a different road segment, Mombasa road between Baricho and Nyayo roundabouts heading SE where the speed of traffic was detected to be 52.5 Km/h with a total speed of 105 Km/h. Only two updates were detected one with a speed of 56 Km/h and the other one with 49 Km/h. A similar convention can be used to interpret the results in figure 4.6.

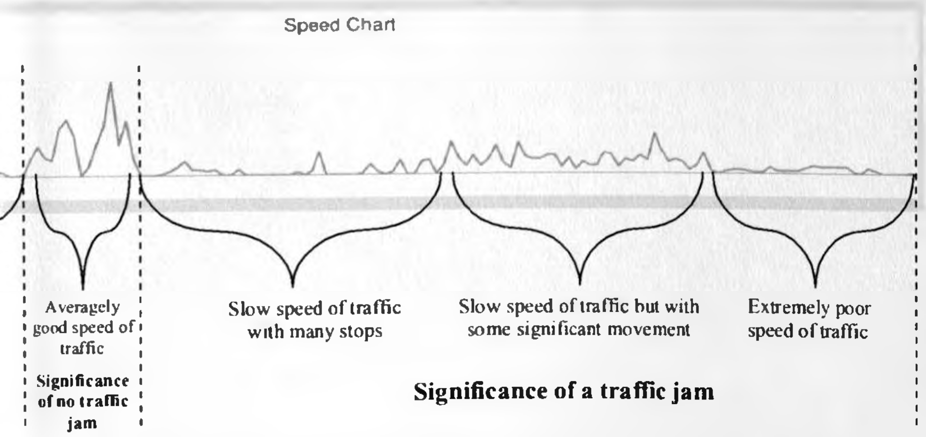


d Speed chart (2012-1-17 18:00:00 to 2012-1-17 18:40:00).

l of Traffic Analysis Results

ded in section 4.3.2 portray a true picture of individual vehicles' movement tected speed, the Google Chart API incorporated in the application automatically (a line graph) to summaries the speeding trend. A speed violation email alert is lue speed compared to the set limit as shown in figure 4.3. This provides a clear ny given driver hence vital information to the Traffic Police Department and any

nterpretation and analysis of a sample speed chart for an analysis done between d 2012-6-11 20:00:00. This is an analysis for the road usage pattern shown in Nairobi CBD to Langata where two major divisions were detected: traffic jam regions with traffic jam can also be divided basing on the wave formed by the e minor divisions for areas with traffic jam i.e. slow speed of traffic with many ffic with some significant movement and extremely poor speed of traffic.



e of a sample speed chart – Nairobi CBD to Langata (2012-6-11 17:45:00 to 2012-6-11 20:00:00).

The speed data for a group of vehicles subjected to the Speed of Traffic algorithm outlined in section 3.9.4 leads to the determination of the speed of traffic for any given road segment within a specified time interval. The output of the speed of traffic analysis as illustrated in section 4.3.2 indicates the following aspects:-

i) The individual vehicle details (the vehicle's registration number, position coordinates, speed, distance from the mapped region, timestamp for the update, option to view the position on Google maps).
ii) The total number of updates involved.
iii) The total speed.
iv) The speed of traffic which is the average speed i.e. the total speed divided by the number of updates.
v) A speed chart (a line graph) that indicates the vehicles speeds.

Using the summary of the speed of traffic data, one can be able to tell the congestion level by checking on the number of updates and/or vehicles involved and the speed of traffic. The vehicles involved depend on the data points detected within the specified road mapped segment. The distance is computed basing on the vehicle's data point with respect to point set in the Road Map table. The computation depends on the given segment's distance in Kilometers. This helps to appropriately narrow down on the point of interest. The Spherical Law of Cosines is heavily used in this analysis.

Section 4.4.1 outlines and discusses a performance analysis that elaborates further on the significance of speed of traffic analysis with two sample cases of predictive speed of traffic analysis.

4.3.3 The Road Usage Analysis

A special algorithm depicted in section 3.9.4 has been developed for use in determination of road usage patterns. This is depicted in form of Google maps for any given point on the map with a specific radius for analysis. Figure 4.8, 4.9, 4.10 and 4.11 show different road usage patterns generated from vehicles that used different roads. The distance from the selected point is computed using the Spherical Law of Cosines.

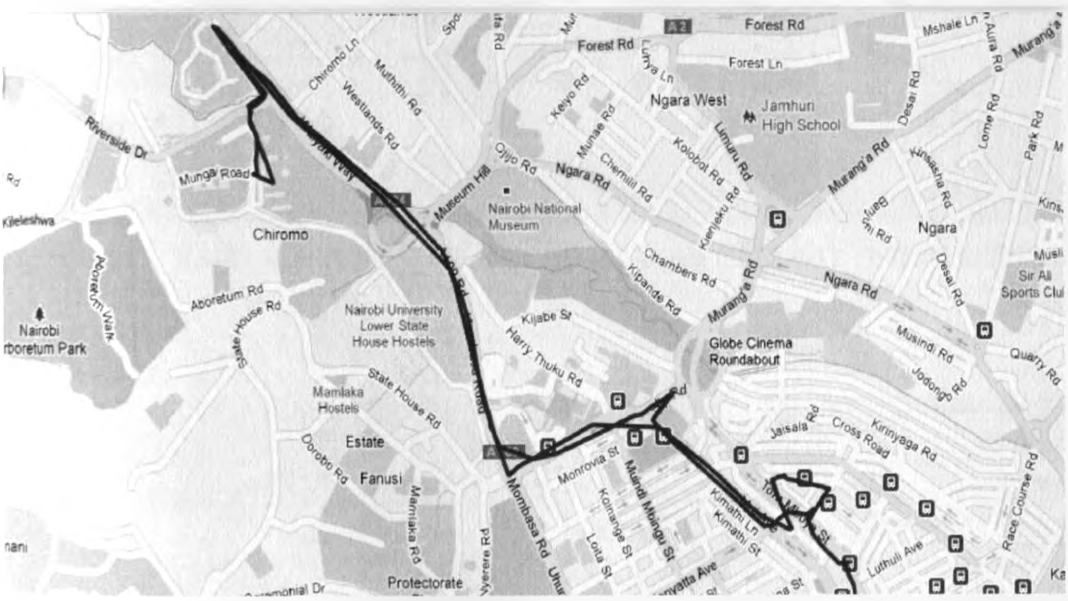


Figure 4.8: Road Usage Pattern – Luthuli Avenue to Chiromo and back (2012-6-13 09:50:00 to 2012-6-13 15:00:00).

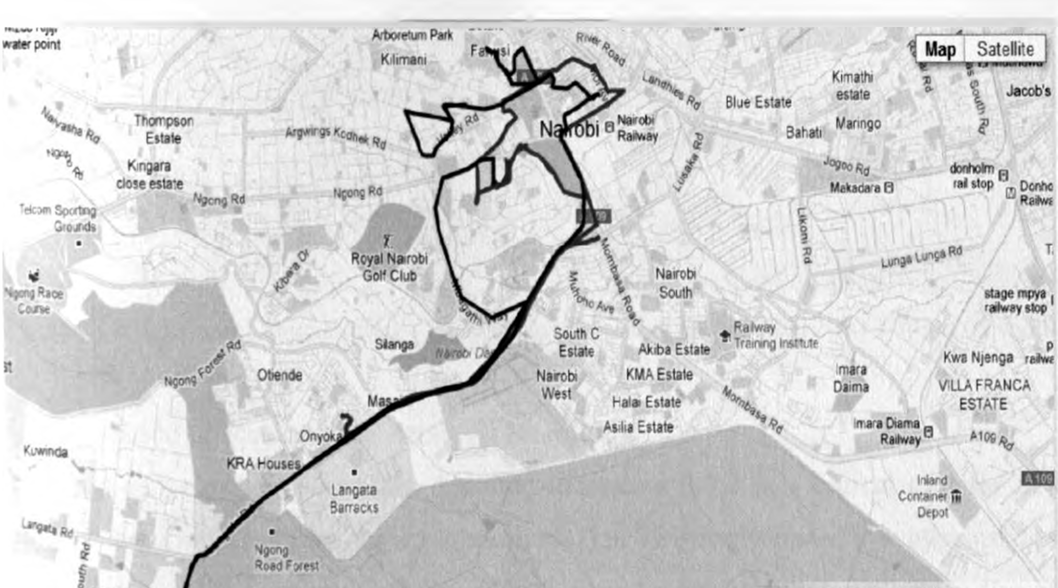


Figure 4.9: Road Usage Pattern – Two vehicles involved (2012-06-20 04:00:00 and 2012-06-22 04:00:00)

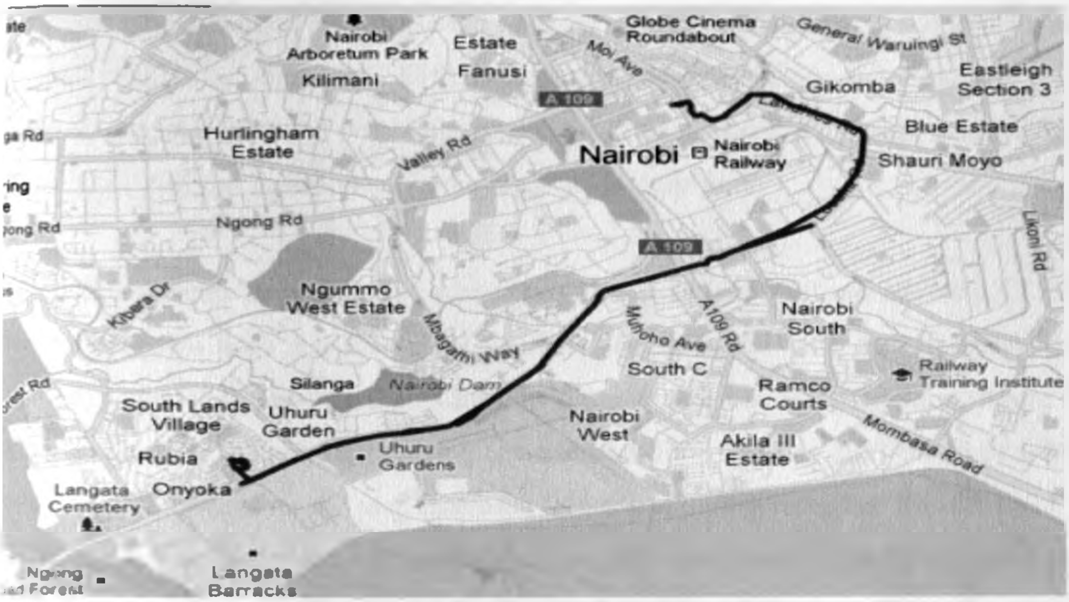


Figure 4.10: Road Usage Pattern – Nairobi CBD to Langata (2012-6-11 17:45:00 to 2012-6-11 20:00:00).

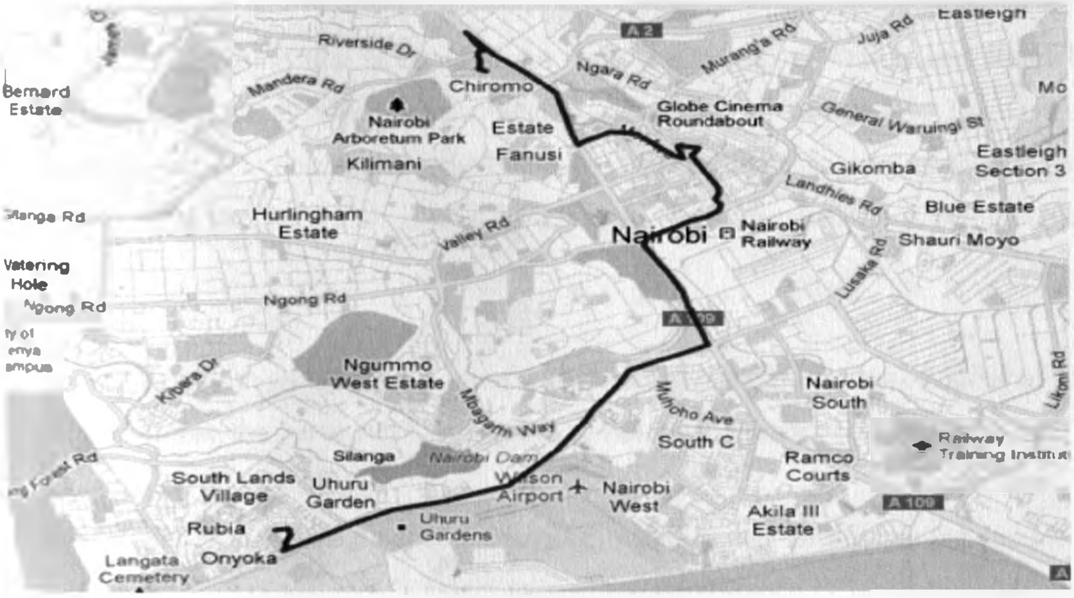


Figure 4.11: Road Usage Pattern – Langata to Chiromo Campus (2012-6-13 09:00:00 to 2012-6-13 11:00:00).

Interpretation of Road Usage Analysis Results

The road usage pattern algorithm depicted in section 3.9.4 is a critical tool for determining the road usage patterns with samples shown in section 4.3.3. This is generated as waypoints for all vehicles found in a given region determined by a specified centroid point, time interval and radius as filter parameters. The selected data points are plotted on Google Maps Overlays (Overlays are objects on the map that are tied to latitude/longitude coordinates, so they move when you drag or zoom the map. Overlays reflect objects that you add to the map to designate points, lines, areas, or collections of objects) using polylines. The output of the road usage patterns as portrayed in figures 4.8 to 4.11 indicates waypoints joined by polylines per vehicle. Different colors as depicted in figure 4.9 indicate the route used by the different vehicles detected within the specified filter parameters. This is achieved by a random color generator algorithm incorporated in the road usage algorithm.

Figure 4.8 outlines an analysis done between 2012-6-13 09:50:00 and 2012-6-13 15:00:00. The trace in red outlines two lines that are almost parallel. This is an indicator that the same route was used from Luthuli Avenue, Nairobi CBD to Chiromo Campus and back. Furthermore, figure 4.9 depicts a road usage pattern for an analysis involving two vehicles done between 2012-06-20 04:00:00 and 2012-06-22 04:00:00. The traces on the map indicated in red and purple represent two different vehicles that were using the region under analysis within the specified time interval. Subjecting the analysis to a region with heavy traffic with all vehicles fit with GPS tracking devices would result in as many traces as the vehicles found in that region. This will provide a clear and useful pattern to planners and road users among other analysts.

The distance attribute that is required for both the Speed of Traffic and Road Usage analysis is computed using the Spherical Law of Cosines depicted in section 3.4. The law uses the specified coordinates and the coordinates in the database to get the distance over a sphere with the Earths radius used. The K-Means algorithm is used in this case where the distance is being subjected to the given centroid.

The KNN algorithm was applied is some way towards the determination of Speed of Traffic as follows:-

1. Determine the parameter k = number of nearest neighbors beforehand. In this case, k can range between 0 to n vehicles.
2. Calculate the distance between the query-instance and all the training samples. In this case, we use the Spherical Law of Cosines to calculate the distance.
3. Sort the distances for all the training samples and determine the nearest neighbors based on the K-th minimum distance. In this case, we determine the minimum distance per vehicle basing on the data point detected among the k parameter.
4. Since this is supervised learning, we get all the categories of the training data for the sorted value which fall under k.
5. Use the majority of nearest neighbors as the prediction value where in this case we are determining the speed of traffic as our prediction.

The K-Means algorithm was also applied in the determination of road usage patterns in a clustering mode as follows:-

1. Determine the centroid coordinates – We specify the coordinates of our centre of interest for analysis.
2. Determine the distances of each data point to the centroid – We subject the centroid to the Spherical Law of Cosines with respect to all vehicles' update data points in question to get the distance between them.
3. Group the data points based on minimum distance – We group the data points basing on the radius initially supplied whereby we are only interested in data points with a distance less than or equal to the radius. At this point, the Google Maps Overlays are then involved for generating the road usage patterns i.e. traces.

4.4 Discussion of Results

The system is meant to operate in a centralized Control Room. All traffic management and final decisions are on the Control Room monitoring team's responsibility. The major tasks at the Control Room include: monitoring the traffic flow by checking on the speed of traffic, analysis of road usage patterns and over speeding email alerts hence estimating, predicting traffic jams and making decision and acting on them accordingly. Analysis on the different traffic information is facilitated by the integrated algorithms outlined in form of flow chart in section 3.9.4 with the corresponding pseudo codes in the appendix section. The algorithms are incorporated with the K-Nearest Neighbor and K-Means algorithms. The Control Room can also find vehicles with any specified registration number, which can be useful for security reasons thus preventing illegal vehicle movement for instance, stolen vehicles or vehicles violating traffic rules.

The GPS Server

The GPS server incorporates a GPS Data Receiver Application developed to captures the following essential information per vehicle per update which was the source of the data for analysis:-

i) Vehicle registration number.
ii) Vehicle Unique identifier.
iii) Vehicle's current location (GPS coordinates).
iv) Vehicle speed received from the GPS tracker's speed sensors.
v) Direction of movement computed as the data streams in.

It should be noted that the interpretation and transformation varies from device to device basing on the manufacturers protocol used.

Algorithms Design and Development

Developing a correct algorithm can be a significant intellectual challenge. Flowcharts and pseudo-codes which are the most widely used notations for developing algorithms were employed in the development of the two algorithms. This was done independent of the programming language that was used to implement them. Section 3.9.4 outlines the flow charts for the two algorithms where the Stepwise Refinement Methodology was used to generate the algorithms' pseudo-codes outlined under appendix B. Section 3.8 outlines the Stepwise Refinement Methodology used.

The design of the algorithms considered the following four key algorithm design aspects:-

i) Preciseness of the algorithms.

ii) The algorithms are executable.

iii) All possible circumstances are handled.

iv) Termination of the algorithms.

A discussion on the two algorithms is as follows:-

1. The Speed of Traffic Algorithm

The speed of traffic algorithm outlined in section 3.9.4 works perfectly by utilizing the captured data parameters where it detects the number of vehicles in any given road segment, their direction of movement, the total speed and the total number of updates. It then computes the average speed which serves as the speed of traffic. This gives traffic flow information for the given mapped road segment. The algorithm is potential for use if subjected to a set of GPS data containing coordinates, speed, direction and timestamp parameter. In this study, the algorithm was subjected to both historical data and the test data collected after the development of the GPS server. The predictive analysis of speed of traffic discussed in section 4.4.1 fully relied on the speed of traffic data generated through the utilisation of the algorithm.

2. The Road Usage Algorithm

The Road Usage pattern algorithm outlined in section 3.9.4 works perfectly by utilizing the captured data parameters whereby it detects vehicles within a given radius and plots the routes they used in form of polylines. Different colors in each case portray different vehicles. This is termed as a route player that depicts the road usage pattern on Google maps. The algorithm is potential for use if subjected to a set of GPS data containing coordinates and timestamp parameter. The sample analysis outlined in section 4.3.3 fully relied on GPS data generated through the utilisation of the algorithm.

4.4.1 Performance analysis basing on results

To rate the performance of the system, predictive and model accuracy tests and analysis were subjected to sample data with conclusions made as follows:

1. **Predictive Accuracy.**

Traffic speed and jams can be termed as dynamic attributes. They can be both predictable and unpredictable basing on the situation at hand. A traffic status (speed of traffic) change for a period of 10 minutes can be so magnificent that predicting it may pose a challenge. In Kenya for example, it has been noted that there is a high possibility of heavy traffic jams whenever it rains hence the model may not predict the traffic under such natural circumstances. Peak and off-peak hours can be predicted since this depends on the hour of the day which always tends to be consistent. The system will permit users to analyse and generate reports for traffic flow for instance speeds of traffic. The following two cases give sample predictive analysis for the speed of traffic for two different road segments:-

Case 1: A Predictive Speed of Traffic

Take a case for Uhuru Highway between Baricho and Haile Selassie Roundabouts. Table 4.2 and 4.3 analyses the Speed of Traffic for the road segment. GPS traffic data gathered in the month of June, 2012 was used as training data sets while May and July, 2012 data sets were used as test data.

Training Data Set (June, 2012)						
Date	**Time bracket**	**Total Speed (Km/h)**	**No. of Updates**	**No. of Vehicles**	**Speed of Traffic(Km/h)**	**Set's Speed of Traffic(Km/h)**
11/06/12	09:00 – 10:00	36	3	2	12	**7**
12/06/12	09:00 – 10:00	61	6	1	10.17	
13/06/12	09:00 – 10:00	15	3	1	5	
22/06/12	13:00 – 14:00	15	18	1	0.83	
Test Data Set (May and July, 2012)						
16/05/12	08:00 – 09:00	27	16	2	6.63	**6.51**
05/07/12	09:00 – 10:00	27	5	1	5.4	
06/07/12	16:00 – 17:00	72	8	1	9	
12/07/12	14:00 – 15:00	30	6	1	5	

Table 4.2: Uhuru Highway between Baricho and Haile Selassie Roundabouts heading NW

The analysis from table 4.2 indicates slow speeds of traffic ranging between 0 and 12 Km/h per given analysis time bracket. The speed of traffic for the test data set resulted to 6.51 Km/h which was closer to the 7 Km/h value for the training data set. Under normal circumstances getting into Nairobi town via Uhuru Highway is a common heavy traffic prone area. Traffic speeds ranging between 0 and 12 Km/h can be concluded to be a true picture on the ground for Uhuru Highway between Baricho and Haile Selassie Roundabouts heading NW (i.e. getting into town).

Training Data Set (June, 2012)						
Date	**Time bracket**	**Total Speed (Km/h)**	**No. of Updates**	**No. of Vehicles**	**Speed of traffic(Km/h)**	**Set's Speed of Traffic(Km/h)**
12/06/12	09:00 – 10:00	7	2	1	3.5	**16.66**
12/06/12	17:00 – 18:00	24	2	1	12	
13/06/12	18:00 – 19:00	58	7	1	8.29	
21/06/12	14:00 – 15:00	66	2	1	33	
23/06/12	15:00 – 16:00	53	2	1	26.5	
Test Data Set (May and July, 2012)						
16/05/12	11:00 – 12:00	127	5	2	25.4	**28.91**
30/05/12	14:00 – 15:00	53	4	1	13.25	
04/07/12	09:00 – 10:00	71	2	1	35.5	
06/07/12	07:00 – 08:00	83	5	1	41.5	

Table 4.3: Uhuru Highway between Baricho and Haile Selassie Roundabouts heading SE

The analysis shown in table 4.3 indicates slightly high speeds of traffic ranging between 12 and 50 Km/h per given time bracket. The speed of traffic for the test data set resulted to 28.91 Km/h which was higher than 16.66 Km/h for the training data set. Under normal circumstances getting out of Nairobi town via Uhuru Highway is a region that experiences fairly good speeds of traffic with minimal stops as compared to the scenario of getting into Nairobi town. Traffic speeds ranging between 12 and 50 Km/h can be concluded to be a true picture on the ground for Uhuru Highway between Baricho and Haile Selassie Roundabouts heading SE (i.e. getting out of town).

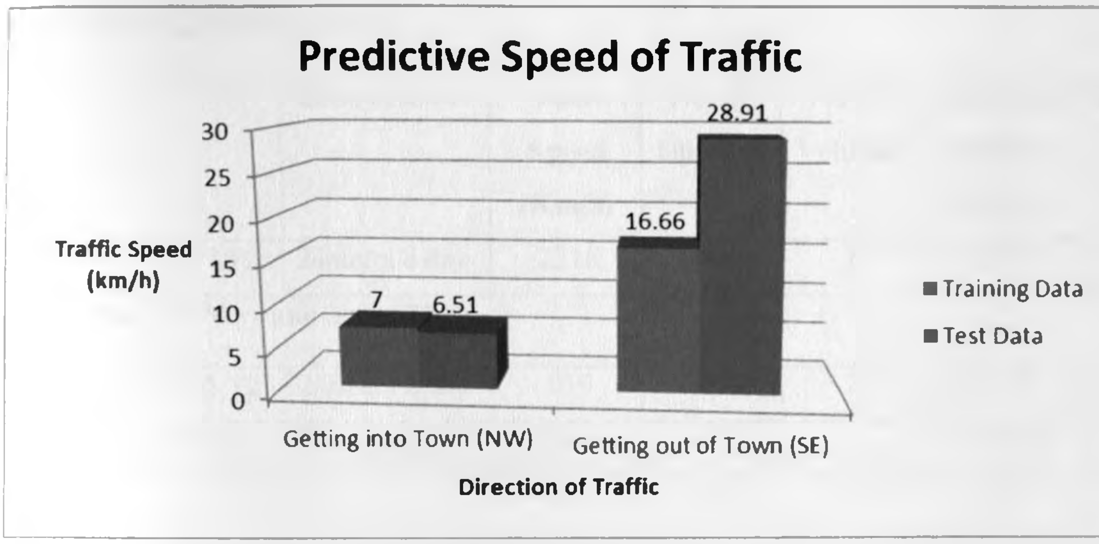


Chart 4.1: Uhuru Highway between Baricho and Haile Selassie Roundabouts

Chart 4.1 depicts a predictive speed of traffic summary for the sample case 1 of Uhuru Highway between Baricho and Haile Selassie Roundabouts. The chart has two main sections i.e. Getting into Town (NW) and Getting out of Town (SE) with a clear comparison of the training and test data sets used. The data set used was similar to the set used for the analysis in figure 4.2 and 4.3. It can be evident that getting into town is much slower than getting out of town.

Case 2 for a Predictive Speed of Traffic

Take a case for Langata road between Kungu Karumba Junction and the Bomas of Kenya Junction. Table 4.4 and 4.5 analyses the Speed of traffic for the road segment. GPS traffic data gathered in the month of July, 2012 was used as training data while May and June, 2012 data sets were used as test data.

Training Data Set (July, 2012)						
Date	**Time bracket**	**Total Speed (Km/h)**	**No. of Updates**	**No. of Vehicles**	**Speed of Traffic (Km/h)**	**Set's Speed of Traffic (Km/h)**
01/07/12 – 22/07/12	24hours a day	1982	120	1	16.52	**16.52**
Test Data Set (May and June, 2012)						
01/05/12 – 31/05/12	24hours a day	2005	97	2	20.67	**17.13**
01/06/12 – 30/06/12	24hours a day	1182	87	2	13.59	

Table 4.4: Langata road between Kungu Karumba Junction and Bomas of Kenya Junction heading SW

Training Data Set (July, 2012)						
Date	**Time bracket**	**Total Speed (Km/h)**	**No. of Updates**	**No. of Vehicles**	**Speed of Traffic (Km/h)**	**Set's Speed of Traffic (Km/h)**
01/07/12 – 22/07/12	24hours a day	2216	99	1	22.38	**22.38**
Test Data Set (May and June, 2012)						
01/05/12 – 31/05/12	24hours a day	939	40	2	23.48	**24.18**
01/06/12 – 30/06/12	24hours a day	945	38	2	24.87	

Table 4.5: Langata road between Kungu Karumba Junction and Bomas of Kenya Junction heading NE

The analysis in table 4.4 indicates slow speeds of traffic along Langata road while moving from Kungu Karumba Junction to the Bomas of Kenya Junction i.e. heading SW as compared to moving in the opposite direction as depicted in table 4.5. Under both circumstances the training data set's speed of traffic value was found close to the test data set's speed of traffic value. This created a better and clear picture of the speeds as expected to be along Langata road between Kungu Karumba Junction and Bomas of Kenya Junction segment in either direction of movement.

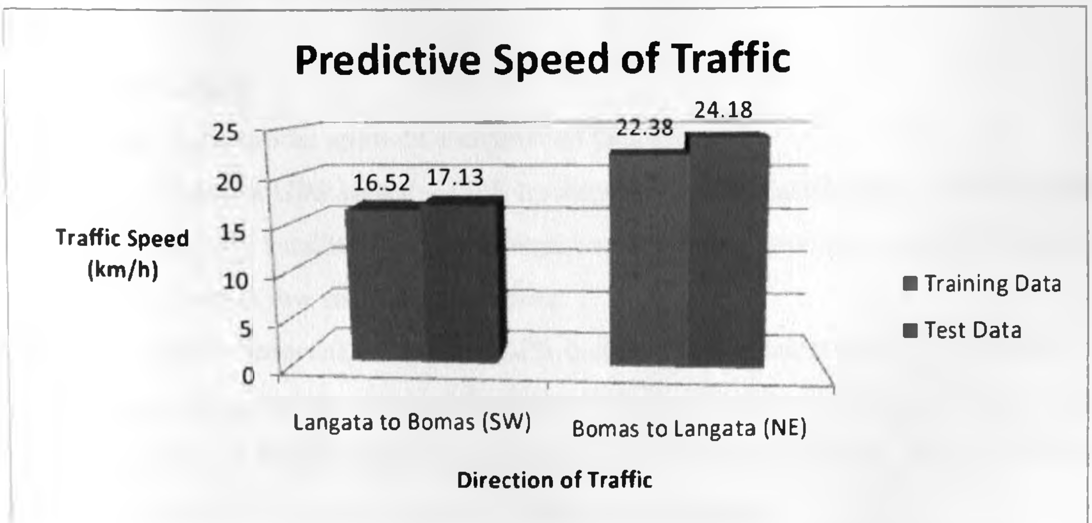


Chart 4.2: Langata road between Kungu Karumba Junction and Bomas of Kenya Junction

Chart 4.2 depicts a predictive speed of traffic summary for the sample case 2 of Langata road between Kungu Karumba Junction and the Bomas of Kenya Junction. The chart has two main sections i.e. Langata to Bomas (SW) and Bomas to Langata (NE) with a clear comparison for the training and test data sets used. The data set used was similar to the set used for the analysis in figure 4.4 and 4.5. It can be evident graphically that moving from Langata to the Bomas of Kenya junction (SW direction) is slower than moving from the Bomas of Kenya junction to Langata (NE direction).

A similar convention outlined by the two cases can be applied to any given road segment with the data gathered by the GPS server developed. It should be noted that the probability of experiencing faster or slower speeds for either road segment as summarized by chart 4.1 and 4.2 may not always be true basing on other unpredictable factors some of which may include:-

1. **Accidents** – In case of an accident, traffic speed will definitely be slowed.
2. **Weather** – As a normal phenomenon in Kenya, whenever it rains, traffic jams increase probably due to poor visibility.
3. **State functions** – Any instances involving state functions for instance where presidential escort services are involved, result in periodical closure of some roads hence congestions leading to slower speeds of traffic.
4. **Driver behaviors** – For instance, overlapping tends to slow the speed of traffic since it results in confusions on the roads.
5. **State of roads** – Poor states of roads result in slower speeds e.g. unrepaired pot holes.
6. **Construction works** – Road works may lead to diversions and/or temporal closure of some lanes hence affecting normal traffic speeds.

2. **Model Accuracy.**

The accuracy of the model relies on a number of factors:-

i) **The available GPS signal** – GPS trackers rely on the availability of GPS satellites in position. At least six satellites make the signal stronger hence a more valid position is established as compared to few satellites in position.

ii) **The update interval set for the GPS trackers** – For better analysis of speed of traffic and road usage patterns, the interval between updates should be set too short possibly less than 30 seconds. A longer interval duration results in a degraded road usage pattern incorporated with poor and/or inaccurate speeds of traffic computation.

iii) **The time bracket for analysis** – Traffic status change for any given road segment can be so significant at some point hence a longer time bracket for analysis of speed of traffic may lead to inaccurate results.

iv) **The GSM network coverage** – The better the network, the more accurate the interval of update such that no updates will be missed and/or lost.

4.5 The Model for the System

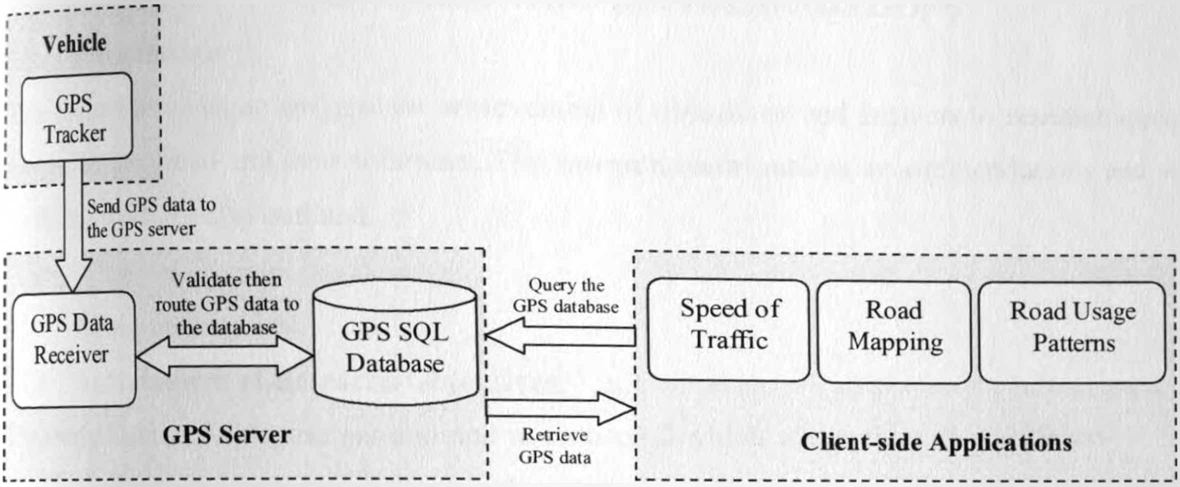


Figure 4.12: The model for the system.

Figure 4.12 depict the resulting conceptual model for the GPS Server with road mapping for traffic analysis. The model has three main sections i.e. the GPS server, client-side applications and the vehicle. The GPS server encompasses the GPS Data Receiver and an SQL database while client-side applications encompass the speed of traffic, road mapping and road usage pattems. The vehicle is installed with the GPS tracker that is configured to send updates to the GPS server. All the aspects in the model were considered towards the achievement of the overall research study results.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

This chapter presents an analysis on achievement of objectives and answers to research questions with the challenges faced and their solutions. The research contributions, recommendations and suggestions for future work are also outlined.

5.1 Achievements

5.1.1 Achievement of Research Objectives

The study had three objectives as outlined in section1.2 which were achieved as follows:-

1. **To research and explore on the developments in GPS technology and GPS data mining.**

This was a state of art objective that majorly targeted an in-depth understanding and knowledge of GPS technology and GPS data mining. The chapter on literature review gave room for the evaluation and analysis of related work with respect to the application of GPS technology in other areas rather than the common security industry. Expounding on the literature review with its division in three sections opened up a clear picture for the developments incorporating GPS technology in tracking systems, data mining and traffic control and management.

2. **To develop a GPS TCP server application that listens to GPS trackers' data, transforms and routes it to a centralised database.**

The developed GPS Data Receiver application addressed the objective through the utilisation of PHP sockets. The application was made in such a way that a server program creates a specific type of socket that is used to listen for client requests (GPS trackers data); in the case of a connection request, the application creates a new socket through which it will receive data from the GPS tracker using an input stream, interpret and/or transform then route it to the database.

3. **To develop client-side applications for analysis of traffic speed and road usage patterns by incorporating a road mapping concept. Speed of traffic and road usage pattern algorithms will be developed in this respect.**

A web based application was developed and setup online to facilitate analysis of the data received by the GPS TCP server application. Two algorithms: the speed of traffic and road usage pattern algorithms developed also addressed the objective with solutions to research questions 2, 3 and 4 as outlined in section 1.3.

5.1.2 Answers to Research Questions

The study had five research questions as outlined in section1.3 which were answered as follows:-

1. **What data formats do GPS trackers use and how can it be interpreted, transformed then routed and stored in a database?**

The research study established that GPS trackers' manufacturers use the NMEA 0183 standard whose 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 a fraction of which apply directly to GPS devices. The study settled on GPS device that use $GPRMC sentence that contains the key information required i.e. position, speed, and unique identifier data. To interpret the $GPRMC sentence, the PHP list() function is first used to split the data sentence into independent variables. Other different sets of PHP functions and computations are used to interpret the variables separately since different outputs and transformations are required before routing the data to the database. The format of the $GPRMC sentence is as outlined in section 4.3.1.

2. **What attributes of raw GPS data are useful for traffic analysis and how can they be transformed to user-friendly format when subjected to road mapping and data mining for analysis?**

The research established that traffic analysis with respect to GPS data relies on the following attributes:-

i) Unique vehicle identity mapped to a registration number.
ii) Position coordinates (latitude and longitude).
iii) Speed.
iv) Timestamp for the update.

One more parameter not send together with the $GPRMC sentence and that was a key requirement for traffic analysis is the direction of movement. The SQL lead and lag concept and a direction trigger were used to compute the direction of movement.

3. **What formula can be used to establish vehicles in a given mapped road segment in a move towards the determination of speed of traffic for any given road?**

The Spherical Law of Cosines was found to be the optimal formula for computing the distance between two coordinates with the radius parameter set to the Earth's radius. The Spherical Law of Cosines incorporates six trigonometric functions in its computation hence results in a more accurate distance value that can be as small as 1 metre as opposed to the Haversine and the Equirectangular Approximation formula. The formula was incorporated in the Speed of Traffic and Road Usage Pattern algorithms hand in hand with the K-Nearest Neighbor and K-Mean data mining algorithms.

4. **What GPS data mining algorithms can be generated for road traffic and road usage pattern analysis on the Kenyan roads using massive GPS data in a centralized database?**

The Speed of Traffic algorithm was developed for the analysis of traffic speeds while the Road Usage Pattern algorithm outlines traces on Google Maps indicating vehicle movement patterns. The K-Nearest Neighbor and K-Mean data mining algorithms were considered towards cluster analysis. The algorithms worked perfectly with the outputs as analysed in chapter four, section 4.3. The distance parameter required for both algorithms was computed using the Spherical Law of Cosines.

5. **Can a centralized GPS data bank with road mapping and GPS data mining substitute the tachograph option proposed in November, 2011 by the Kenyan Ministry of Transport to be used for traffic management for Public Service Vehicles after the collapse of the speed governor rule? Can this add value towards the achievement of the Kenyan Vision 2030?**

The establishment of a centralized GPS data bank offers a better opportunity for the development of different applications to query the data for different viable purposes with the authority from the Government who should own the data hence increasing efficiency in the Public Service Vehicle sector, Ministry of Transport, Ministry of Roads and Public Works, Traffic Police Department, road users among others. The GPS data can be easily accessed both in real-time and historically for any analysis as opposed to the tachograph that relies on data provision by the driver upon request. Tachograph data may be easily tampered with by a guilty driver to distort the evidence.

In a move by the Kenyan Government to achieve Vision 2030, there is growing need for digitalization and automation of most operations. GPS technology adoption in traffic analysis can be a better move towards traffic control and management. GPS speed data can substitute the speed governor option by far basing on the fact that traffic speed violation data is automatically captured and stored in addition to the generation of speed graphs for evidence. Automatic over speeding email alert messages serves as an added advantage towards detection and acting on traffic violation. For instance, the recent speed limit of 50 Km/h that was set out for Thika Road Super Highway users due to the increase in road accidents that led to erection of bumps on the road could have been alleviated by the GPS system with email alert for over speeding set to 50 Km/h together with a thorough analysis of other speed data for the road users.

5.2 Limitations

The major constraint was the small sample size that could not allow for a clear demonstration of real-time speed of traffic analysis. Only two devices were used for testing hence could not provide a better room for real-time analysis.

5.3 Challenges

GPS technology being a relatively new area particularly with respect to GPS data mining, a number of challenges were experienced in the course of the research study. The first challenge was that of learning and understanding the data formats used by GPS trackers basing on the fact that different manufacturers use different protocols. This was mitigated through constant research with the involvement of online discussion forums.

Maps to be incorporated in the system was also another major challenge for instance, Google maps license has to be purchased for one to be able to use maps with rich features. The challenge was mitigated through the use of OpenStreetMaps and Google Maps free versions designed for non-commercial use. The only disadvantage is that these maps are not detailed and may lack some points.

Another major challenge experienced was that GPS tracking devices' data sentences did not include a direction parameter yet it was a key requirement in the determination of the speed of traffic. Furthermore the data sourced for assistance in historical analysis for the data mining part of the study also lacked the direction parameter. The challenge was mitigated by employing the lead and lag concept of SQL statements to generate the direction basing on coordinate changes. A direction trigger was developed to help generate the direction parameter for any incoming data point basing on the last data point received per asset unique identifier.

Lastly the historical data had a larger time interval (5 minutes) set between updates hence distorting the maps while attempting to determine the road usage patterns. The challenge was mitigated by not subjecting the data to road usage analysis. The data was only used to determine the speed of traffic since this analysis doesn't heavily rely on the update intervals. Note that, a 5 minutes time interval is too large hence if road usage patterns are subjected to data with such an interval, then the trace pattern gets out of the road sections of the map. This results in a totally distorted pattern.

Iterative agile methodology was used to bridge any emerging gaps noticed during the development phase. Moreover, usability testing and evaluation was also employed constantly to help detect and solve any usability related aspects of the system.

5.4 Research Contribution

The application of GPS and GIS in traffic analysis is proving to be the most effective solution compared to other existing traffic management methods like safety cameras, human inspection, speed governors and tachographs. 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 GPS 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 any future analysis.

The research resulted in the development of a model for GPS system with road mapping and data mining for traffic analysis. This is a major contribution towards automation of traffic management and analysis for any given Nation. Additionally, two algorithms have been developed in the course of the study i.e. the Speed of Traffic algorithm and the Road Usage Pattern algorithm both of which are vital towards the analysis of traffic conditions on any given road segment and region. Automatic email alerts on over speeding drivers also provide better and clearer evidence to any Traffic Police Department as opposed to the speed governor and tachograph options.

5.5 Recommendations

The results from the GPS traffic analysis show considerable variations in both the scope of impacts and effectiveness of the measures. For successful implementation of traffic management and control measures, not only should the technical perspectives of the measure be properly considered, but also the integration with existing infrastructure and policies. Based on the test results of the concept, it is recommended that:

1. GPS traffic management are effective, reliable and efficient, with a flexible scale of implementation at relatively small investment, which should be considered as one of priority traffic management and control measures compared to other different types of traffic management and control measures.
2. Any Nation intending to adopt the model must put in place a policy and/or law prompting all vehicle owners to fit their vehicles with GPS trackers. Furthermore the taxes on importation of the devices should be scraped with prices regulated to help make the devices affordable to the citizens.
3. Some traffic management measures use sophisticated systems such as safety cameras that cannot be installed all over since they require substantial infrastructure support and fine tuning which turns out to be costly to implement. Furthermore, safety cameras can only serve well in cities i.e. a smaller region. GPS systems are less sophisticated to implement hence the optimal options.

5.6 Suggestions for Further Research

Future research may lie in investigating several enhancements to the current implementation, including automatic traffic prediction with auto alerts to traffic police and drivers detected closer and/or heading towards congested areas. This prompts for the incorporation of mobile SMS alerts as opposed to email alerts.

Further future research should focus on enhancing the system towards driver behavior analysis and reporting with specific real-time warnings and advanced mapping for instance, the ability to send a real-time warning alert to the reckless and over speeding drivers.

References

Agamennoni, G., Nieto, J. and Nebot, E. 'Mining GPS Data for Extracting Significant Places', Paper for Australian Center for Field Robotics, Australia.

Auld, J., Williams, C., Mohammadian, A. and Nelson, P. (2008) 'An automated GPS-based prompted recall survey with learning algorithms', Paper for the Journal of Transportation Letters, Chicago.

Battelle Transport Division, 1997, Lexington Area Travel Data Collection Test, Final report prepared for the FHWA.

Biem, A., Bouillet, E., Feng, H., Ranganathan, A., Riabov, A., Verscheure, O. (2010) 'Real-Time Traffic Information Management using Stream Computing', IBM TJ Watson Research Center, New York, USA.

Ehmke, J. and Mattfeld, D. (2010) 'Data allocation and application for time-dependent vehicle routing in city logistics', European Transport n. 46, pp. 24 – 25, Germany.

Hasan, K., Rahman, M., Haque, A. L., Rahman, M. A., Rahman, T. and Rasheed, M. M. (2009) 'Cost Effective GPS-GPRS Based Object Tracking System', Proceedings of the International MultiConference of Engineers and Computer Scientists 2009, Vol I IMECS 2009, 18 – 20 March, Hong Kong.

Jan, O., Horowitz, A., Peng, Z. (2000) 'Using Global Positioning System Data to Understand Variations in Path Choice', Transportation Research Record 1725, TRB, National Research Council, Washington, D.C., pp.37-44.

Kardashyan, A. (2011) 'New Concept of the Urban and Inter-Urban Traffic', Proceedings of the World Congress on Engineering and Computer Science 2011, Vol II, WCECS 2011, 19 – 21 October, San Francisco, USA.

Kochan, B., Bellemans, T., Janssens, D. and Wets, G. (2006) 'Dynamic activity-travel data collection using a GPS-enabled personal digital assistant', Proceedings of the ninth International Conference on AATT, pp.319-324.

Marketos, G. (2009) 'Mobility Data Warehousing and Mining', 09 Ph D Workshop, 24 August, Lyon, France.

Meier, J. (2008) Software Engineering, Project Management and Effectiveness, Agile Architecture Methods, J. D. Meier's blog, 5 November 2008. Available at: http://blogs.msdn.com/jmeier/archive/2008/11/06/agile-architecture-method.aspx (Accessed 10th December 2011).

Mukesh, P. and Muruganandham (2010) 'Real Time Web based Vehicle Tracking using GPS', World Academy of Science, Engineering and Technology 61.

Owusu, J., Afukaar, F. and Prah, B. E. K. (2006) 'Urban Traffic Speed Management: The Use of GPS/GIS', Shaping the Change, XXII FIG Congress, Munich, Germany 8-13.

Thianniwet, T., Phosaard, S., member, IAENG, and Pattara-Atikom, W. (2009) 'Classification of Road Traffic Congestion Levels from GPS Data using a Decision Tree Algorithm and Sliding Windows', Proceedings of the World Congress on Engineering 2009, Vol I, WCE 2009, 1 – 3 July, London, U.K.

Tripathi, J. (2010) Algorithm for Detection of Hot Spots of Traffic through Analysis of GPS Data. Thesis submitted in partial fulfillment of the requirement for the award of degree of Master in Engineering in Computer Science and Engineering. Thapar University, Patiala.

Williams, L. (2007) 'A Survey of Agile Development Methodologies', pp. 209 - 227.

Ye, Q., Chen, L. and Chen, G. (2009) 'Personal Continuous Route Pattern Mining', Journal of Zhejiang University Science A, 2009 10(2): pp. 221 – 231, Hangzhou 310027, China.

Yoon, J., Noble, B. and Liu, M. (2007) 'Surface Street Traffic Estimation', Electrical Engineering and Computer Science University of Michigan, Ann Arbor, MI 48109-2122, MobiSys ' 07, 11 – 14 June, San Juan, Puerto Rico, USA.

Appendices

Appendix A: User Manual

Starting the System

To launch the system:

- Open any browser
- Type "localhost" as the URL

Login

On the login screen:

- Type the username and password then click login
- Accept the security message that appears or logout

Once logged in the homepage will be a map showing the positions of all vehicles and their distribution in different regions. Zoom in and out narrows down to a given area of view.

The side menu has three main categories of items outlined as follows:

1. **Mapping**

There are two map types: Google Maps and OpenStreetMaps.

- Select a map type you want to use.
- Select a vehicle you want to view from the drop down menu then click "Show on Map".
- Pointing the cursor to a vehicle marker on the map pops out a window with the given vehicle's details.
- The "Refresh" button resets the entries in the fields.

2. **Reports**

There are two sections: Current Status and History

a) Current Status
   - The output area will show the real-time position for all vehicles.
   - "Export to Excel" allows you to export the details to an excel file.

b) History
   - Select a vehicle you want to view from the drop down menu.
   - Specify the time bracket for analysis using the date picker for both "date from" and "date to" fields.
   - Click "Generate Report"
   - The "Refresh" button resets the entries in the fields.

3. Traffic speed analysis

There are three sections: Real-time Traffic Speed, Selective Traffic Speed and Historic Traffic Speed

a) Real-time Traffic Speed
   - Select a road segment for analysis from the drop down menu.
   - Click "Analyse Traffic Speed"
   - The "Refresh" button resets the entries in the fields.

b) Selective Traffic Speed
   - Select a road segment for analysis from the drop down menu.
   - Specify the time bracket for analysis using the date picker for both "date from" and "date to" fields.
   - Click "Generate Report"
   - The "Refresh" button resets the entries in the fields.

c) Historical Traffic Speed – This is similar to "Selective Traffic Speed" however it is subjected to historical data specifically sourced to assist in demonstration of the data mining concept.
   - Select a road segment for analysis from the drop down menu.
   - Specify the time bracket for analysis using the date picker for both date from and date to fields.
   - Click "Generate Report"
   - The "Refresh" button resets the entries in the fields.

4. Road usage

There is only one option: Road Usage

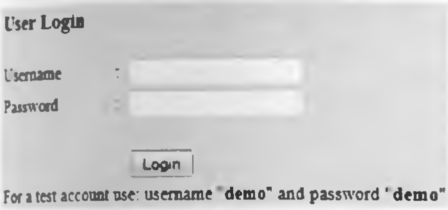
- Specify the coordinates for the centroid: latitude and longitude.
- Specify the radius for analysis in Kilometers.
- Specify whether to analyse all or an individual vehicle using the drop down menu.
- Specify the time bracket for analysis using the date picker for both date from and date to fields.
- Click "Play the Route"
- The "Refresh" button resets the entries in the fields.

The top menu list has four menu items:-

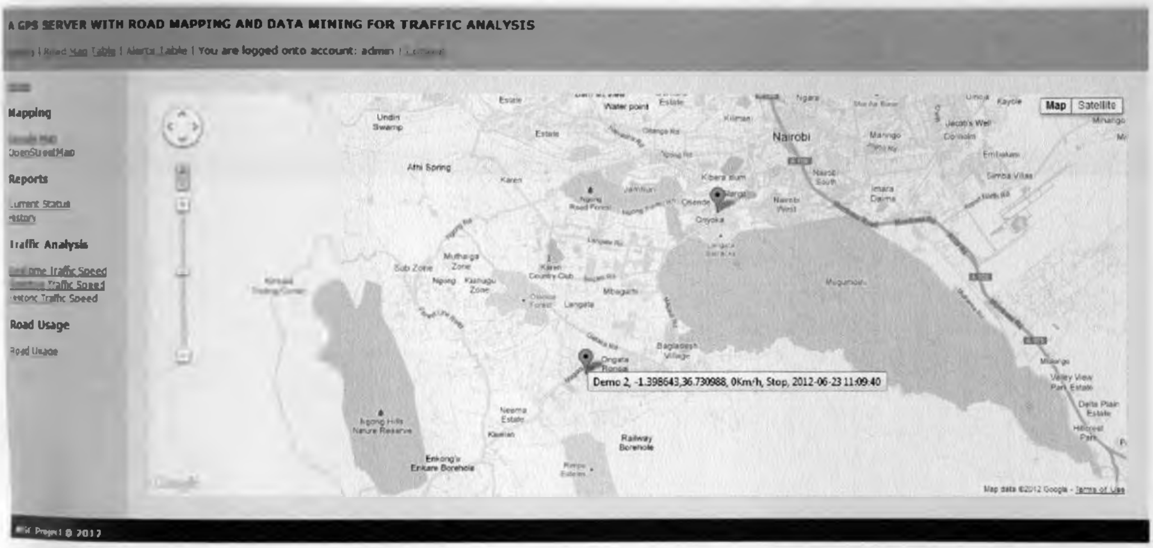
1. **Assets**
   - In the form that appears, enter the asset details for the asset to be added.
   - Click "Add"
   - The "Refresh" button resets the entries in the fields.
   - "Export to Excel" allows you to export the details to an excel file.
2. **Road Map Table**
   - In the form that appears, enter the road map details for the road segment to be added. The distance filed should be in Kilometers and should be the total distance for the entire road segment.
   - Click "Add"
   - The "Refresh" button resets the entries in the fields.
   - "Export to Excel" allows you to export the details to an excel file.
3. **Alerts Table**
   - In the form that appears, enter the maximum speed limit and email address to receive the alert.
   - Click "Update"
   - The "Refresh" button resets the entries in the fields.
4. **Users**
   - In the form that appears, enter the Username and Password.
   - Click "Add"
   - The "Refresh" button resets the entries in the fields.
5. **Logout**
   - Click "Logout" to sign out of the system.

Appendix B: Samples

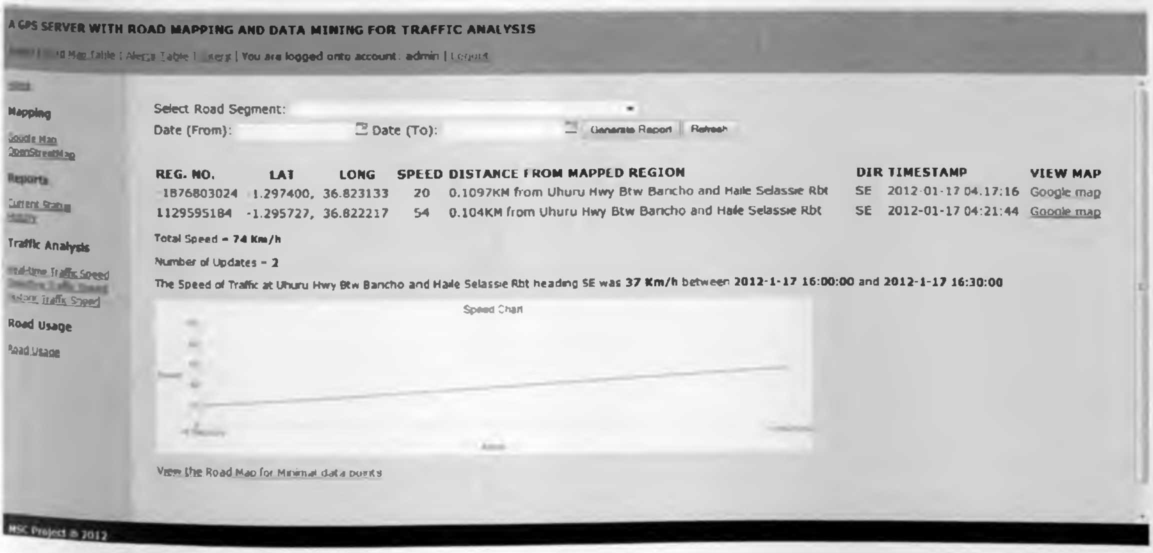
Login Screen



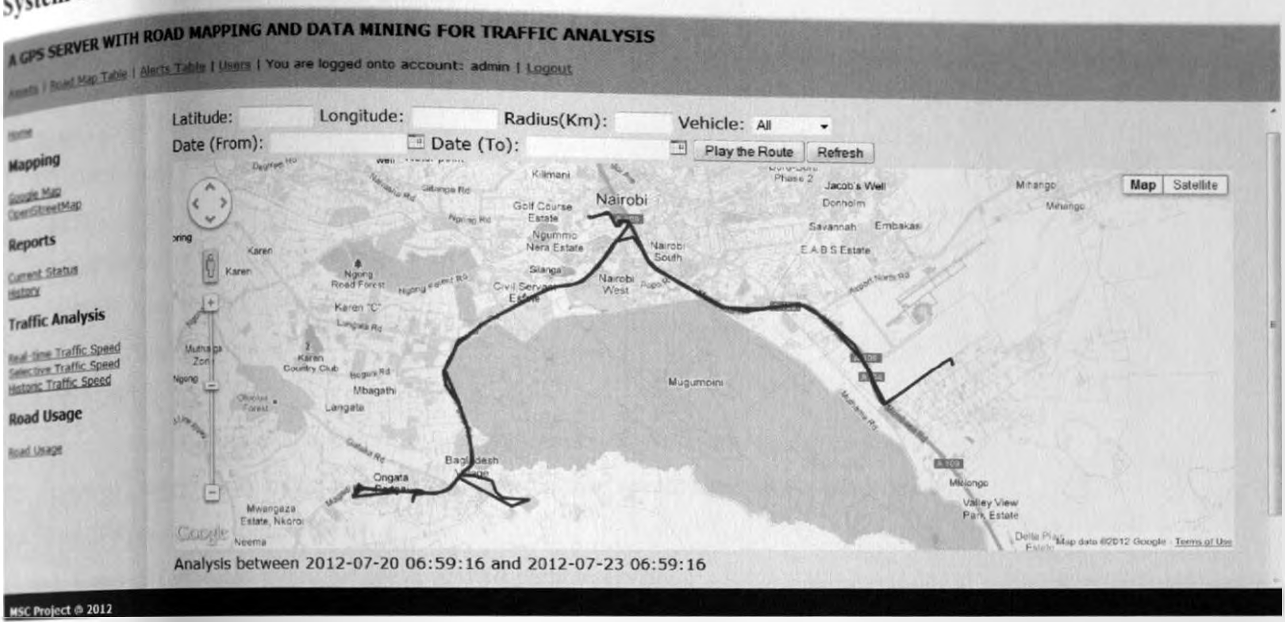
System Screen (Homepage)



System Screen (Historic Speed of Traffic)



System Screen (Road Usage Pattern)



Sample Code

Pseudo Code for the Speed of Traffic Algorithm (section 3.9.4)

1. Declare Counter and TotalSpeed and initialize them to 0.
2. IF the analysis is real-time THEN
    2.1 Get the road segment.
3. ENDIF
4. ELSEIF the analysis is not real-time THEN
    4.1 Get the time bracket for analysis.
    4.2 Get the road segment.
5. END ELSEIF
6. Extract the road segment's data: *segment name, direction, coordinates and distance.*
7. Extract GPS data records: *asset's unique ID, coordinates, speed, direction and timestamp.*
8. IF no GPS records are found THEN
    8.1 Display "No records found for the specified parameters".
    8.2 IF another analysis THEN
        8.2.1 GOTO 2.
    8.3 ENDIF
    8.4 ELSEIF no other analysis THEN
        8.4.1 Stop.
    8.5 END ELSEIF
9. ENDIF
10. ELSEIF GPS data records are found THEN
    10.1 WHILE not end of records
        10.1.1 Compute the distance: *Subject the extracted coordinates and the road segment's coordinates to the Spherical Law of Cosines*
        10.1.2 IF distance > the specified value THEN
            10.1.2.1 The record is eliminated from the analysis.
            10.1.2.2 GOTO 10.1.
        10.1.3 ELSEIF distance < the specified value THEN
            10.1.3.1 Evaluate the direction of movement with respect to the road segment.
            10.1.3.2 IF the direction is not valid THEN
                10.1.3.2.1 GOTO 10.1.
            10.1.3.3 ENDIF
            10.1.3.4 ELSEIF the direction is valid THEN
                10.1.3.4.1 Counter ++
                10.1.3.4.2 TotalSpeed = TotalSpeed + Speed.
            10.1.3.5 END ELSEIF
        10.1.4 END ELSEIF
    10.2 END WHILE
    10.3 Compute the Average Speed as TotalSpeed divided by Counter.
    10.4 Display the Average Speed and Speed graph.
11. END ELSEIF

Pseudo Code for the Road Usage Pattern Algorithm (section 3.9.4)

1. Determine the centroid coordinates, time bracket and radius for analysis.
2. Extract GPS records basing on the time bracket.
3. IF no GPS records are available THEN
    3.1 Display "No records found".
    3.2 IF you want to do another analysis THEN
        3.2.1 GOTO 1.
    3.3 ENDIF
    3.4 ELSEIF no other analysis THEN
        3.4.1 Stop.
    3.5 END ELSEIF
4. ENDIF
5. ELSEIF GPS data records are found THEN
    5.1 Set the default color code to #FF0000 (red).
    5.2 Identify a unique identifier not analysed.
    5.3 WHILE not end of records for the unique identifier
        5.3.1 Extract the coordinate values for the data point.
        5.3.2 Compute the distance: *Subject the extracted coordinates and the centroid coordinates to the Spherical Law of Cosines.*
        5.3.3 IF distance <= radius THEN
            5.3.3.1 Plot GPS trace/waypoint on the map.
            5.3.3.2 Generate a color code for the next set of records.
        5.3.4 ENDIF
        5.3.5 ELSEIF distance > radius THEN
            5.3.5.1 The record is eliminated from the analysis.
        5.3.6 END ELSEIF
        5.3.7 GOTO 5.3.
    5.4 END WHILE
    5.5 Load and Display the map with the traces/waypoints joined by polylines using the Google Maps Overlays.
    5.6 Infer the road usage pattern.
6. END ELSEIF

The code for the GPS Data Receiver Application

```
<?php
set_time_limit(0);
$address = '127.0.0.1';
$port = 8500;
$max_clients = 100;
$client = array();
$sock = socket_create(AF_INET, SOCK_STREAM, 0);
socket_bind($sock, $address, $port) or die('Could not bind to the IP Address');
socket_listen($sock);
while (true) {
  $read[0] = $sock;
  for ($i = 0; $i < $max_clients; $i++)
  {
    if ($client[$i]['sock']  != null)
      $read[$i + 1] = $client[$i]['sock'] ;
  }
  $ready = socket_select($read,$w = null,$e = null,$t = null);
    if (in_array($sock, $read)) {
    for ($i = 0; $i < $max_clients; $i++)
    {
      if ($client[$i]['sock'] == null) {
        $client[$i]['sock'] = socket_accept($sock);
        break;
      }
      elseif ($i == $max_clients - 1)
        print ("too many clients");
    }
    if (--$ready <= 0)
      continue;
  } // end if in_array

  for ($i = 0; $i < $max_clients; $i++) // for each client
  {
    if (in_array($client[$i]['sock'] , $read))
    {
      sleep(1);
      $input = socket_read($client[$i]['sock'] , 1024);
      if ($input == null) {
        unset($client[$i]);
      }
      $n = trim($input);
      if ($input == 'exit') {
        socket_close($client[$i]['sock']);
      } elseif ($input) {


        list($input1,$input2,$input3,$input4,$input5,$input6,$input7,$input8,$input9,$input10,$input11,$input12,$input13,$input14,$in
put15,$input16,$input17,$input18) = split('[&,#]', $input);
                $space = array("$","?");
                $input1 = str_replace($space, '', $input1);

                $latdegree=(int)($input5/100);
                $latminutes= $input5-($latdegree*100);
                $latdotdegree=$latminutes/60;
                $latdecimal=$latdegree+$latdotdegree;
                $input5=$latdecimal;

                $longdegree=(int)($input7/100);
                $longminutes= $input7-($longdegree*100);
                $longdotdegree=$longminutes/60;
                $longdecimal=$longdegree+$longdotdegree;
                $input7=$longdecimal;

                if ($input6 == 'S') {
                $input5=($latdecimal*(-1));
                }
```

```
                if ($input8 == 'W') {
                $input7=($longdecimal*(-1));
                }

                if ($input4 == 'A') {
                $input4='1'; //Valid fix
                }
                else if ($input4 == 'V') {
                $input4='0'; //Invalid fix
                }

                $input9=round((($input9)*(1.852)),0);

                $hour=substr($input3,1,-8);
                $min=substr($input3,3,-6);
                $sec=substr($input3,5,-4);
                $input3=$hour.$min.$sec;

                $day=substr($input11,0,-4);
                $month=substr($input11,2,-2);
                $year=(2000 + (substr($input11,4)));
                $input11=$year.$month.$day.$input3;

                $input11=strtotime($input11) + 10800;
                exec('php update_db.php "'. addslashes($input1) .'" "' . addslashes($input4) .'" "' . addslashes($input11) .'" "'.
addslashes($input7) .'" "' . addslashes($input5) .'" "' . addslashes($input9) .'"');
        }
     } else {
        //Close the socket
                        if ($client[$i]['sock'] != null){
                        socket_close($client[$i]['sock']);
                        unset($client[$i]);
}
      }
   }
 } // end while
 // Close the master sockets
 socket_close($sock);
 ?>
```