Design and Implementation of Intelligent Traffic Congestion System

Monitoring and Analysis Approach

تصميم وتطبيق نظام ذكي لاختناقات المرور

مراقبة وتحليل

Prepared By

ShireenY.khazneh

Supervised By

Prof. Dr. MuzhirShaban. Al-Ani

This thesis is submitted in fulfillment of the requirements for the Master's degree of science in Computer Science

Department of Computer Science

Collage of Computer Science and Informatics

Amman Arab University

2012

Delegation

I am Shireen Yousef Mohammad Khazneh delegates Amman Arab University to make copy of my dissertation to libraries institutions, or people when asked.

Name: Shireen Yousef Mohammad Khazneh

Signature:

Date:3/12/2012

Discussion Committee's Decision

This Dissertation was discussed under the title:

"Design and Implementation of Intelligent Traffic Congestion System
Monitoring and Analysis Approach"

It was approved in: 3/12/2012

Dissertation Committee Approval:		**Signature**
Prof. Dr. **Prof. Reyadh Naoum**	Chairman	R. Naoum
Prof. Dr. Alaa. Al-Hammami	Member	Alahamami
Prof. Dr. Muzhir Al-Ani	Member/Supervisor	Muzhir Ani

Abstract

Through this thesis, the primary focus was to reduce the complexities in the information available from the mobile telecommunication companies, from their telecommunication towersabout the road users, in order to be able to notice and predict traffic jams in roads, reduce these jams and provide the ability to proactively interact with certain roads, at certain period of times from the historical data analyzed.

The followed methodology worked on utilizing the existing infrastructure in a proper manner to increase the collection of information that is available at the Telecommunications Companies, and scattered data available in an intelligent manner that allows smooth usage of the roads at all times.

In this thesis we analyzedthe traffic load within Jamal Abdulnasser Street and ALdakhleya Circle for two months and collected the readings at different time periods which explain the load of vehicles utilizing this street.At the end of this thesis, the complexities in the data collected from telecommunication operators were solved through cleansing it within an automated process, which covered different algorithms to control data at different stages. Thisprocess provides the ability to benefit from this valuable information that is available historically for analysis and review.

The conclusion was that the algorithms used in cleansing data turned the huge unmanageable data into intelligent information that produced business intelligence reports which provide the ability to analyze and forecast information collected.Moreover, the cost of implementing this solution is not high due to the dependency on available infrastructure.

الملخص

الهدف الأساسي من هذه الرسالة هو التركيز على تحليل وإعادة استخدام المعلومات المتوفرةلدى شركات الهواتف (المتنقلة) وذلك لمعرفة حركة السيارات وحالة الطرق في أوقات مختلفةوهذا بدوره يؤدي إلى تخفيف مدة الاحتقانات وذلك من خلال توفير معلومات واضحة عن حالة الطرق في وقت و منطقة ما معتمدين على المعلومات التي توفرها أبراج شركات الهواتف المتنقلة.

تم استخدام عدد من الوسائل في تحليل البنية التحتية الحالية لشركات الهواتف (المتنقلة) لاستخراج أكبر كم من المعلومات المفيدةوالتي تؤدي في النهاية إلى نتيجة تفيد في تخفيف حدة احتقان الطرق في الأوقات المختلفة.

أوضحت التحليلات التي تمت خلال شهرين من مدة العمل على هذه الرسالة لشارع جمال عبد الناصر ودوار الداخلية في أوقات مختلفة من النهار مدى كفاءة وفاعلية النموذج المقترح في هذه الرسالة والتي اعتمدت بشكل أساسي على إيجاد خوارزمية ذكية لتحليل البيانات الواردة من شركات الهواتف المتنقلة وتحويلها إلى معلومات مفيدة تصب في مصلحة المواطن وتحقق أهداف الرسالة.

خلاصة الدراسة :تم تطوير خوارزميات تحول البيانات الغير منظمة إلى معلومات ذكية بتكلفة معقولة نظراً لاستغلال البنية التحتية الموجودة لشركات الهواتف (المتنقلة).

Acknowledgments

First and foremost, I have to thank almighty God for giving me the courage and the determination, as well as guidance in conducting this thesis, despite all difficulties.

Also, I would like to acknowledge the inspirational instruction and guidance of Prof. Dr. MuzhirShaban. Al-Ani who has been the ideal thesis supervisor. His sage advice, insightful criticisms, and patient encouragement aided the writing of this thesis in innumerable ways.

Thank you to my wonderful husband, Amin N.Muwafi, for being patient, supportive and helping at every stage of this thesis. To my parents who have always said I could do anything that I wanted,and to my mother-in-low for helping and supporting me.

To all my friends, thank you for your understanding and encouragement in my many, many moments of crisis. Your friendship makes my life a wonderful experience. I cannot list the entire name here, but you are always on my mind.

Last, but not least, I extend thanks and appreciation to everyone who helped directly and indirectly to get this work done.

Table of contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

Suffix	Prefix
AGPS	Assisted Global Positioning System
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traveler Management Systems
AUC	AUthentication center
BCH	Broadcast Channel
BI	Business Intelligence
BRT	Bus Rapid Transport
BS	Base Station
BSC	Base Station Controllers
BSS	Base Station System
BTS	Base Transceiver Station
BTS	Base Transceiver Stations
CCTV	Closed-Circuit Television
CFCD	Cellular Floating Car Data
CS	Circuit Switched
EIR	Equipment Identity Register
ETC	Electronic Toll Collection
FCD	Floating Car Data
FPD	Floating Phone Data
GLR	Gateway Location Register
GPS	Global Positioning System
GPRS	General Packet Radio Service
GSM	Global System for Mobile communication
HLR	Home Location Register
IT	Information Technology
ITS	Intelligent Transportation Systems
ISA	Intelligent Speed Adaptation

KSA	**Kingdom of Saudi Arabia**
LAs	Location Areas
MAC	Machine Access Control
MMS	Multimedia Messaging Service
MSC	Mobile Switching Center
MTs	Mobile Terminals
OMC	Operations Maintenance Center
OSS	Operation And Support System
PCS	Personal Communications Services
PS	Packet Switched
PSTN	Public Switching Telephone Network
RFID	Radio Frequency Identification
SGSN	Serving GPRS Support Node
SMS	Short Message Service
SS	Switching System
TDMA	Time Division Multiple Access
VLR	Visitor Location Register
UAE	United Arab Emirates
UITP	International Association of Public Transport
UMTS	Universal Mobile Telecommunications System

Chapter1: Introduction

1.1. Overview

The usage of transportation within the day to day activities has become demanding and required within our life. However, traffic jam is an annoying fact that affects the usage of vehicles.

Transportation changed to be a highly demanding requirement in today's modern cities. Human beings used to consume a considerable Period of time and effort while moving from one location to another. Consumed time may reach to several months while transferring from one location to the other. As part of human's nature in trying to develop and improve their lifestyle, transportation mechanisms have developed over the time. It started with the usage of animals that can tolerate long distances and weights through the movement from one location to another, to the development of mechanical and electrical vehicles that provide the methodology of movement through the utilization of natural or chemical resources. Different vehicle types and methods have been introduced allowing the movement from one location toanother such as bicycles, motor-cycles, cars, buses, planes, or ships.All those transportation vehicles and mechanisms introduced several advantages and disadvantages to the lifestyle of people dealing with different activities. Having different transportation mechanisms reduced the consumed time when moving from one location to the other considerably. This could be enough to consider the usage of those vehicles in our daily life due to the increased value of time, huge reduction of lost time, and the ability to move between different locations dynamically, quickly, and safely[1].

However, the introduction of transportation mechanisms introduced the huge need for availableresources such as petrol, metals, energy, and different other natural resources. The increased eagerness towards those resources built markets at different countries and continents. It introduced the need for money, exchange of goods between traders and governments, and a high demand of supply and demand. The unmanaged demand of resources and the desire of quick and huge profit turned into the possibility of changing the blessing to a curse. Weather pollution, water pollution, struggle to gain resources, and other scenarios were huge disadvantages of this revolution [2].

1.2. Congestion Effects

Vehicles transportation and the increasing demand on vehicles usage of the roads caused a huge disadvantage other than those mentioned above which is traffic jams. Traffic Jam or traffic congestion is a condition on road networks that occurs as use increases, and is characterized by slower speeds, longer trip times, and increased vehicular queuing. The most common example is the physical use of roads by vehicles. When traffic demand is great enough that the interaction between vehicles slows the speed of the traffic stream, this results in some congestion. As demand approaches the capacity of a road (or of the intersections along the road), extreme traffic congestion sets in. When vehicles are fully stopped for periods of time, this is colloquially known as a traffic jam or traffic snarl-up[3].

Traffic congestion occurs when a volume of traffic or modal split generates demand for space greater than the available road capacity; this point is commonly termed saturation.

There are a number of specific circumstances which cause or aggravate congestion; most of them reduce the capacity of a road at a given point or over a certain length, or increase the number of vehicles required for a given volume of people or goods[4].

Traffic jams have several disadvantages and negative impacts such as[5]:

- Wasting time of motorists and passengers ("opportunity cost"). As a non-productive activity for most people, congestion reduces regional economic health.
- Delays, which may result in late arrival for employment, meetings, and education, resulting in lost business, disciplinary action or other personal losses.
- Inability to forecast travel time accurately, leading to drivers allocating more time to travel "just in case", and less time on productive activities.
- Wasted fuel increasing air pollution and carbon dioxide emissions owing to increased idling, acceleration and braking.
- Wear and tear on vehicles as a result of idling in traffic and frequent acceleration and braking, leading to more frequent repairs and replacements.
- Stressed and frustrated motorists, encouraging road rage and reduced health of motorists.
- Emergencies: blocked traffic may interfere with the passage of emergency vehicles traveling to their destinations where they are urgently needed.
- Spillover effect from congested main arteries to secondary roads and side streets as alternative routes are attempted ('rat running'), which may affect neighborhood amenity and real estate prices.

- Car accidents which cause loss in money and lives of people involved within an accident in addition to time and involvement of other parties.

It is not expected that the number of cars will decrease; the number of cars will rather increase due to increasing urbanization. Traffic jams turned the beauty and advantages of transportation and how it changed our lives into an annoying and very demanding fact. Having the traffic jams led to the need to identify alternative scenarios of transportation, inventing more money and time on thesis and studies to reduce the traffic jams in the roads, acquiring more investments from governments and officials to ensure the transport network is better, safer, and serves more vehicles within it. This required envisioning and forecasting mechanisms to be introduced as well in order to properly plan the development and implementation of roads and transport. Moreover, citizens turned into considering their governments as being unsupportive and lack the proper method of planning and investment in order to ensure the decent life for their population[6].

To manage traffic jams, the invention of traffic lights appeared. Traffic lights are implemented in all countries and cities around the world in order to help reducing the problem of traffic jams.

1.3 Intelligent Transportation System (ITS)

Information Technology (IT) is one of the most modern techniques that are utilized within most, if not all, sectors and business activities required within the day to day activities. The Intelligent Transportation Systems (ITS) Institute is a national University Transportation Center headquartered at Minnesota.The interest of Intelligent Transportation System (ITS) comes from the problems caused by traffic congestion and a synergy of new information technology for simulation, real-time control, and communications networks. Congestion has been increasing worldwide as a result of increased motorization, urbanization population, and changes in population density. Congestion reduces efficiency of transportation infrastructure and increases travel time, air pollution, and fuel consumption[7].

Recent governmental activity in the area of ITS is further motivated by an increasing focus on homeland security. Many of the proposed ITS systems also involve surveillance of the roadways, which is a priority of homeland security [8].

The urban infrastructure is being rapidly developed, providing an opportunity to build new systems that incorporate ITS at early stages[7].

Intelligent transport systems vary in technologies applied, from basic management systems such as car navigation, traffic signal control systems; container management systems, variable message signs, automatic number plate recognition or speed cameras to monitor applications, such as security Closed-Circuit Television (CCTV) systems, and to more advanced applications that integrate live data and feedback from a number of other sources, such as parking guidance and information systems, weather information, bridge deicing systems, and the like. Additionally, predictive techniques are being developed to allow advanced modeling and comparison with historical baseline data[9].

1.4 Congestion Solutions

Solving theCongestion problem in the roads may be implemented through different approaches[10]:

1.4.1Vehicle Re-Identification

Vehicle re-identification methods require sets of detectors mounted along the road. In this technique, a unique serial number for a device in the vehicle is detected at one location and then detected again (re-identified) further down the road. Travel times and speed are calculated by comparing the time at which a specific device is detected by pairs of sensors. This can be done using the MAC (Machine Access Control) addresses from Bluetooth devices,or using the RFID (Radio Frequency Identification) serial numbers from Electronic Toll Collection (ETC) transponders (also called "toll tags")[11].

1.4.2.GPS Based Methods.

An increasing number of vehicles are equipped with in-vehicle GPS (satellite navigation) systems that have two-way communication with a traffic data provider. Position readings from these vehicles are used to compute vehicle speeds [13].

1.4.3.Traffic Loops and Cameras

Production of data through fixed monitoring equipment typically deployed in a discontinuous way and mostly limited to highways or main roads so not common on theunderlying network[14].

1.4.4. Triangulation Method

In developed countries, a high proportion of cars contain one or more mobile phones. The phones periodically transmit their presence information to the mobile phone network, even when no voice connection is established. In the mid-2000s, attempts were made to use mobile phones as anonymous traffic probes. As a car moves, so does the signal of any mobile phones that are inside the vehicle. By measuring and analyzing network data using triangulation, pattern matching or cell-sector statistics (in an anonymous format), the data was converted into traffic flow information. With more congestion, there are more cars, more phones, and thus, more probes. In metropolitan areas, the distance between antennas is shorter and in theory accuracy increases. An advantage of this method is that no infrastructure needs to be built along the road; only the mobile phone network is leveraged. But in practice the triangulation method can be complicated, especially in areas where the same mobile phone towers serve two or more parallel routes (such as a freeway with a frontage road, a freeway and a commuter rail line, two or more parallel streets, or a street that is also a bus line). By the early 2010s, the popularity of the triangulation method was declining[12].

Using this method, almost all vehicles within the roads will be covered within the analysis since all vehicles utilizing the roads have passengers with mobile phones and the mobile towers are available in all roads.

Although the usage of the triangulation method has decreased due to its complexity, our thesis is to reduce those complexities since this method has the lowest cost of implementation due to the readiness of its infrastructure. This method depends on telephone towers assembled in the roads and mobile phones available within the road users. Since most drivers already have mobile phones and telephone towers are already assembled, the infrastructure readiness is almost complete and cost is at the minimum.Through this thesis we will work on solving all complexities that prohibit the adoption of this method in order to allow the seamless implementation of it within our region ensuring that the least cost and overhead are required within those implementations.Later on, we will provide useful information to a driver as predictive information. Since, it allows the road users toestimate the travel time of their trips under consideration of future developments of the traffic conditions [15].

This thesis will present a methodology towards reducing the traffic jams caused by vehicles over the roads through the forecasting the number of vehicles that will be served at a certain road at any given time.

1.5 Importance and Objectives

The implemented method within this thesis uses the advantage of an already available technique which reflects in lower start-up costs, less time in initiation and execution, reduced opportunity of loss of time and cost, and ability to take advantage of a huge amount of historical data that is already collected and not used to serve this approach. The availability of such huge amount of information will provide the system with the instructions and methods to learn from the historical data it has and allow the experiment of the system's intelligence and recommendations prior to its actual adoption and implementation which results in real scenarios to test[9].

Within the thesis, the objective is to resolve the complexity of this approach and its usage in order to be able to notice and predict traffic jams in roads, reduce these jams and provide the ability to proactively interact with certain roads at certain period of times from the historical data analyzed. All actions and behaviors executed through this method are based on the historical data collected and are not real-time information. The roam of error when depending on historical data shall be reduced by excluding the anomaly factors that might persist within the collected data. The analysis provided will be for the number of vehicles consuming roads at a certain period of time with the exclusion of traffic jams that occur due to accidents, fire, or any abuse of the roads.

From the data collected in such systems, a Floating Car Data (FCD) market is only now growing worldwide with a wide range of applications and benefits. This would not only improve traffic management but would also help to satisfy the growing demand of drivers who are willing to pay service providers as long as they have access to relevant real-time information: will there be any congestion on my usual route today? How to avoid it? If not, how long will it last? [16].

1.6. Literature Reviews

World Road Association **(2002)** focuses within their handbook "on the “soft” on engineering approaches and tools available to the network operator to improve network operations. The handbook discuses:

• The shift from the traditional building and maintaining the road network to a service oriented policy towards the road user.

• The road network operator's tasks and measures.

• ITS solutions for network monitoring, maintaining road serviceability and safety, traffic control, travel aid and user information and demand management.

• The institutional and organizational aspects of network operations.

• Performance indicators for network operations.

The handbook, identified details about traffic lights and how they intend to be one of the main solutions for traffic jams. Solutions applied in practice at different countries towards the deduction of traffic jams are. Architecture and algorithm of mobile structure is detailed towards the inclusion of this important component within the research [17].

Wahle **(2002)** combined the contributions of some disciplines with emphasis on the methods of physics. The most important achievements of this workare:

- Giving insights into the concepts of Intelligent Transportation Systems.
- Discussing potential benefits and drawbacks of providing information in traffic systems.
- Analyzing and evaluate the usefulness of multi-agent system for modeling traffic system.
- Developing a general agent-based traffic model, which is capable to include the reaction of drivers to information?
- Studding and modeled the route choice behavior of road users on different time-scales, e.g., from day to day or within a day.
- Demonstrating the usefulness of the proposed approaches in practice, especially in combination with data provided on-line.

It concluded that “the current developments in the field of ITS are summarized. The focus lies on ATIS (Advanced Traveler Information Systems) and ATMS (Advanced Traveler Management Systems), which provide real-time information to the road users in order to distribute traffic among the road network more efficiently [9].

Fabian **(2004)** discusses the dynamics of using vehicle probes for detecting link

participate in traffic flow, also known as Floating Car Data, offers a solution to this problem. This thesis investigates Floating Car Data as a tool for acquiring traffic information. The research identifies that Real-life implementation of such a system is closer than one might think;all components of the system rely on devices widely available today. The system will provide a steady stream of traffic information for the benefit of all road users, even the ones who do not directly participate in the system. The huge benefit to the society is obvious – time spent in traffic jams can be much better spent either on leisure or on economic activities [15].

Bham and others **(2004)** showed that a high fidelity cell based traffic simulation model (CELLSIM) has been developed for simulation of highvolume of traffic at the regional level. Straightforward algorithms and efficient use of computational resources make the model suitable for real time traffic simulation. They also detailed that the model is formulatedusing concepts of cellular automata and car-following models, but is more detailed than cellular automata models and simpler than car-following models. Results of comparison of macroscopic parameters with current car-following models showed that even with the simplicity in modeling. The solution was validated both microscopically and macroscopically in detail using two sets of field data. Graphical, statistical analyses and several error tests were conducted to test the validityof the model. It is also able to simulate the hysteresis effect as observed in real world traffic data. Moreover, the model performed satisfactorily in a most severe stability test [18].

Bar-Gera**(2007)** defined the purpose of this paper is to examine the performance of a new operational system for measuring traffic speeds and travel times which is based on information from a cellular phone service provider. According to these analyses:

The correspondence between the two systems is good. Floating car travel time measurements provided additional assurance for data accuracy.

The cellular phone data appears to be suitable for usage in practical applications, especially for ATIS (Advanced Traveler Information Systems) as well as modeling, planning and management of transportation infrastructure investments.

‘‘Noise’’ is the main potential limitation for the Users of data from the cellular phone system at its current status that accompanies the measurements.

It seems reasonable to anticipate increasing usage of cellular phone-based system for measuring speeds and travel times in the coming future considering the positive results and potential for further improvements and enhancements [19].

Guillaume**(2008)** described their "extended data collection system, in which vehicles are able to collect data about their local environment, namely the presence of road works and traffic slowdowns, by analyzing visual data taken by a looking forward camera and data from the on-board Electronic Control Unit.

Upon detection of such events, a packet is set up containing time, position, vehicle data, results of on-board elaboration, one or more images of the road ahead and an estimation of the local traffic level". The report aims to make a snapshot of the recent developments and discuss the potentials and bottlenecks related to new technologies as well as some short-term perspectives”.The report summarized the advantages of fixed sensors method. Moreover, it summarized the advantages of FCD (cellular phones) method [7].

McLin**(2008)** mentioned that traffic congestion in metropolitan areas is a major problem because of its negative effects which include increased delay, fuel consumption, and pollution. Congestion remedies now center on traffic management solutions which better utilize the existing infrastructure. The research focuses on resolving the control issues in order to achieve regional traffic control. The issues include: obtaining system-wide traffic data, identifying the presence of incident conditions, and determining the best control actions in a timely manner. As a summary of the research a regional control algorithm was developed in the dissertation to evaluate the potential and benefits of integrated regional traffic control. Experimental results show that an automated regional traffic control system has the potential to reduce the negatives consequences of non-recurrent traffic incidents [2].

Friedrich and others **(2008)**mentioned that Floating Phone Data recorded in the cellular phone network are used to derive time-space trajectories of car travelers. Data recorded during the call of a mobile phone user serve as an alternative to floating car data for a cheaper and broader traffic state recognition.

They introduced that FPD (Floating Phone Data), an alternative to FCD which is a considerable data source for traffic state detection. Moreover, main disadvantages of FCD are the low rate of equipped vehicles and the still relatively high communication costs. The research project intends to prove that FPD contributes to the following objectives:

- Provision of traffic state information for motorways and the secondary road network.
- Improving methods for detection and forecast of the traffic state by applying origin-destination matrices derived from FPD.
- Continuously monitoring the service quality (travel times) and the travel behavior (travel demand, route choice) in the network for planning purposes and for evaluating telematics control strategies [20].

Stefano**(2009)** stressed on the fact that "the development of Intelligent Transportation Systems (ITS) requires high quality traffic information in real-time. For several years, under growing pressure for improving traffic management, collecting traffic data methods have been evolving considerably and the access to real-time traffic information is becoming routine worldwide".
The main result of the project has been the design and implementation of a distributed architecture composed of a set of moving platforms, endowed with sensors and processing units, connected through a wireless channel to a Main Center. The distributed architecture paradigm permits to balance the on-board computational load for management of reactive behaviors, the usage of the communication channel and the load for tasks performed in the Main Center [16].

Valerio**(2009)** addressed the problem of using cellular network signaling for inferring real-time road traffic information. They surveyed and categorized the approaches that have been proposed in the literature for a cellular-based road monitoring system and identified advantages and limitations. he outlined a unified framework that encompasses UMTS(Universal Mobile Telecommunications System) and GPRS(General Packet Radio Service) data collection in addition to GSM, and prospectively combined passive and active monitoring techniques.

In the report he explored the possibility of using cellular networks for inferring road traffic condition and proposed an extended framework for the collection and the analysis of signaling traces both from the CS(Circuit Switched) and the PS(Packet Switched)domains of a mobile operator. Finally he explored the collected data and preliminary results showed the potentiality of such a system for mapping cellular network signaling patterns into road anomalies [21].

TomTom**(2009)**outlined the basic technological background, coverage and system design of TomTom's traffic information and navigation service, which is already available in a number of European countries such as the Netherlands, Germany, the UK, France, Belgium, Switzerland and Portugal under the name -HD Traffic.

The paper outlined TomTom's revolutionary concept of time-dynamic navigation using dynamic, historic and real-time traffic and travel-time information gathered from GPS and GSM probes. The accuracy and precision of the travel time and delay information calculating the Estimated Time of Arrival of a planned trip is highly accurate. Furthermore, the routing behavior and detour optimization greatly improved using these technologies. This is beneficial both to TomTom's navigation and routing excellence and to other related applications for traffic planning and management [22].

Erlangung**(2011)** in his research deals with cooperative traffic information systems, which support the driverof a car in selecting a route, based on traffic information collected by other cars. System participants contribute measurements of the traffic situation in their vicinity and use the measurements made by other drivers to find the fastest route to their destination with regard to the current conditions to avoid traffic jams, highly congested roads, place of accidents and other unexpected deterioration.

The research covered continuously updated data mechanism based on a peer-to-peer paradigm and cellular networks, to efficiently distribute traffic informationThe research concluded that "the proposed algorithms may also be applied in other applications where structured data have to be stored and processed in an efficient, robust fashion [1].

Chen **(2011)** presented an iterative travel time forecasting scheme, named the Advanced MultilanePrediction based Real-time Fastest Path (AMPRFP) algorithm.

It summarized that the study extends the empirical prediction model by investigating the influence of real-time traffic flow fluctuations that were caused by congestions and incidents that happened in related arcs to improve the quality of the forecasted travel time. Compares the performance of the traditional coordinates based empirical learning algorithm with the distance based learning algorithm in localizing mobile terminals in outdoor cellular networks. Finally, it extends the work of Chandrasekaran's research by utilizing the 1-D curve matching technique and the fast normalized cross-correlation algorithm [23].

1.7. Statement of the Problem

Within the design of roads in each city, the consideration of vehicles inflation might not be expected accurately. This increases the chance of having more traffic jams when unexpected numbers of vehicles are served at the same time within the same road. Proper analysis on the usage of roads and the number of vehicles being served at a certain time is not available at the moment in most countries. In order to resolve such problems, many countries implement and are implementing different analytics through manual approaches.

The thesis is based upon implementing an automated approach for such analysis. The collection of information available from mobile operators through the data gathering of information from their telecommunication towers shall be followed. The suggestion is to use the data from mobile operators, while the solution will be resolving the complexity in the gathered information due to unwanted information available within it and providing an automated analytics approach on this data in order to reduce traffic jams and provide proactive actions on them. Complexities that may arise are such as:

- A group of mobile phones using the same operating tower within the same vehicle.
- Inactive users that are utilizing the same tower but are not using the road itself such as population within a building.
- Walkers on the road that are not affecting the traffic through vehicles utilizing the road.
- Vehicles within traffic jam which are already idle for a long time and are similar in a certain extent to the walkers within the road.

Chapter2: Cellular Mobile Network Structure

2.1 Mobile Network Architecture

Mobile network is built from a set of components that, together, make the correct function and usage of it as explained in figure 2.1 the components are:

2.1.1 The Switching System

The switching system (SS) is responsible for performing call processing and subscriber-related functions. The switching system includes the following functional units.

- Home Location Register (HLR) —The HLR is a database used for storage and management of subscriptions. The HLR is considered the most important database, as it stores permanent data about subscribers, including a subscriber's service profile, location information, and activity status. When an individual buys a subscription from one of the PCS operators, he or she is registered in the HLR of that operator[24].

- Mobile Services Switching Center (MSC) —The MSC performs the telephony switching functions of the system. It controls calls to and from other telephone and data systems. It also performs such functions as toll ticketing, network interfacing, common channel signaling, and others[25].

- Visitor Location Register (VLR) —The VLR is a database that contains temporary information about subscribers that is needed by the MSC in order to service visiting subscribers. The VLR is always integrated with the MSC. When a mobile station roams into a new MSC area, the VLR connected to that MSC

- will request data about the mobile station from the HLR. Later, if the mobile station makes a call, the VLR will have the information needed for call setup without having to interrogate the HLR each time[24].

- Authentication Center (AUC) —A unit called the AUC provides authentication and encryption parameters that verify the user's identity and ensure the confidentiality of each call. The AUC protects network operators from different types of fraud found in today's cellular world.

- Equipment Identity Register (EIR) —The EIR is a database that contains information about the identity of mobile equipment that prevents calls from stolen, unauthorized, or defective mobile stations. The AUC and EIR are implemented as stand-alone nodes or as a combined AUC/EIR node.

2.1.2 The Base Station System (BSS)

All radio-related functions are performed in the BSS, which consists of base station controllers (BSCs) and the base transceiver stations (BTSs)[26].

- BSC — The BSC provides all the control functions and physical links between the MSC and BTS. It is a high-capacity switch that provides functions such as handover, cell configuration data, and control of radio frequency (RF) power levels in base transceiver stations. A number of BSCs are served by an MSC[27].

- BTS — The BTS handles the radio interface to the mobile station. The BTS is the radio equipment (transceivers and antennas) needed to service each cell in the network. A group of BTSs are controlled by a BSC[25].

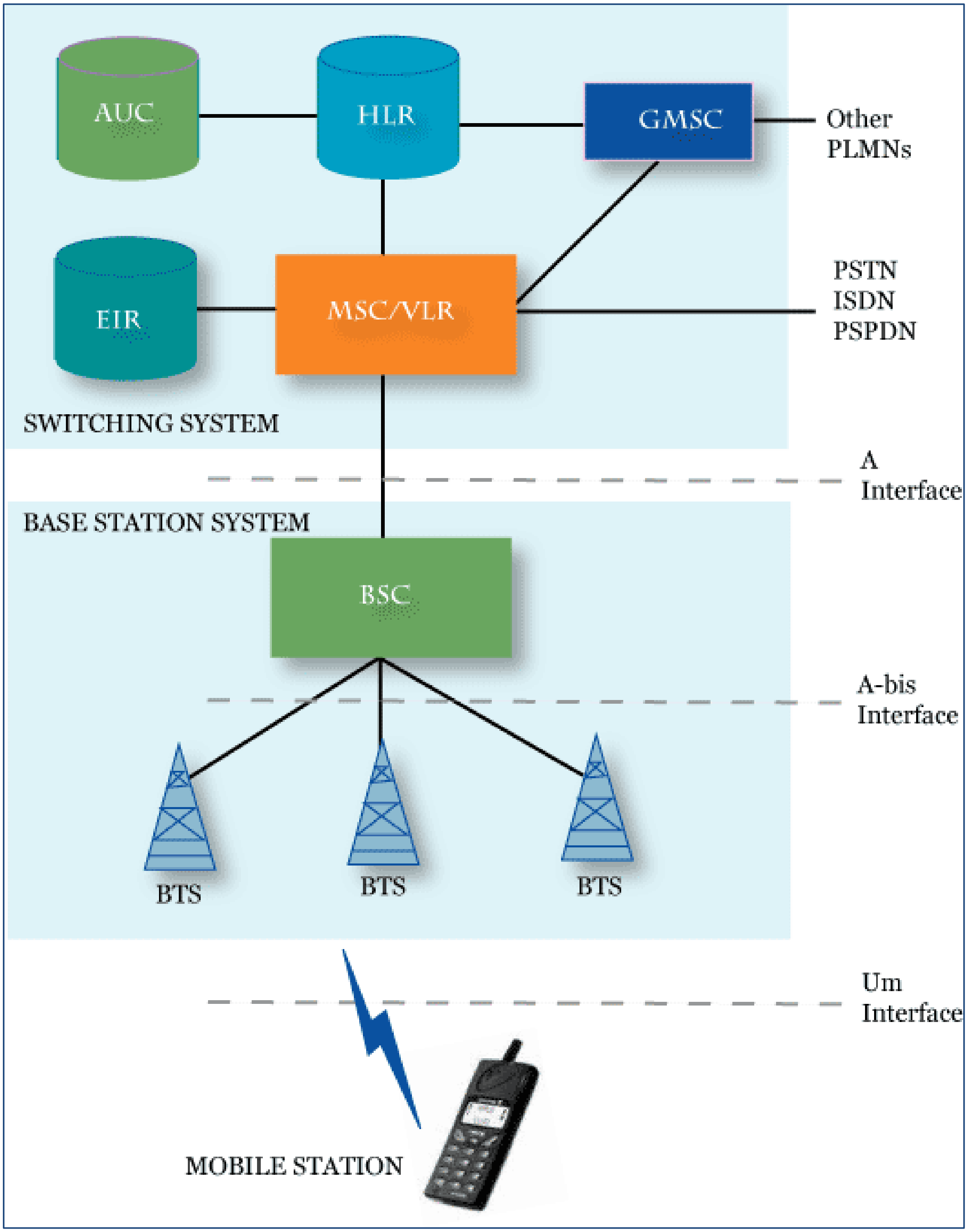

Figure 2.1 Mobile Network Architecture [25]

2.2Implemented Solutions for Traffic Jams

Traffic jams are one of the main problems in modern cities. Traffic jams are a main reason for lots of problems, pollution, and waste of money. Lots of techniques and technologies were used in order to stop or at least reduce this problem.

2.2.1 Traffic lights

One of the solutions for this problem is traffic lights. Traffic lights, which may also be known as stoplights, traffic lamps, traffic signals, signal lights, robots or semaphore, are signaling devices positioned at road intersections, pedestrian crossings and other locations to control competing flows of traffic. Traffic lights were first installed in 1868 in London, and today are installed in most cities around the world. Traffic lights alternate the right of way of road users by displaying lights of a standard color (red, yellow/amber, and green), using a universal color code (and a precise sequence to enable comprehension by those who are color blind)[17].

In the typical sequence of colored lights:

- Illumination of the green light allows traffic to proceed in the direction denoted,
- Illumination of the yellow/amber light denoting, if safe to do so, prepare to stop short of the intersection, and
- Illumination of the red signal prohibits any traffic from proceeding.

Usually, the red light contains some orange in its hue, and the green light contains some blue, for the benefit of people with red-green color blindness, and "green" lights in many areas are in fact blue lenses on a yellow light (which together appear green).

Traffic lights are one of the main (basic) solutions introduced in the world in order to reduce the traffic jams and control the flow of vehicles within all roads.However, due to the increased number of vehicles utilizing the roads, the increase of population, additional demand on traffic, and moving towards urban cities, traffic lights became a traditional solution that can't cope with the increased demands over the roads. Moreover, the concept of deducting number of vehicles is not an option due to the long and various distances between different locations which makes it very hard to reach certain locations through public transportation. Based on that, new techniques were furnished towards the control and deduction of traffic jams in different countries. Those techniques are now in action and have already good feedback and results due to their implementation[28].

2.2.2 Radio Frequency Identification (RFID) technology

This technique executed is in Dubai, United Arab Emirates (UAE) through a system called "Salik", meaning open or clear, is Dubai's new electronic toll collection system which was launched in July 2007. It emphasizes the system's congestion management objectives as well as the choice of technology for the toll system[29].

Salik utilizes the latest technology to achieve free flow operation with no toll booths, no toll collectors, and no impact to traffic flow, allowing vehicles to move freely through the tolling point at highway speeds. There are 4 toll points,each time you pass through a Salik tolling point,a specific amount will be deducted from your prepaid toll account using advanced Radio Frequency Identification (RFID) technology.Your vehicle will be identified to the system through communication with the small, thin Salik sticker tag affixed to your windshield as shown in figure 2.2 .

Figure 2.2 Radio Frequency Identification (RFID) in Dubai [25]

2.2.3 Intelligent Speed Adaptation (ISA)

Another implementation is at London, UK. Technology known as Intelligent Speed Adaptation or ISA which was tested by Transport for London, started in the summer of 2009 on all roads inside the M25. Drivers who have the device fitted will be able to select an option that stops them driving over the speed limit.

The device will automatically slow the vehicle down if the driver unknowingly exceeds the limit and is likely to be offered for sale next year to private motorists, too.

Transport for London believes that the execution of ISA could have a dramatic effect in reducing the number of road casualties within the London area by at least 10%. It is also believed that the improvement in road safety could reduce congestion and vehicles that drive slower produce less pollution[30].

2.2.4 Joining the International Association of Public Transport (UITP)

At Kingdom of Saudi Arabia (KSA), statistics show that 9,400,000 vehicles on the road at present, and half a million new cars registered yearly. Kingdom’s car market worth more than $9 billion, Saudi Arabia will have over four million traffic accidents a year by 2030 as shown in figure 2.3[31].

Figure 2.3Traffic jam in Kingdom of Saudi Arabia (KSA) [31]

Lower road density, a higher proportion of motor vehicles, and the driving culture are reasons behind the large number of road accidents in RIYADH. The General Administration of Traffic has joined the International Association of Public Transport (UITP) in a bid to gain expertise on solving the nation's traffic and public transport problems because it wants to tackle traffic congestion and accidents, and reduce the effect they have on the Kingdom's cities.

According to its website, UITP is the international network for public transport authorities and operators, policy decision-makers, scientific institutes and the public transport supply and service industry. It is a platform for worldwide cooperation, business development and the sharing of know-how between its 3,400 members from 92 countries andit is the global advocate of public transport and sustainable mobility, and the promoter of innovations in the sector[32].

2.2.5A Bus Rapid Transport (BRT) system

At Cape Town, South Africa, building more roads and widening existing ones will never solve Cape Town's traffic crisis, a leading international expert advising the city on a bus rapid public transport plan, has warned.

The international experience had been thatas cities built more roads, traffic also increased. The problem was alleviated, if at all, only temporarily.There is no such thing as a natural level of car use. The number of cars used in the city is a political decision. Traffic problems don't come from more cars, they come from more roads.

Cape Town would have to change the negativity concerning public transport and view the system as "urban improvement". Major cities in the world had committed to reducing the number of cars on their roads by at least 40 percent.

A good transport system has to be low-cost and high-frequency. If the goal is to have a more compact city, car use has to be restricted[33].

Chapter3: Implemented System

3.1 Introduction

It is a fact that traffic jams have a huge effect on our daily life. All countries, governments, and individuals aim towards reducing this problem or cancelling it completely.

However, the efforts towards such resolutions are complicated. Lots of issues are facing the reduction of traffic jams whether in missing information related to the current infrastructure, or proper and scientific studies forecasting the expansion in vehicles usage and roads utilization.

3.1.1 Mobile Innovation

The innovation and growth on the mobile phones front is astonishing. The top-end phones available now have the processing power and storage available in desktop computers just 4-5 years ago. In unfamiliar neighborhoods, they tell us where we are. The address book and contacts list on phones is our social interface. Without the phone, many of us would be quite lost in connecting with other people. The calendar function on the mobile phones can help us track our lives. Phones can also function as radios. For some, the mobile phone also becomes a notepad, send a Short Message Service (SMS) to one-self and make it a reminder service. Owners also have tended to customize phones, with their own ringtones, themes and wallpapers.

The use of a mobile phone is not limited to speaking alone it is being used in making video, recording information and transmitting them to a phone or a computer as was done by a computer. Mobile phone can be connected to a computer to download information from it or vice versa.

Other facilities like online chatting, conferencing, sending text, transferring Multimedia Messaging Service (MMS) information by a mobile phone are compatible with a computer.

Relating to the number of mobile users, it is announced that in Jordan with a population of around 6 million, number of mobile subscribers within the three mobile operators are 7.49 million subscribers which results in having people with more than one mobile subscription. The distribution of those subscribers is divided as follows:

- 2.69 million Subscribers for Zain at 36% of market share.
- 2.5 million Subscribers for Orange at 33.4% of market share.
- 2.3 million Subscribers for Umniah at 30.6% of market share.

3.1.2Proposed Solution

The thesis aims towards providing information related to roads usage, during different timeframes based on the information and data collected from mobile operators based on individuals carrying their mobile phones and utilizing those roads.Through the data collection, information related to each mobile user is defined, information about the change in direction, time consumed, speed is automatically calculated from these values allows the identification of the path used by each single mobile device (which will refer to a vehicle) which will provide the information of the usage of the road at different periods of time. The more information is collected and stored, the more historical knowledge is gained in which the ability to forecast and analyze the usage of roads at different periods of times will be possible, in addition to the easy identification of peak and off-peak times on a certain road. Such information will provide the knowledge and ability to suggest the proper path to follow by each road utilized as a future expansion of this thesis.

The selection of this method has been done based on the fact that most citizens have a mobile which has turned into a fact that a mobile phone may almost be considered mandatory with each person. Moreover, the collection of information related to mobile users and their locations is already collected and stored by mobile operators.

Utilizing mobile will simplify the efforts done and will be one of the most close methods to provide the most accurate results due to having all the infrastructure ready for this implementation.

Within the thesis, the process will cover(1) dynamic collection of data available at mobile operators, (2) dynamic and intelligent algorithms to validate, (3) compile, and (4) cleanse data collected from different mobile operators and (5) identifying what could be considered as wrong data, not used data, corrupted data, or even duplicated data.

3.2 Problem Effects Analysis

As Mentioned previously, traffic jams have a huge effect on our daily life. To identify the size of traffic jam and its effects on traffic accidents, I have visited the Traffic Institute which provided different reports and analytics that specify the accidents' analysis based on different factors. Analysis of those reports and the outcomes for them are detailed in Appendix A.

3.3 Steps for the Solution

The main output of the solution is to get results identifying the usage of different roads at different periods of time in order to provide the intelligence on which roads to follow at different periods of times to ensure the minimal traffic jams and expenses. The solution aims towards providing this information upon request and providing recommendations on roads utilization.

Within the solution, the steps will take action as shown in figure 3.1.

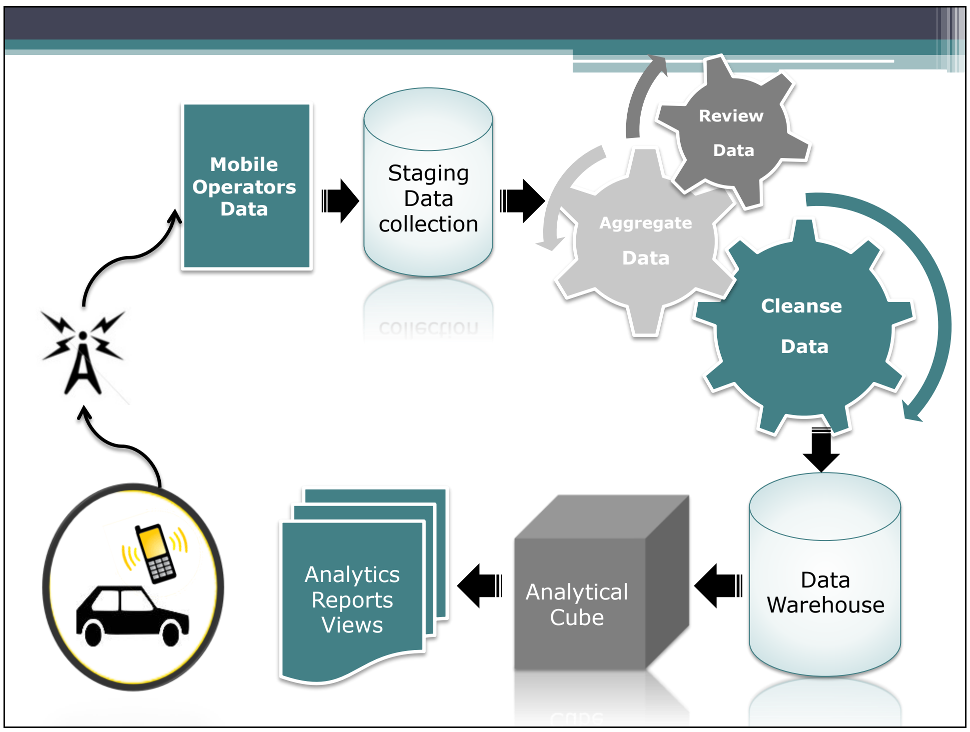

Figure3.1 High-level Implementation

3.3.1Data collection from mobile operators

Each mobile operator will provide data collected for its subscribers consuming the telecommunication services at different locations. Information is collected through the telecommunication towers installed at different locations on the roads; those towers are available to ensure the continuity of the service to those subscribers at different locations.

The data includes the tower from where the data was collected, the date and time this reading was done, with subscriber number optionallyidentified as shown in figure 3.2.

Figure 3.2Example of handover footprints generated by a moving vehicle.

3.3.2Towers Information

Information from each mobile operator related to their telecommunication towers installed shall be collected and defined. Information related to a tower includes the latitude and longitude of the tower. This allows the ability to know the location of each subscriber (vehicle) at a certain period of time based on defining the tower that read the subscriber's location as shown in figure 3.3.

Figure 3.3Row Database + Tower information's for telecom operators

3.3.3 Duplication of data cleansing algorithm

Within the collected information, more than one person might be within the same vehicle. The algorithm shall define those people by automatically deleting records for subscribers that have the same readings from the same towers at the same time and date. This means that all those readings are for more than one person utilizing the same vehicle or for one person that has more than one mobile from the same mobile operator Algorithm 3.1.

```
Define RS as RECORD SET
Define Position as pair (latitude, longitude)
Define n as Integer
Define Time as DateTime
Define MSISDN as String
Load all records within a certain period of time to RS
For each record in RS
    SET Array arr_n to pairs (Position, Time) for MSISDN_n
    For each record in RS
        IF MSISDN_n not equal MSISDN_n+1
            SET Array arr_n+1 to pairs (Position, Time) for MSISDN_n+1
            If all records in arr_n mach arr_n+1
                Delete arr_n+1
            End If
        End If
    End Loop
End Loop
```

Algorithm 3.1 Duplication of data cleansing algorithm

3.3.4 Inactive subscribers' data cleansing algorithm

Information for residents living beside a certain tower will be always collected and provided within the raw data. Those residents are actually not road utilizers, don't affect the usage of the roads, and don't cause traffic jam. The algorithm will automatically remove those readings by defining the repetitive information for certain subscribers that have the same readings from the same tower and different times and dates as shown in Algorithm 3.2.

Define RS as RECORD SET

Define Count as Number

Define distance as Number

Define time as DateTime

Define as Speed as Number

Define Speed Sum as Number

Define Average Speed as Number

Load all records within a certain period of time to RS

For each record in RS

Set Count to Count + 1

Set Distance to sqrt$((Lat_n – Latn_{-1})^2 + (Long_n - Longn_{-1})^2)$

Set Time to $t_n - T_{n-1}$

Set Speed to Distance / Time

Set Speed Sum to Speed Sum + Speed

Set average Speed to Speed Sum / Count

If Speed is less than to **0.5** and average speed is greater than **5.3**

Delete Record

Algorithm 3.2 Inactive subscribers' data cleansing algorithm[١]

3.3.5 Group subscribers' data cleansing algorithm

Partially covered in the "duplication of data cleansing algorithm", subscribers within public transport such as bus, taxi, or service cars, shall be removed from the provided information. Those subscribers may be

[١] 0.5 Error margin for position readings, 5.3 Average walking person speed identified through the fact that all those subscribers have the same reading information from the same tower at the same date and time. Information such as this should be deleted from the data and ignored since the utilization of the road is from one vehicle and the number of vehicles is not based on the number of subscribersas shown in Algorithm 3.3.

```
Define RS as RECORD SET
Define Position as pair (latitude, longitude)
Define n as Integer
Define Time as DateTime
Define MSISDN as String
Load all records within a certain period of time to RS
For each record in RS
        SET Array arr_n to pairs (Position, Time) for MSISDN_n
        For each record in RS
                IF MSISDN_n not equal MSISDN_n+1
                        SET Array arr_n+1 to pairs (Position, Time) for MSISDN_n+1
                        If all records in arr_n mach arr_n+1
                                Delete arr_n+1
                        End If
                End If
        End Loop
End Loop
```

Algorithm 3.3Group subscribers' data cleansing algorithm

3.3.6 Non-vehicle users' data cleansing algorithm

Walking and running people are not vehicle users. Based on that, those don't cause a traffic jam within the road as they don't actually use it. Those readings are identified through having the readings with minor changes at different periods of times. Those users are not utilizing the roads and don't cause a traffic jam as shown in Algorithm 3.4.

Define RS as RECORD SET

Define Count as Number

Define distance as Number

Define time as DateTime

Define as Speed as Number

Define Speed Sum as Number

Define Average Speed as Number

Load all records within a certain period of time to RS

For each record in RS

    Set Count to Count + 1

    Set Distance to $\text{sqrt}((Lat_n – Latn_{-1})^2 + (Long_n - Longn_{-1})^2)$

    Set Time to $t_n - T_{n-1}$

    Set Speed to Distance / Time

    Set Speed Sum to Speed Sum + Speed

    Set average Speed to Speed Sum / Count

    If Speed is less than 5.3and speed is less than ¼ average speed

    Delete Record

End

Algorithm 3.4.Non-vehicle users' data cleansing algorithm[۲]

[۲] The equations are based on Jordan traffic institute

3.3.7 Data transformation from raw data store to staging store

Each mobile operator's data shall be stored in a separate data store. The previously explained algorithms shall be executed individually on each data store. Once those operations are executed, the transformation of data from the mobile operators' store shall be executed towards a staging store (database) which includes a collection of all data cleansed from the preliminary databases as shownin Figure 3.4.

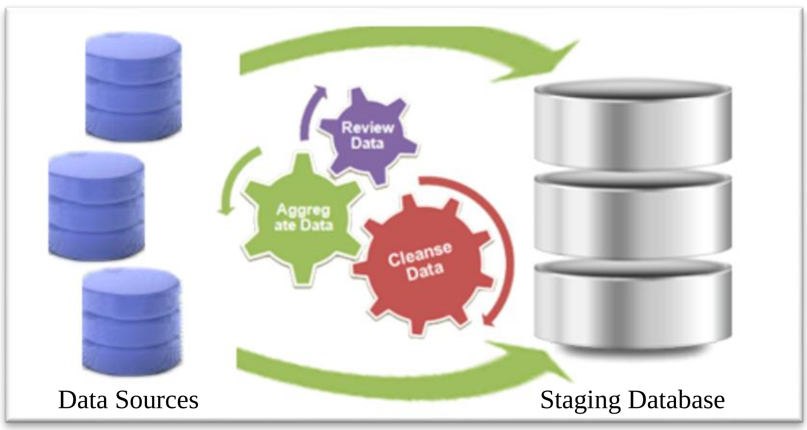

Figure3.4Data transformation from raw data store to staging store

3.3.8 Data cleansing on the staging database

Algorithms previously executed on the raw data shall be executed on the staging database in order to ensure no duplication of data exists in information from different operators.

3.3.9 Data transformation from staging database to data warehouse

After cleansing all data and consolidating them into one location (staging database), transformation activities areapplied on them to move them from the staging database to the data warehouse. A data warehouse is a normal database with a de-normalized structure which allows building the structure of an analytical cube that provides the ability to analyze data and produce dynamic and intelligent reports. The process will transform the data from the staging database to the data warehouse taking into consideration that current data on the data warehouse should not be affected in order to ensure historical information is perceived and available for future reporting and analytics as shown in Figure 3.5.

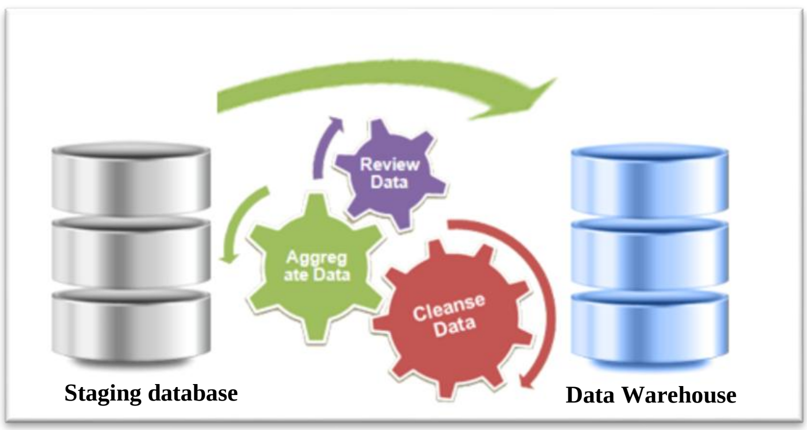

Figure 3.5 Data transformation from staging database to data warehouse

3.3.10 Populate data cubes

As explained earlier, data cubes provide the aggregated collection of data and information in one location. Cubes are built in a manner that allows quick and easy analytics of data and information through the analysis of data in top-down manner. Moreover, within a cube, dimensions, which provide the analytical views are defined which provide the option of identifying on which item analytics are calculated and based as shown in Figure 3.5.

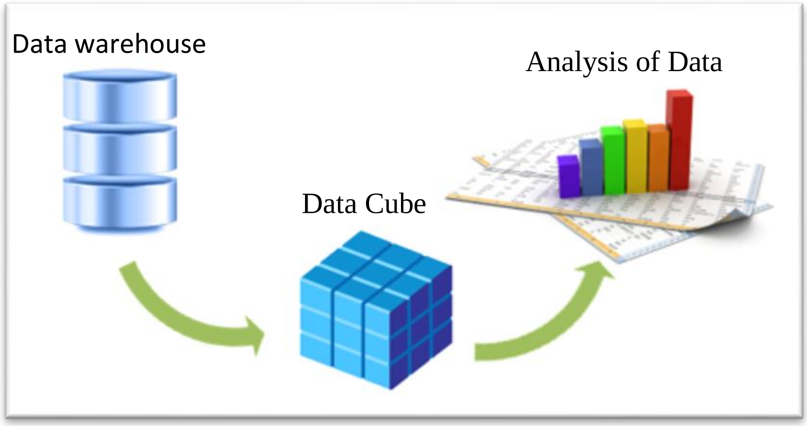

Figure3.6Populate data cubes

3.4. Amman Traffic Map

Since the thesis is based on traffic jams and their effect on the roads, we had to study a certain area during different period of times and seasons to provide the required readings and analytics for this thesis.

We focused on analyzing this information within the region of Amman. Through its roads, different readings and analyes are collected, at such; those readings provide the intelligence towards providing the correct suggestions and recommendations for those areas.

In the course ofthesis, we're covering a specific area in Amman as the utilization of its roads and number of vehicles accessing those roads at different periods of times.

In this thesiswe considered different object types available within a map that are known in traffic jams through roads, highways, intersections, circles, and other objects such as Sports City Circle, Queen Rania Street, 4th Circle, 7th Circle, 8th Circle, City Center, and Al Madina Al Monawwara Street.

We have analyzed the map of Greater Amman Municipality, and through the analyses we drilled down to the specific road and related objects within that road for our thesis, data collection, and analyticsas shown in Figure 3.6.

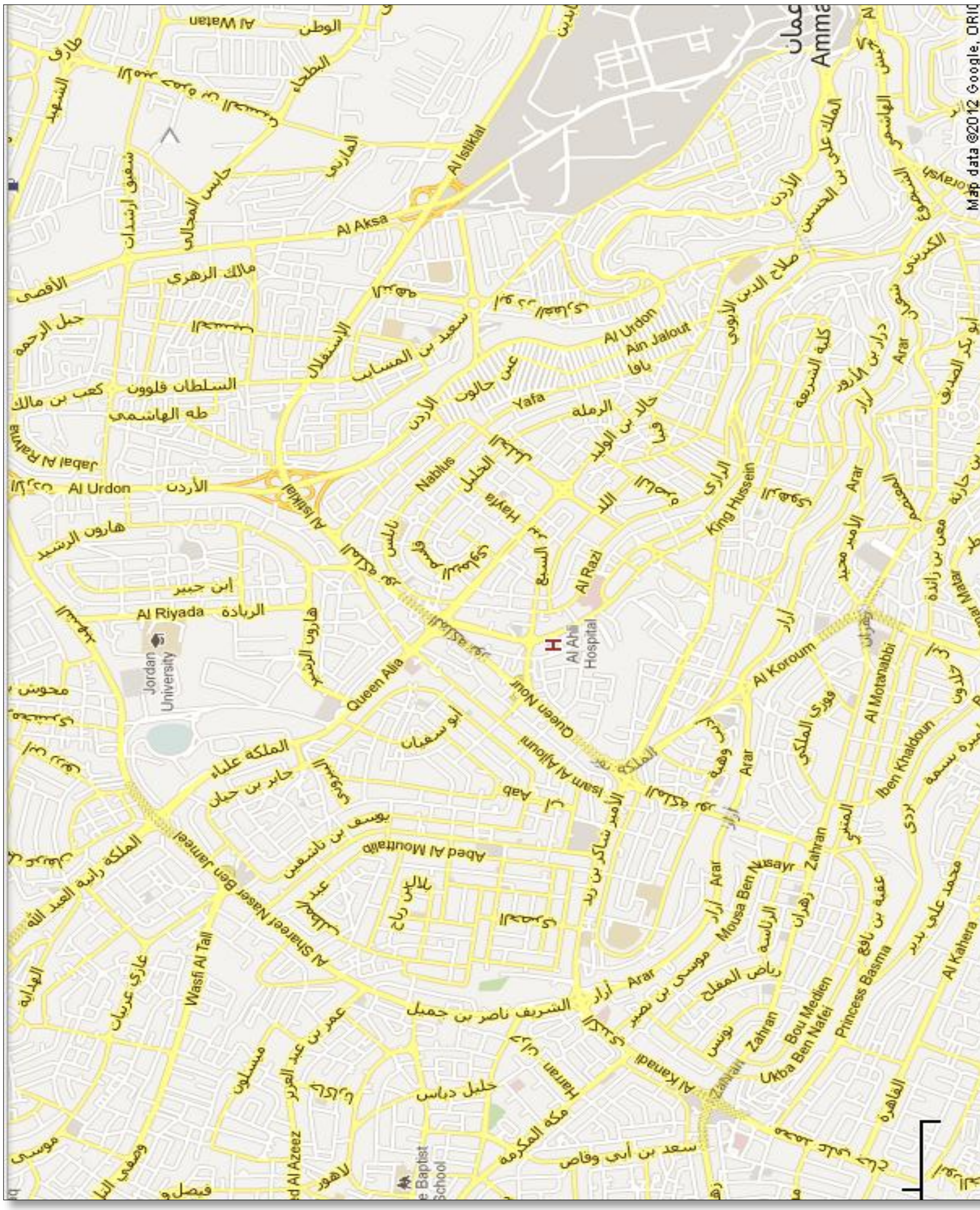

Figure 3.7 Amman Traffic Map [Google Earth]

The main focus within this thesis was in studying and analyzing information related to Jamal Abdulnasser Street andALdakhleya Circle. This circle has lots of traffic jams at different periods of time, in addition to different roads related to this circle that cover lots of routes used by vehicles.

3.5. Database Design

The database design for Amman Map, considered the following tables as shown in figure 3.11

1. Object types (intersection, road, bridge, circle …).
2. Cells (towers) locations for mobile operators within the areas.
3. Area covered by each cell.
4. And location of objects within the map (i.e. location of a road, bridge, circle …).

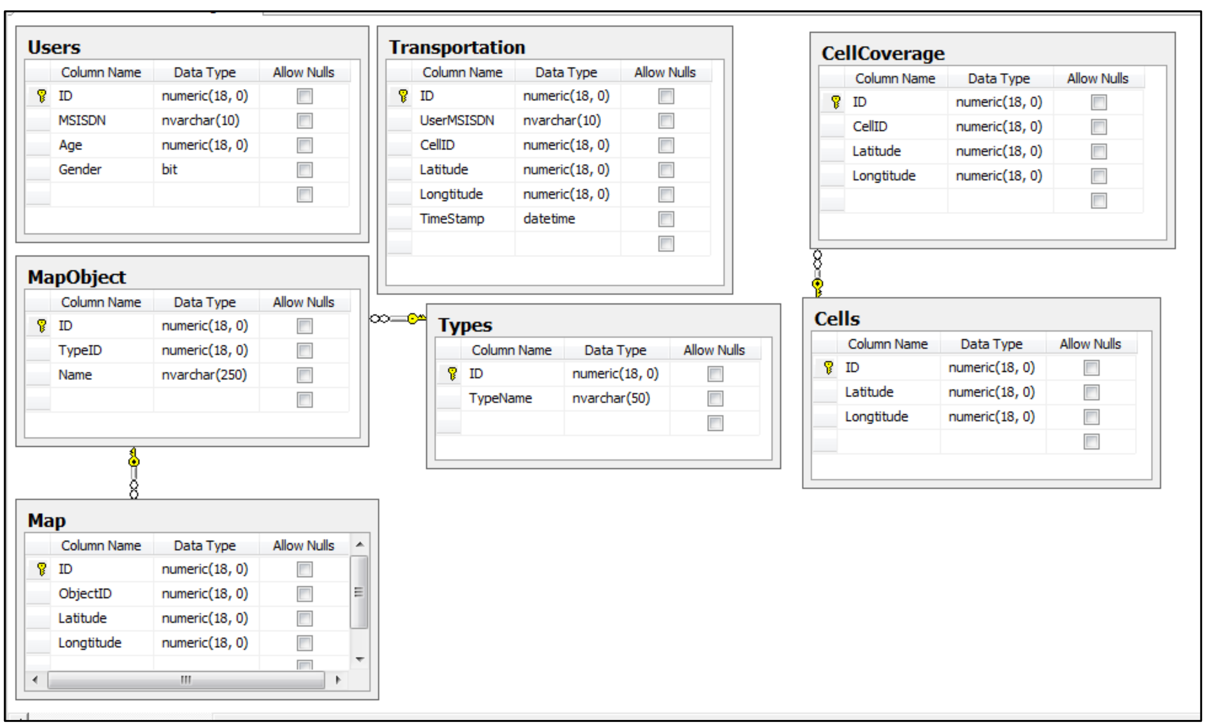

Figure 3.8Database Design

3.6. Methods and sources for obtaining information

Getting the information of roads utilization during a long period of time for historical analyses and details was not an easy task. Moreover, the combination of the need towards having information related to vehicles passing certain roads, number of mobile users, and normal actions of citizens within the city increased the number of sources that are needed to be communicated for this information.

During the thesis, the following organizations have been contacted for information:

3.6.1Department of Traffic

It gives Information about traffic, number of vehicles, roads usage and utilization, and other details. We have completed a number of interviews, meetings, and discussions with involved parties within the department in order to identify this information. Moreover, details of the roads and intersections in Amman were reviewed and summarized in order to ensure the analytics are properly managed and reviewed.

3.6.2. Telecommunication Companies

We have met with the involved people in the three telecommunication companies in Jordan (Zain, Orange, and Umniah), We have identified the needs towards the thesis, required information to be retrieved from those companies, and details needed within the data provided. Based on our meetings, certain information was identified as unavailable within those companies due to the huge information history which intersects with available resources. Moreover, certain details were not provided due to telecommunication towers capabilities. However, other information was identified as available based on our discussions and were used within the thesis and analytics done.

3.6.2Questionnaire

A questionnaire (Appendix B) was spread to a randomly selected group of individuals collecting information about demographical information related to them such as age, number of family members, and number of family members with mobile phones. As a result of retrieving the data collected from the questionnaire, information about the number of possible mobile users with their age periods was available to us as a supporting tool within the thesis and analytics.

The questionnaire was used within the thesis due to the fact that it wasn't possible to retrieve information from the mobile operators.

3.6.2.1Questionnaire Analysis

The questionnaire was distributed to 270 individuals, returned from 260 of them. Errors in filling the questionnaires were in 23 items. Remaining valid ones were 237. Based on the selected questionnaires, number of mobile users is 1,072 users as shown in Figure B.1 in Appendix B

Figure 3.8explains the theoretical part including the Questionnaire's variables and the dimensions it was measured through.

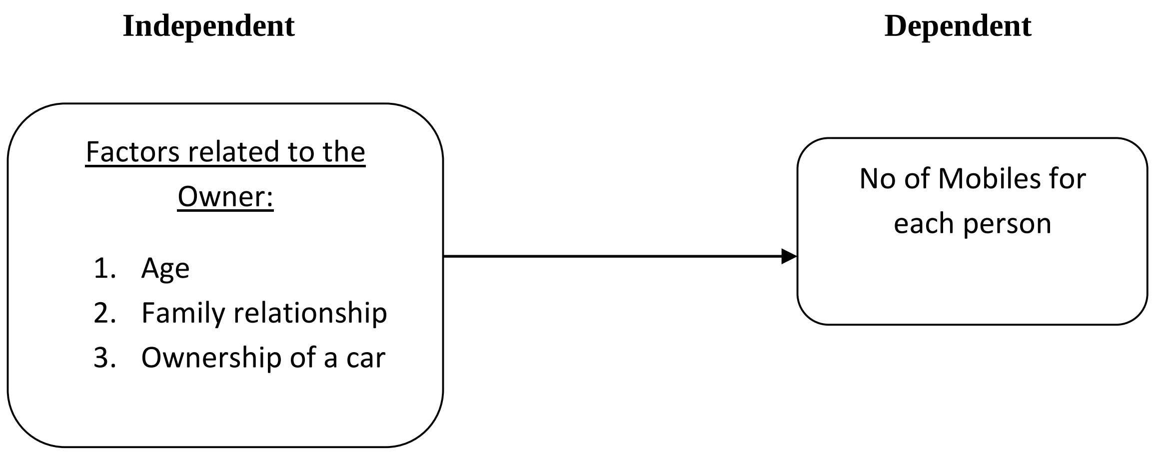

Figure 3.9Theoretical parts for Questionnaire's variables Model

One dependent factor, number of mobiles for each person is affected by the independent factors related to the mobile owners, which use mobile phones, interact with them, and utilize services available within the mobile phones. Those independent factors are:

1- **Age**: The age of the mobile owner utilizing the mobile and its services.
2- **Family relationship**: The level of family relationship between the person that filled the questionnaire and the mobile owner. For example, older parents use one mobile phone for each.
3- **Ownership of a car**: If the person utilizing the mobile is a car owner.

After studying the relationships between the independent variables and the dependent variable,based on table 3.1, there is a significant relationship between all the independent variables together and the dependent variable since the significance value is less than (**0.005**).

Table 3.1 significant relationship between all the independent variables together and the dependent variable

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	31.717	3	10.572	25.104	.000[a]
	Residual	450.626	1070	.421		
	Total	482.343	1073			

a. Predictors: (Constant), Car_Owner, Rel_Degree, Age

b. Dependent Variable: No_Tel

After studying the relation between the independent factors and the dependent factor,based on table 3.2there is a significant relationship only between the car ownership and number of mobiles.

Table 3.2.significant relationship between every independent vari able and the dependent variable

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.027	.108		18.710	.000
	Age	-.024-	.020	-.036-	-1.161-	.246
	Rel_Degree	.010	.013	.022	.737	.461
	Car_Owner	-.365-	.043	-.264-	-8.492-	.000

a. Dependent Variable: No_Tel

```
DESCRIPTIVES VARIABLES=Rel_Degree Age No_Tel Car_Owner
```

The questionnaire aims towards figuring out:

1- In which **age period** does the most ownership of mobiles exist?

2- How many **mobile phones** for each person?

3- How many **mobile phones** in each car?

Those results are utilized within the algorithms applied on the data for cleansing purposes.

From the table 3.3, the**mean** for car ownership (Car_Owner) is **1.38** which means that almost every person owns a car.

The mean for number of telephone users (No_Tel) is **1.49** which means that every person at least has two mobile phones within each single car.

The range for telephone users (No_Tel) is **3**, from the questionnaire.The maximum number of mobiles owned by one person was **4**.

Based on all the above, each car, and each person has two mobile phones which results that each car has two mobile phones within it.

Table 3.3 Descriptive statistics for all the independent variable

Descriptive Statistics

	N	Range	Sum	Mean	
	Statistic	Statistic	Statistic	Statistic	Std. Error
Car_Owner	1074	1	1478	1.38	.015
No_Tel	1074	3	1598	1.49	.020
Age	1074	4	3022	2.81	.032
Rel_Degree	1074	5	3308	3.08	.046
Valid N (listwise)	1074				

Now fromtable 3.4, we noticed that the age period with most mobile phones is (2), which is (18 – 25 years) from the questionnaire.

Table 3.4 The relation between age period and no of telephones the person owned

Age * No_Tel Crosstabulation

Count

		No_Tel				Total
		1	2	3	4	
Age	1	66	12	0	0	78
	2	246	151	19	7	423
	3	114	95	18	5	232
	4	187	92	20	4	303
	5	28	8	2	0	38
Total		641	358	59	16	1074

3.7. How did we build the Conceptualization?

Based on the information wehave analyzed and reviewed in the questionnaire, wehave identified the number of vehicles, mobile users, and utilization within the roads. This provides information on how to identify more details about vehicles, roads utilization, and possibility of traffic jams during the day. Through this information, the importance of those factors within the real-life was mapped identifying that this data is needed in order to produce the required results.

3.8. Proposed Solution Details

Upon all the above rules and facts, and since the collection of data from the telecommunication companies was not available due to the unavailability of information at those companies, security reasons, and confidentiality requirements, it was a must to generate the data through a simulating application.

The application suggests a certain number of vehicles passing through a certain route during a certain period of time (Algorithm 3.5). Moreover, the application uses as an input, the number of mobile users utilizing the same vehicle. Based on that, data generated from the application includes all available real-time options that might occur in the real life. This provides the ability to execute and run the algorithms on this data for cleansing and finalization purposes. Later on, data is grouped and analyzed. Dynamic analytics are generated and output is furnished for those items.

1. Vehicles passing within a road:

a.Inputs:

1. **Path** of the vehicle (latitude, longitude) for checkpoints
2. Reading **time** (start and end of reading)
3. The **Avg**(Number of mobiles within the vehicle)generate from the questioner

b. Outputs:

1. **Number of vehicles** passing the path during the reading time.
2. **Number of mobiles** will range from 1 – 5. However, they might reach 50 in some vehicles (such as busses).
3. **Vehicles speed** will be identified from the different readings through the formula **speed = distance / time.**

Results of the algorithm will **provide** the following information:

- Possibilities of traffic **jam** at a certain period of time.
- **Number** of vehicles and **road users** within the reading time.

Algorithm 3.5 Reading data from the roads (Generating data)

The Figure 3.9is for a cell related to a mobile operator that covers part of a road. The road is connected with the cell in two locations, **x** and **y**.

A car that passes through the road arrives to point **x** at time $\mathbf{t_1}$ and gets out of the road at point y at time $\mathbf{t_2}$.

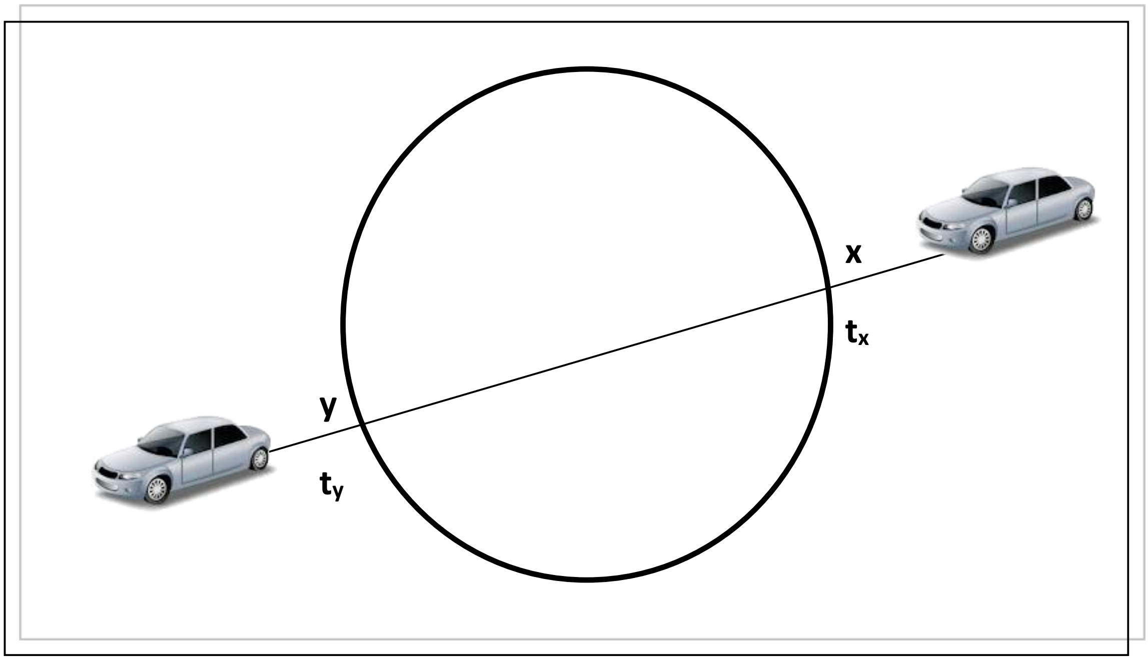

Figure 3.10 Traffic jam analysis – Basic algorithm

The Algorithm 3.6 explains how to calculate if the road has a traffic jam or not. Jam detection algorithm is one of the main algorithms used within this thesis. Through this algorithm, identification of the speed of all vehicles passing two points within a road at different period of times and being able to calculate their speed provides the ability to identify whether there's a jam on the road, it's clear, or it's medium.

```
Define D as Number

Define Speed as Number

Define Road Speed as Number

Define Average Speed as Number

Define KBI as String

Load all records within a certain period of time to RS

For each record in RS

Set Distance of the road (d) to square root((latitudex) - (latitudey)2) + (longtitudex) - (longtitudey)2 )

Set Speed to D (distance of the road) / (ty – tx).

IF (average speed (S) is less than Road Speed) and (Number of Cars greater than or equal to 100) Then

IF Average Speed is between 90% of Road Speed to 100% of Road Speed THEN

Set KBI to Low Traffic Jam END IF

IF Average Speed is between 70% of Road Speed to 90% of Road Speed THEN

Set KBI to Medium Traffic Jam END IF

IF Average Speed is less than 70% of Road Speed THEN

Set KBI to High Traffic Jam END IF

END IF

END LOOP
```

Chapter 4: Results and Analysis

Algorithm 3.6Jam Detection

During the implementation of the results in this thesis, and due to the lack of data from the telecommunication operators, the researcher able to generate data through the simulation of similar information that would exist in real life collected by those operators.

The simulated data generated includes problematic information. The way to resolve then was explained during the explanation of the algorithms.

Data cleansing algorithms were applied on this data through different phases by applying a certain algorithm on each phase. At the end of these phases, the cleansed data was available for analysis and review.

During my research, the researcher able to solve the complexities in the data collected from telecommunication operators as shown in table 4.1 and cleansing them through the automated process of data cleansing which covered different algorithms to control data at different stages. This provided the ability to benefit from this valuable information that is available historically without using it in such scenarios as shown in table 4.2.

Table 4.1 Data before cleansing

ID	UserMSISDN	CellID	Latitude	Longitude	TimeStamp
1	795555447	1	23456	876543	01/06/2012 13:00
2	795555447	1	23459	876545	01/06/2012 13:02
3	795555447	1	23461	876547	01/06/2012 13:03
4	795555447	1	23464	876550	01/06/2012 13:06
5	795555447	1	23469	876558	01/06/2012 13:08
6	795551123	1	23456	876543	01/06/2012 13:00
7	795551123	1	23459	876545	01/06/2012 13:02
8	795551123	1	23461	876547	01/06/2012 13:03
9	795551123	1	23464	876550	01/06/2012 13:06
10	795551123	1	23469	876558	01/06/2012 13:08
11	797523456	1	23456	876543	01/06/2012 13:00
12	797523456	1	23459	876545	01/06/2012 13:02
13	797523456	1	23461	876547	01/06/2012 13:03
14	797523456	1	23464	876550	01/06/2012 13:06
15	797523456	1	23469	876558	01/06/2012 13:08
16	796655430	1	23456	876543	01/06/2012 13:00
17	796655430	1	23459	876545	01/06/2012 13:02
18	796655430	1	23461	876547	01/06/2012 13:03
19	796655430	1	23464	876550	01/06/2012 13:06
20	796655430	1	23469	876558	01/06/2012 13:08
21	796655430	2	23475	876562	01/06/2012 13:11
22	796655430	2	23479	876567	01/06/2012 13:15
23	794445551	1	23445	876645	01/06/2012 13:03
24	794445551	1	23445	876645	01/06/2012 13:04
25	794445551	1	23445	876645	01/06/2012 13:05
26	794445551	1	23445	876645	01/06/2012 13:06
27	790001213	1	23444	876644	01/06/2012 13:00
28	790001213	1	23444	876644	01/06/2012 13:02
29	790001213	1	23445	876645	01/06/2012 13:04
30	790001213	1	23445	876645	01/06/2012 13:06
31	790001213	1	23446	876646	01/06/2012 13:08
32	790001213	1	23446	876646	01/06/2012 13:10

Duplicated data (same users at the same readings).

User within same route but then completes alone.

User at home or standing all time.

Walking user.

Table 4.2 Data after cleansing

ID	UserMSISDN	CellID	Latitude	Longitude	TimeStamp
1	795555447	1	23456	876543	01/06/2012 13:00
2	795555447	1	23459	876545	01/06/2012 13:02
3	795555447	1	23461	876547	01/06/2012 13:03
4	795555447	1	23464	876550	01/06/2012 13:06
5	795555447	1	23469	876558	01/06/2012 13:08
16	796655430	1	23456	876543	01/06/2012 13:00
17	796655430	1	23459	876545	01/06/2012 13:02
18	796655430	1	23461	876547	01/06/2012 13:03
19	796655430	1	23464	876550	01/06/2012 13:06
20	796655430	1	23469	876558	01/06/2012 13:08
21	796655430	2	23475	876562	01/06/2012 13:11
22	796655430	2	23479	876567	01/06/2012 13:15

Thefigures 4.1 and 4.2 explain the data simulated through the process which automatically generated, cleansed, and then produced analysis for end users. This information has passed through an evaluation and review mechanisms to ensure data is correct and not duplicated. Through this process, this ensured to have correct reports and analytics for end users.

1	Date	16/7/2012	17/7/2012	18/7/2012
2	8:00 - 9:00	100	101	100
3	9:00 - 10:00	120	130	125
4	10:00 - 11:00	130	130	133
5	11:00 - 12:00	105	107	109
6	12:00 - 13:00	97	100	95
7	13:00 - 14:00	143	140	166
8	14:00 - 15:00	127	121	133
9	15:00 - 16:00	88	97	83
10	16:00 - 17:00	92	90	98
11	17:00 - 18:00	102	100	102
12	18:00 - 19:00	99	99	99
13	19:00 - 20:00	94	82	99
14	20:00 - 21:00	87	88	102
15	21:00 - 22:00	86	100	94
16	22:00 - 23:00	54	50	52
17	23:00 - 00:00	34	36	32
18	00:00 - 1:00	32	32	27
19	1:00 - 2:00	24	26	25

Figure 4.1 Analytic of peak hours

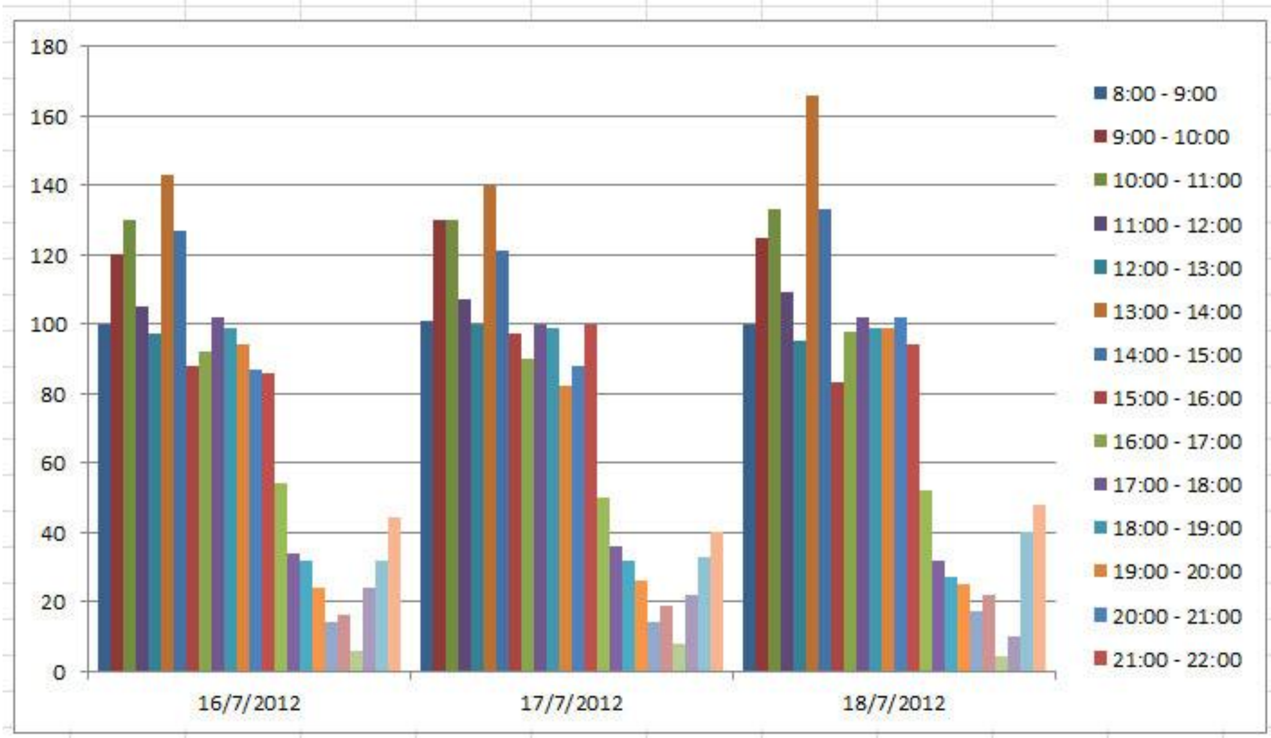

Figure 4.2 Analytics of peak hours

The traffic load within Jamal Abulnasser Street and ALdakhleya Circle for one month have been analyzed and the readings at different time periods which explain the load of vehicles utilizing this street have been collected. The figures 4.3, 4.4, 4.5, and 4.6 explain those readings.

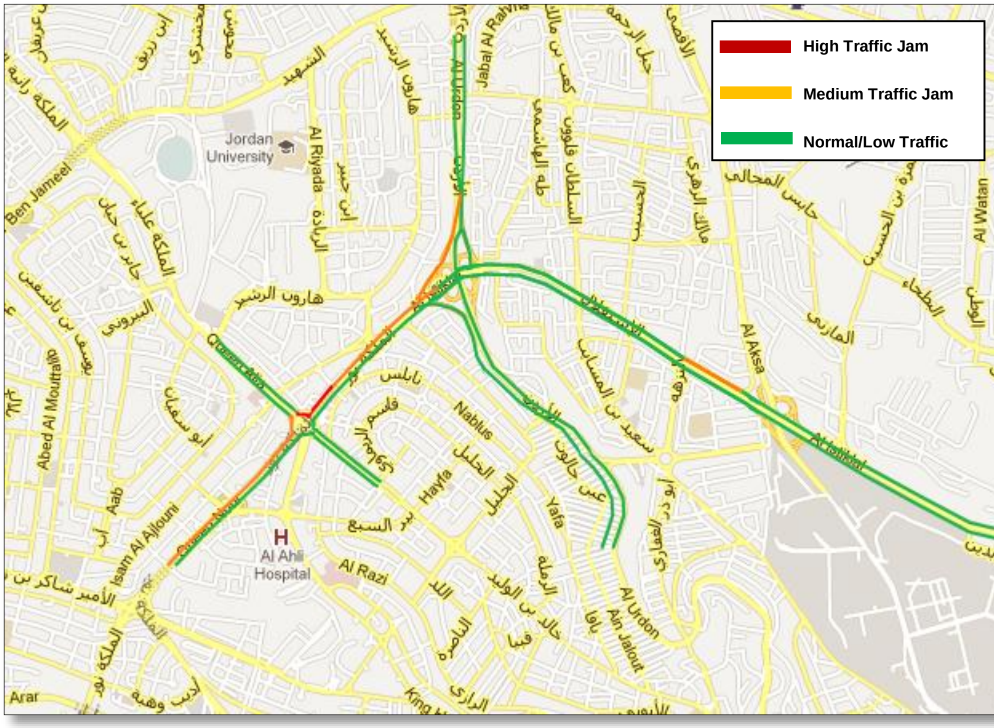

Figure 4.3Presentation of traffic for Jamal Abdulnasser Street and ALdakhleya Circle at (7 AM)

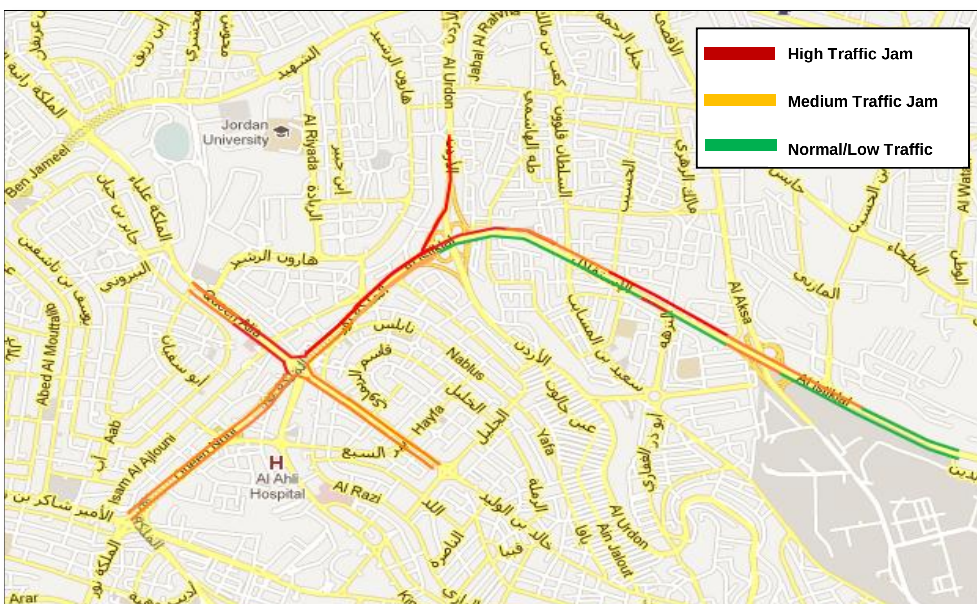

Figure 4.4Presentation of traffic for Jamal Abdulnasser Street and ALdakhleya Circle from (8 AM)

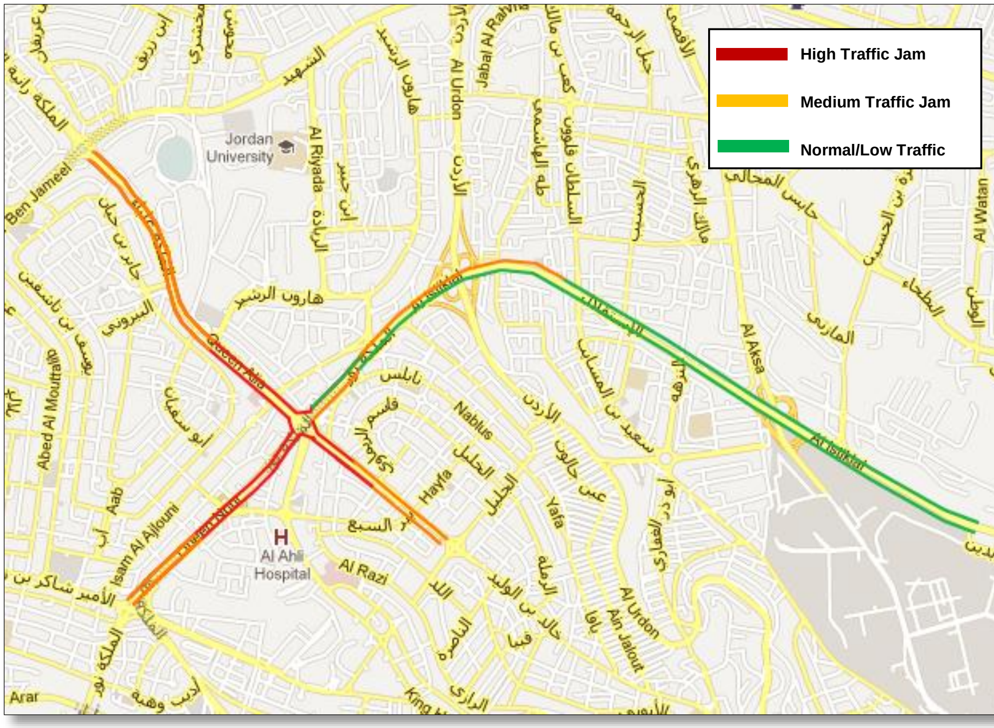

Figure 4.5Presentation of traffic for Jamal Abdulnasser Street and ALdakhleya Circle from (12 PM)

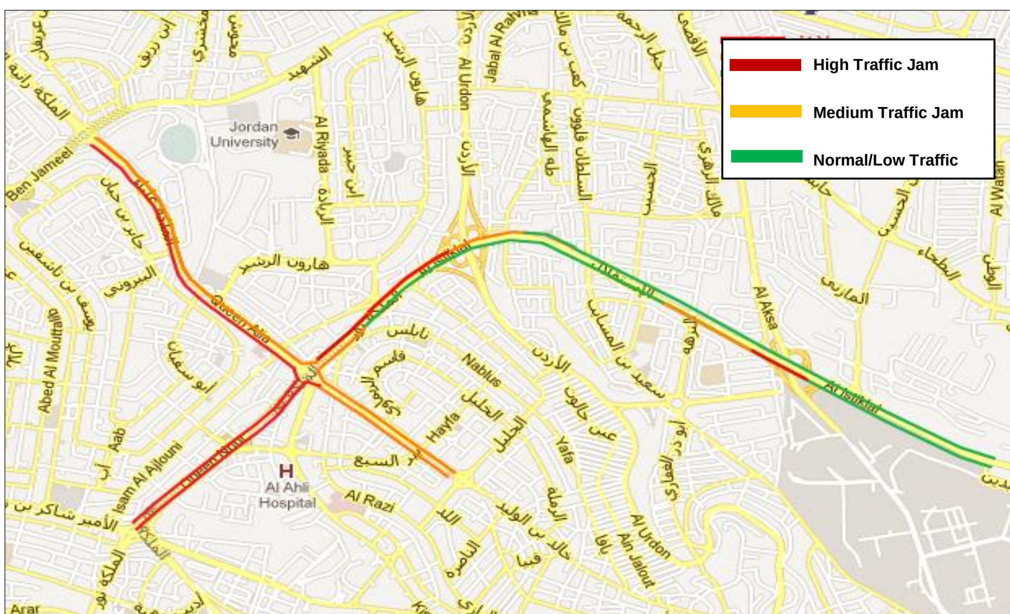

Figure 4.6Presentation of traffic for Jamal Abdulnasser Street and ALdakhleya Circle at (5 PM)

Chapter 5: Conclusion and Future Work

5.1 Conclusion

The thesis got benefit from the already available infrastructure within all roads through the telecommunication towers installed by the different mobile operators, and went through cleansing and analysis of data collected from this information.

Through the thesis, benefiting from the algorithms followed in cleansing the data made it available for review and analysis. Those algorithms showed success in resolving the problems within the data collected from the sources. Through those algorithms, unmanageable huge data that was collected became simple and available for analysis.

Dynamic analyses were prepared providing information regarding the usage of the roads at different periods of time and providing information on possible traffic jams due to the usage of those roads.

5.2 Future Work

The research on this thesis could be carried on, in the future, by any interested party.The room is still open towards involving the analyses built through Business Intelligence in terms of furnishing dynamic reports that provide the ability for decision support, ad-hoc analytics, and root-cause analysis.

Moreover, collecting, cleansing, and consolidating information in real-time format would add value to the thesis. However, real-time information requires high end technologies, infrastructure and services that manage the different scenarios perceived during a very short period of time.

Making this information available to the citizens through any accessible method would add value to the thesis. Citizens would like to get suggestions on which route to follow, where to go each day to their jobs, and how to drive through an intelligent application that provides this information at a correct and timely manner.

Finally, an additional factor in case of implementing the real-time scenario shall be required in adding information related to external factors that would affect the traffic flow such as accidents, motorcades, or any external incident that would occur on a street.

References

[1] Von. V, Rybicki.J and Poznan, "**Cooperative Traffic Information Systems Based on Peer-to- Peer Network:"**,Unpublished Doctoral,Dissertation, Universities Dusseldorf, March 2011.

[2] Mclin.J, "**A regional traffic control system using emerging intelligent transportation systems (ITS) technology"**, ProQuest Dissertations and Theses, December 2008.

[3] Randolph W.Hall, "Non-Recurrent Congestion: **How Big is the Problem? Are traveler Information System The Solution?"** Transportation Research Part C,1(1):89-103,1993.

[4] Sarah B.Medley and Michale J.Demetsky, "**Development of Congestion Performance Measures Using ITS Information"**, Technical report, Virginia Transportation Research Council, Jan 2003, VTRC 03-R1, University of Virginia.

[5] Wikipedia, the free encyclopedia, "Traffic congestion", http://en.wikipedia.org/wiki/Traffic_jam, last modified on 30 July 2012.

[6] Lomax.T and Schrank.D.,“**The 2007 Urban Mobility Report”**, Technical Report, Texas Transportation Institute, May 2007.

[7] Ledus.G, “**Road Traffic Data: Collection Methods and Applications”, Working Papers on Energy**, Joint Research Center Technical Notes, 37, (2008). JRC 47967 – 2008.

[8]Anjuman.T, “**Evaluation of an Online Intelligent Transportation System (ITS) Asset Management System**”, A Thesis Presented to the Graduate School of Clemson University, December 2009.

[9] Wahle.J, “**Information in Intelligent Transportation Systems”**, Unpublished Doctoral Dissertation, Technologie der Universitat, Jan 2002.

[10] Torin.M. "War Rooms" of the Street: Surveillance Practices in Transportation Control Centers”, The Communication Review 10 (4). 2007.

[11] Torp.K, Lahrmann.H, "**Floating Car Data For Traffic Monitoring, Aalborg University"**, Denmark, 2001.

[12] Tarnoff, John.P, Bullock, Darcy M, Young, Stanley E, et al. "**Continuing Evolution of Travel Time Data Information Collection and Processing"**, Transportation Research Board Annual Meeting, 2009.

[13] Fastenrath.U,"**Floating Car Data on a Larger Scale"**, Report Verkehrsdatenmb H, 2007.

[14] Dedene.N, Kenis.E, "**Cellular Floating Car Data trial in Flanders"**,A Case study, Aug 2004- Nov 2005.

[15] Fabin, P,**"TheEnd of Congestion: Developing a Large Scale "Floating Car Data" System"**, Unpublished Bachelor degree,Princeton University,April 2004.

[16] Messelodi.S,Modena.C, Zanin.M, Natale. F, Granelli. F,Betterle.E and Guarise,"**Intelligent extended floating car data collection"**,Expert Systems with Applications, 36(3), 4213–4227, (2009).

[17] World Road Association, "Technical Committee on Network Operations", (C16), Road Network Operations Handboook, V116, 2002.

[18] Bham. G, andBenekohal,R. **"A high fidelity traffic simulation model based on cellular automata and car-following concepts"**, Transportation Research Part C 12 (2004) 1–32, 2004.

[19] Gera,H, **"Evaluation of a cellular phone-based systemfor measurements of traffic speeds and travel times",**A case study from Israel, Transportation Research Part C 15 (2007) 380–391, June2007.

[20] Friedrich, M. Jehlicka, P. Otterstatter, T. Schlaich, "**Monitoring Travel Behavior and Service Quality in Transport Networks with Floating phone Data"**, Proceedings of 4rd International symposium networks for mobility, Stuttgart,2008.

[21] Valerio.D, "**Road Traffic information fromCellular Network Signaling, ftw"**, Technical Report Forschungszentrum Telecommunication Wien,Vienna, AUSTRIA, March 2009.

[22] Atkinson, I., Adam, B. Dixon, J., "**How TomTom's HD Traffic™ and IQ Routes™ data provides the very best routing"**, White paper for Traffic Monitoring System, EP 1 348 208 B1, 2009.

[23]Chen. K**,"Mitigating Congestion by Integrating Time Forecasting and Realtime Information Aggregationin Cellular Networks"**, UnpublishedDoctoral Dissertation**,**Florida International University, Miami, Florida**,** Nov 2011.

[24]Wu, Jie and Dai, F. "**Efficient Broadcasting with Guaranteed Coverage in Mobile Ad-hoc Networks**", IEEE Transactions on Mobile Computing, Vol.4, No.3, 2005, pp. 259-270.

[25]Al-Kasasbeh Basil M., Muzhir S. Al-Ani, Rafa E.Al-Qutaish, Khalid T. Al-Sarayreh,"**Mea Measuring the Optimal Transmission Power of GSM Cellular Network**", A Case Study', Communications of the IBIMA, Vol. 4, pp. 216-221, 2008.

[26] Samfat D., Devernay V., Bonnet C., 1995, "**A GSM simulation platform for intrusion detection**", In:Proceedings of the IEEE international conference on communications (ICC'95), vol. 2. pp. 766-770

[27] I.V. Kashtanov and O.D. Arhelyuk, "**Architectural Solutions Of Testing Network For Mobile Operators"**, International Conference on Antenna Theory and Techniques, October, 2009, Lviv, Ukraine pp. 225-227.

[28]Carlos Gershenson.C, "**Self-organizing Traffic Lights"**, Centrum Leo Apostel,Vrije Universiteit Brussel, Krijgskundestraat 33 B-1160 Brussels, Belgium,Complex Systems, 16 (2005) 29–53.

[29] "**Dubai's Road and transport authority"**, http://www.salik.gov.ae, last modified: 5/07/2012

[30] Prison Planet Educational Forum and Library, "**London's New Technology Can Solve London's Traffic Problems chip n grid 2"**, July 02, 2009, http://forum.prisonplanet.com.

[31] "**Saudi gazette"**, http://www.saudigazette.com.sa.

[32] International Association of Public Transport, http://www.uitp.org.

[33] JOL News, http://www.iol.co.za, "**More roads won't solve traffic problem**", August 18 2008.

[34] Markos Papageorgion etal."**Review of Road Traffic Control Strategies"**, Proceedings of the IEEE, 91(12):2043-2067, Dec 2003.

1- Traffic Accidents Types by time

Traffic accidents, issues, and problems mainly occur during day time, they to the peak between 12:00 and 16:00 in which most people working leave work, non-working people go out for shopping or their daily activities.

On the opposite, number of accidents incrementally decreases during late night hours (1:00 – 5:00) which is normally a time people don't go out and utilize the roads as shown in table A.1 and figure A.1.

Table A.1 Traffic Accidents Type by Time

تدهور Turn Over	مشاة Pedestrian	صدم Collision	Total	Time
22	39	1376	1437	00:00 -00:59
34	51	1843	1928	01:00 -01:59
22	41	1469	1532	02:00 -02:59
30	13	735	778	03:00 -03:59
22	28	604	654	04:00 -04:59
23	23	591	637	05:00 -05:59
33	41	950	1024	06:00 -06:59
61	150	3474	3685	07:00 -07:59
74	132	5680	5886	08:00 -08:59
67	146	5380	5593	09:00 - 09:59
97	178	6856	7131	10:00 -10:59
117	253	9054	9424	11:00 -11:59
120	302	10597	11019	12:00 - 12:59
107	324	9715	10146	13:00 -13:59
108	307	11495	11910	14:00 -14:59
122	279	10901	11302	15:00 - 15:59
116	291	10335	10742	16:00 -16:59
74	307	8939	9320	17:00 -17:59
89	324	9352	9765	18:00 - 18:59
63	276	7097	7436	19:00 -19:59
63	195	5139	5397	20:00 -20:59
48	195	5622	5865	21:00 - 21:59
38	119	3929	4086	22:00 -22:59
45	77	3195	3317	23:00 -23:59
1595	4091	134328	140014	Total £>«a-«S

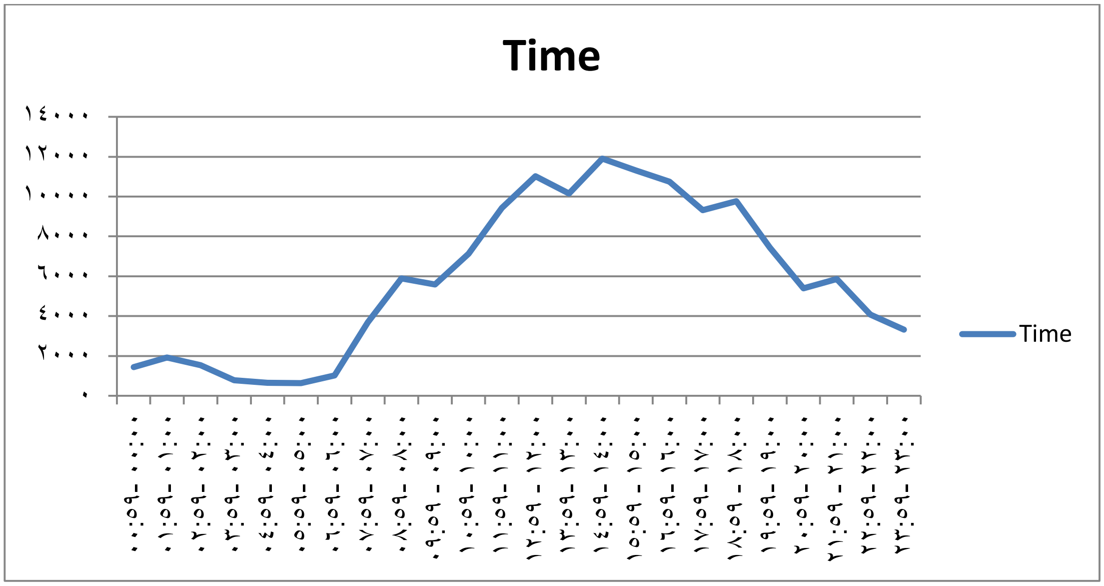

Figure A.1 Traffic Accidents Type by Time

2- Traffic Accidents Type by Weekdays

Traffic accidents mainly occur during week days, during weekends, accidents decrease since the number of vehicles and people within roads get less. In addition, public transportation decreases during weekends which decrease the number of vehicles utilizing the roads. As a result, accidents get less during those days as shown in table A.2 and figure A.2.

Table A.2. Traffic Accidents Type by Weekday

Turn Over	Pedestrian	Collision	Total	Weekday
234	586	18624	19444	Saturday
218	594	20706	21518	Sunday
223	539	20244	21006	Monday
174	594	19715	20483	Tuesday
228	560	20160	20948	Wednesday
286	663	22223	23172	Thursday
232	555	12656	13443	Friday
1595	4091	134328	140014	Total

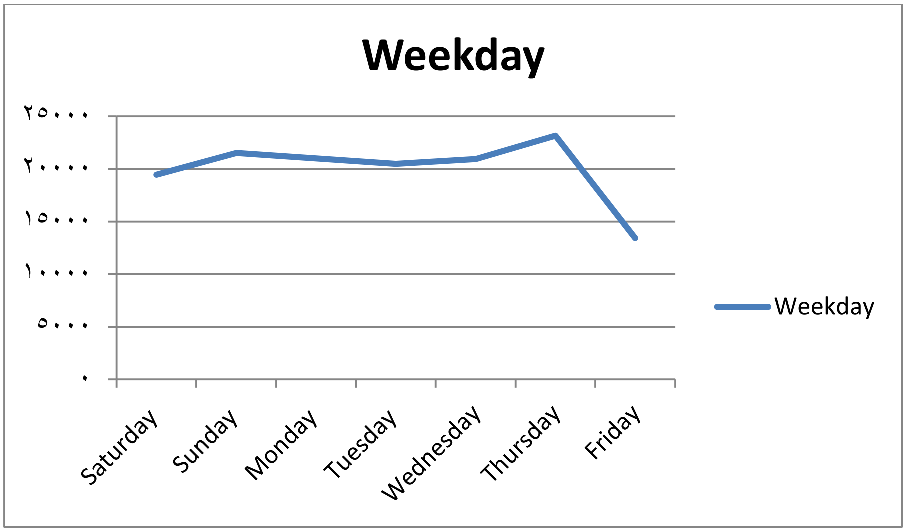

Figure A.2. Traffic Accidents Type by Weekday

3- Traffic Accidents Type by Month

Accidents increase during summer, mainly between July and September. During those months, most educational institutions have their summer vacation, number of citizens not attending to their schools, universities, or colleges increases which results in having more people outside utilizing the roads. The more people within the roads, the more possibility for accidents to occur as shown in (**Table A.3**) and (**Figure A.3**).

Table A.3. Traffic Accidents Type by Month

Turn Over	Pedestrian	Collision	Total	Month
93	310	9856	10259	January
109	252	9369	9730	February
91	369	10111	10571	March
121	394	10479	10994	April
117	355	10472	10944	May
162	381	11102	11645	June
150	372	12337	12859	July
185	327	13502	14014	August
153	359	12460	12972	September
135	333	11974	12442	October
143	354	10989	11486	November
136	285	11677	12098	December
1595	4091	134328	140014	Total

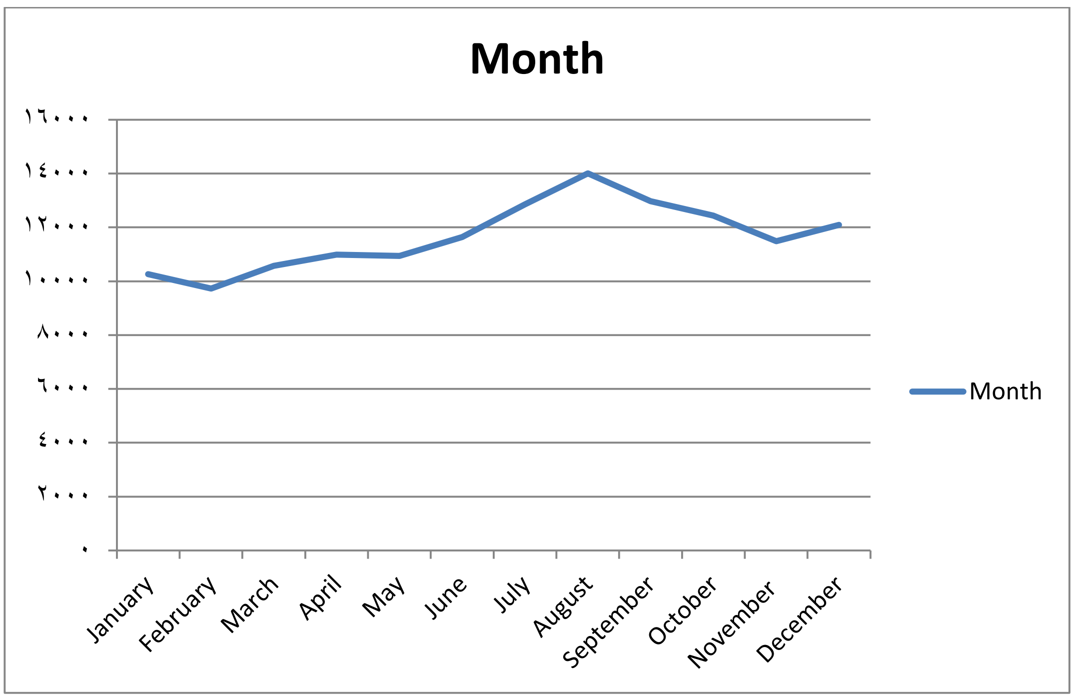

Figure A.3. Traffic Accidents Type by Month

4- Traffic Accidents Type by Speed Limits

Normally roads with low speed limits (40–50) are small roads within suburbs, those roads are narrow, have lots of population, don't provide a good scene for the driver while passing through. Due to that, number of accidents increase within roads of those speed limits as shown in table A.4 and figure A.4.

Table A.4. Traffic Accidents Type by Speed Limits

Turn Over	Pedestrian	Collision	Total	Speed Limits
5	19	1191	1215	20
30	76	1654	1760	30
385	2021	61518	63924	40
232	884	32077	33193	50
358	782	27147	28287	60
113	143	5108	5364	70
249	120	3806	4175	80
64	21	709	794	90
85	20	493	598	100
73	3	400	476	110
1	2	225	228	120
1595	4091	134328	140014	Total

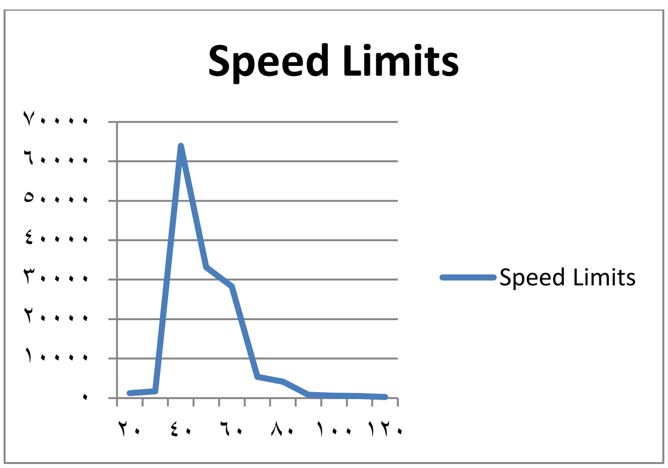

Figure A.4. Traffic Accidents Type by Speed Limits

استبيان حـــول مستخدمي الهواتف الخلوية لدى افراد المنزل

يحتوي الاستبيان على معلومات نرجو تعبئتها وهي معلومات متعلقة بالهواتف الخلوية المستخدمة لدى أفراد منزل الشخص "المقيم في عمان" المعبيء للاستبيان من حيث عددها والعمر التقريبي لأفراد المنزل ممن يملكون هذه الهواتف الخلوية مع ذكر درجة القرابة لكل منهم وتحديد ان كان كل شخص يملك سيارة خاصة به/بها.

عنوان السكن في عمان :---

التسلسل	درجة القرابة						العمر (تقريبا)					عدد الهواتف الخلوية				يملك سيارة	
	نفسه	أم	أب	أخ	أخت	غير ذلك	أقل من 18	25-18	40-26	60-40	60<	1	2	3	3<	نعم	لا
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	

Figure B.1 Questionnaire

Questionnaire About cellular phone users of house members

The questionnaire contains information (please fill) which is related to cell phones used by the members of a person's family, "resident in Amman", in terms of the number and approximate age of the members own these phones with degree of relationship for each of them and determine if each person owns a car

Residential address in Amman:---

SN	Degree of relationship						Age					No of Mobiles				Owns a car	
	SELF	MOM	DAD	BRO	SIS	OTHER	<18	25-18	40-26	60-40	<60	1	2	3	>3	Yes	No
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	

A paper **"Control and Monitoring of an Efficient Traffic Congestion System"**was written based on this research, the paper was published on IJESET (International Journal of Engineering Sciences and Emerging Technologies), following is the new paper.

Triangulation Method. In developed countries a high proportion of cars contain one or more mobile phones. The phones periodically transmit their presence information to the mobile phone network, even when no voice connection is established. By measuring and analyzing network data using triangulation, pattern matching or cell-sector statistics (in an anonymous format), the data was converted into traffic flow information. With more congestion, there are more cars, more phones, and thus, more probes. In metropolitan areas, the distance between antennas is shorter and in theory accuracy increases [9].

An advantage of this method is that no infrastructure needs to be built along the road; only the mobile phone network is leveraged. But in practice the triangulation method can be complicated, especially in areas where the same mobile phone towers serve two or more parallel routes (such as a freeway with a frontage road, a freeway and a commuter rail line, two or more parallel streets, or a street that is also a bus line) [6].

Using this method, almost all vehicles within the roads will be covered within the analysis since all vehicles utilizing the roads have passengers with mobile phones and the mobile towers are available in all roads. Although the usage of the triangulation method has decreased due to its complexity, my research was to reduce those complexities since this method has the lowest cost of implementation due to the readiness of its infrastructure. Through the research we worked on solving all complexities that prohibit the adoption of this method in order to allow the seamless implementation of it within our region ensuring that the least cost and overhead are required within those implementations.

Within the paper we will explain the mobile network structure and how we got advantage from it within the implementation of the solution at section 2. In section 3, an explanation for the problem that this paper resolves is identified and discussed. Section 4 defines the importance of information collected, methodologies of collecting it, and the needs gained from this collection. Within section 5, we have briefed previous work executed by others within the same scope. Both sections 6 and 7 explained the implemented work, activities executed, and algorithms followed in cleansing collected information. Finally, sections 8 and 9 detail the summary and conclusion of this paper.

II. MOBILE NETWORK ARCHITECTURE

Mobile network is built from a set of components that, together, make the correct function and usage of it. The components are:

2.1. The Switching System

The switching system (SS) is responsible for performing call processing and subscriber-related functions. The switching system includes functional units, which are, Home Location Register (HLR) which is a database used for storage and management of subscriptions. The HLR is considered the most important database, as it stores permanent data about subscribers, including a subscriber's service profile, location information, and activity status [10]. Mobile Services Switching Center (MSC) that performs the telephony switching functions of the system. It controls calls to and from other telephone and data systems. It also performs such functions as toll ticketing, network interfacing, common channel signaling, and others [11]. Visitor Location Register (VLR) is a database that contains temporary information about subscribers that is needed by the MSC in order to service visiting subscribers. The VLR is always integrated with the MSC. When a mobile station roams into a new MSC area, the VLR connected to that MSC will request data about the mobile station from the HLR. Later, if the mobile station makes a call, the VLR will have the information needed for call setup without having to interrogate the HLR each time [10]. Authentication Center (AUC) provides authentication and encryption parameters that verify the user's identity and ensure the confidentiality of each call. The AUC protects network operators from different types of fraud found in today's cellular world. Equipment Identity Register (EIR) a database that contains information about the identity of mobile equipment that prevents calls from stolen, unauthorized, or defective mobile stations. The AUC and EIR are implemented as stand-alone nodes or as a combined AUC/EIR node.

2.2. The Base Station System (BSS)

All radio-related functions are performed in the BSS, which consists of base station controllers (BSCs) and the base transceiver stations (BTSs) [12]. BSC provides all the control functions and physical links between the MSC and BTS. It is a high-capacity switch that provides functions such as handover, cell configuration data, and control of radio frequency (RF) power levels in base transceiver stations. A number of BSCs are served by an MSC [13]. BTS handles the radio interface to the mobile station. The BTS is the radio equipment (transceivers and antennas) needed to service each cell in the network. A group of BTSs are controlled by a BSC [11].

III. PROBLEM OF THE STUDY

Within the design of roads in each city, the consideration of vehicles inflation might not be expected accurately. This increase the chance of having more traffic jams when unexpected numbers of vehicles are served at the same time within the same road. Proper analytics on the usage of roads and the number of vehicles being served at a certain time is not available at the moment in most countries. In order to resolve such problems, lots of countries implemented and are implementing different analytics through manual approaches [14].

The research is based upon implementing an automated approach for such analysis; the collection of information available from mobile operators through the data gathering of information from their telecommunication towers shall be followed. The suggestion is to use the data from mobile operators, while the solution will be resolving the complexity in the gathered information due to unwanted information available within it and providing an automated analytics approach on this data in order to reduce traffic jams and provide proactive actions on them. Complexities that may arise are such as parallel roads that are covered through the same tower. A group of mobile phones using the same operating tower within the same vehicle. Inactive users that are utilizing the same tower but are not using the road itself such as population within a building. Walkers on the road that are not affecting the traffic through vehicles utilizing the road. And vehicles within traffic jam which are already idle for a long time and are similar in a certain extent to the walkers within the road.

IV. IMPORTANCE AND OBJECTIVES OF THE STUDY

The implemented method within this research takes advantage of an already available technique which reflects in lower start-up costs, less time in initiation and execution, reduced opportunity of loss of time and cost, and ability to take advantage of a huge amount of historical data that is already collected and not used to serve this approach.

The availability of such huge amount of information will provide the system with the instructions and methods to learn from the historical data it has and allow the experiment of the system's intelligence and recommendations prior to its actual adoption and implementation which results in real scenarios to test.

Within the research, the objective is to resolve the complexity of this approach and its usage in order to be able to notice and predict traffic jams in roads, reduce these jams and provide the ability to proactively interact with certain roads at certain period of times from the historical data analyzed. All actions and behaviors executed through this method are based on the historical data collected and are not real-time information.

The roam of error when depending on historical data shall be reduced by excluding the anomaly factors that might persist within the collected data. The analytics provided will be for the number of vehicles consuming roads at a certain period of time with the exclusion of traffic jams that occur due to accidents, fire, or any abuse of the roads.

From data collected in such systems, a Floating Car Data (FCD) market is only now growing worldwide with a wide range of applications and benefits. This would not only improve traffic management but would also help satisfy the growing demand of drivers who are willing to pay service providers as long as they have access to relevant real-time information: will there be any congestion on my usual route today? How to avoid it? If not, how long will it last? [15].

V. RELATED WORK

Erlangung "deals with cooperative traffic information systems, which support the driver of a car in selecting a route, based on traffic information collected by other cars. System participants contribute measurements of the traffic situation in their vicinity (e. g., current traffic flow speed) and use the measurements made by other drivers to find the fastest route to their destination with regard to the current conditions. Such systems help avoid traffic jams, highly congested roads, place of accidents and other unexpected deterioration" [16].

Guillaume describe their "extended data collection system, in which vehicles are able to collect data about their local environment, namely the presence of road works and traffic slowdowns, by analyzing visual data taken by a looking forward camera and data from the on-board Electronic Control Unit. Upon detection of such events, a packet is set up containing time, position, vehicle data, results of on-board elaboration, one or more images of the road ahead and an estimation of the local traffic level" [2].

McLin mentions that "traffic congestion in metropolitan areas is a major problem because of its negative effects which include increased delay, fuel consumption, and pollution. In the past, capacity expansion was identified as the solution for congestion, however, budget restrictions and other concerns now limits the growth in new infrastructure. Congestion remedies now center on traffic management solutions which better utilize the existing infrastructure. This dissertation describers the development of a regional traffic control system which is designed to provide integrated control in a large traffic network during non-recurrent congestion events" [1].

Stefano stressed on the fact that "the development of Intelligent Transportation Systems (ITS) requires high quality traffic information in real-time. For several years, under growing pressure for improving traffic management, collecting traffic data methods have been evolving considerably and the access to real-time traffic information is becoming routine worldwide" [11].

World Road Association focuses within their handbook "on the “soft” engineering approaches and tools available to the network operator to improve network operations. The handbook discusses:

The shift from the traditional building and maintaining of the road network to a service oriented policy towards the road user; The road network operator's tasks and measures; ITS solutions for network monitoring, maintaining road serviceability and safety, traffic control, travel aid and user information and demand management; The institutional and organizational aspects of network operations; and Performance indicators for network operations" [4].

Fabian "discusses the dynamics of using vehicle probes for detecting link travel times in a large scale road networks. Data from vehicle probes that actively participate in traffic flow, also known as Floating Car Data, offers a solution to this problem. Highly distributed sensor population provides scalability as compared to traditional methods of travel time estimation. This thesis investigates Floating Car Data as a tool for acquiring traffic information. Individual components of Floating Car Data systems are examined along with various alternatives to build such components efficiently" [17].

Wahle "is to combine the contributions of some disciplines with emphasis on the methods of physics. The main objectives of this work are:

to give insights into the concepts of Intelligent Transportation Systems, to discuss potential benefits and drawbacks of providing information in traffic systems, to analyze and evaluate the usefulness of multi-agent system for modeling traffic system, to develop a general agent-based traffic model, which is capable to include the reaction of drivers to information?, to study and model the route choice behavior of road users on different time-scales, e.g., from day to day or within a day and to demonstrate the usefulness of the proposed approaches in practice, especially in combination with data provided on-line [18].

Dubai's Road and transport authority, executed The first technique in Dubai, United Arab Emirates (UAE) through a system called “Salik”, meaning open or clear, is Dubai's new electronic toll collection system was launched in July 2007, which emphasizes the system's congestion management objectives as well as the choice of technology for the toll system. Salik utilizes the latest technology to achieve free flow operation with no toll booths, no toll collectors, and no impact to traffic flow, allowing vehicles to move freely through the tolling point at highway speeds. There are 4 toll points, each time you pass through a Salik tolling point, a specific amount will be deducted from your prepaid toll account using advanced Radio Frequency Identification (RFID) technology, and your vehicle will be identified to the system through communication with the small, thin Salik sticker tag affixed to your windshield [19].

VI. IMPLEMENTED SYSTEM

The main output of the solution is to get results identifying the usage of different roads at different periods of time in order to provide the intelligence on which roads to follow at different periods of times to ensure the minimal traffic jams and expenses. The solution aims towards providing this information upon request and providing recommendations on roads utilization.
Within the solution, the following steps will take action:
Data collection for raw information available at the mobile operators. Each mobile operator will provide data collected for its subscribers consuming the telecommunication services at different locations. Information is collected through the telecommunication towers installed at different locations on the roads; those towers are available to ensure the continuity of the service to those subscribers at different locations. The data includes the tower from where the data was collected, the date and time this reading was done, with subscriber number optionally identified.
Towers information. Information from each mobile operator related to their telecommunication towers installed shall be collected and defined. Information related to a tower includes the latitude and longitude of the tower. This allows the ability to know the location of each subscriber (vehicle) at a certain period of time based on defining the tower that read the subscriber's location.
Duplication of data cleansing algorithm. Within the collected information, more than one person might be within the same vehicle. The algorithm shall define those people by automatically deleting records for subscribers' that have the same readings from the same towers at the same time and date. This means that all those readings are for more than one person utilizing the same vehicle or for one person that has more than one mobile from the same mobile operator.
Inactive subscribers' data cleansing algorithm. Information for residents living beside a certain tower will be always collected and provided within the raw data. Those residents are actually not road utilizers, dot affect the usage of the roads, and don't cause traffic jam. The algorithm will automatically remove those readings by defining the repetitive information for certain subscribers that have the same readings from the same tower and different times and dates.
Group subscribers' data cleansing algorithm. Partially covered in the "duplication of data cleansing algorithm", subscribers within public transport such as bus, taxi, or service cars, shall be removed from the provided information. Those subscribers may be identified through the fact that all those subscribers have the same reading information from the same tower at the same date and time. Information such as this should be deleted from the data and ignored since the utilization of the road is from one vehicle and the number of vehicles is not based on the number of subscribers.
Non-vehicle users' data cleansing algorithm. Walking and running people are not vehicle users. Based on that, those don't cause a traffic jam within the road as they don't actually use it. Those readings are identified through having the readings with minor changes as different periods of times. Those users are not utilizing the roads and don't cause a traffic jam.
Data transformation from raw data store to staging store. Each mobile operator's data shall be stored in a separate data store. The previously explained algorithms shall be executed individually on each data store. Once those operations are executed, the transformation of data from the mobile operators' store shall be executed towards a staging store (database) which includes a collection of all data cleansed from the preliminary databases.
Data cleansing on the staging database. Algorithms previously executed on the raw data shall be executed on the staging database in order to ensure no duplication of data exists in information from different operators.
Data transformation from staging database to data warehouse. After cleansing all data and consolidating it into one location (staging database), transformation activities applied on it to move it from the staging database to the data warehouse. A data warehouse is a normal database with a de-normalized structure which allows building the structure of an analytical cube that provides the ability to analyze data and produce dynamic and intelligent reports. The process will transform the data from the staging database to the data warehouse taking into consideration that current data on the data warehouse should not be affected in order to ensure historical information is perceived and available for future reporting and analytics.
Populate data cubes. As explained earlier, data cubes provide the aggregated collection of data and information in one location. Cubes are built in a manner that allows quick and easy analytics of data

and information through the analysis of data in top-down manner. Moreover, within a cube, dimensions, which provide the analytical views are defined which provide the option of identifying on which item analytics are calculated and based as shown in Figure 1.

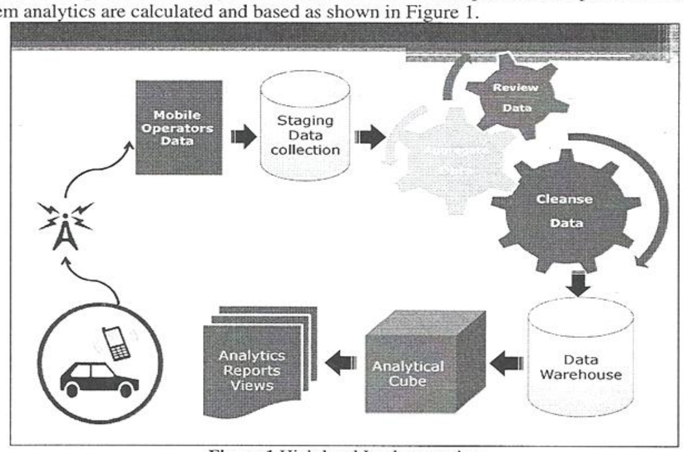

Figure 1 High-level Implementation

VII. Proposed Solution

Upon all the above facts, and since the collection of data from the telecommunication companies was not available due to the unavailability of information at those companies, security reasons, and confidentiality requirements, it was a must to generate the data through a simulating application.
The application suggests a certain number of vehicles passing through a certain route during a certain period of time as shown in Algorithm 1. Moreover, the application uses as an input for it, the number of mobile users utilizing the same vehicle. Based on that, data generated from the application includes all available real-time options that might occur in the real life. This provides the ability to execute and run the algorithms on this data for cleansing and finalization purposes. Later on, data is grouped and analyzed. Dynamic analytics are generated and output is furnished for those items.

1. Vehicles passing within a road:

a. Inputs:

1. **Path** of the vehicle (latitude, longitude) for checkpoints
2. Reading **time** (start and end of reading)
3. The **Avg**(Number of mobiles within the vehicle)generate from the questioner

b. Outputs:

1. **Number of vehicles** passing the path during the reading time.
2. **Number of mobiles** will range from 1 – 5. However, they might reach 50 in some vehicles (such as busses).
3. **Vehicles speed** will be identified from the different readings through the formula **speed = distance / time.**

Results of the algorithm will **provide** the following information:

- Possibilities of traffic **jam** at a certain period of time.
- **Number** of vehicles and **road users** within the reading time.

Algorithm 1 Reading data from the roads (Generating data)

The algorithm 2 explains how to calculate if the road has a traffic jam or not. Jam detection algorithm is one of the main algorithms used within this research. Through this algorithm, identification of the speed of all vehicles passing two points within a road at different period of times and being able to calculate their speed provides the ability to identify whether there's a jam on the road, it's clear, or it's medium.

- From time t_x to time t_y.
- Loop for all available cars.
- Distance of the road **(D) = square root((latitude$_x$)-(latitude$_y$)^2) + (longtitude$_x$)-(longtitude$_y$)^2**)
- Speed = D **(distance of the road)/($t_y - t_x$).**
- If (average speed **(S)** <**Road Speed**) and (**Number of Cars**>= 100), Then there is a traffic jam.
- Key Performance Indicators (**KPIs**):
  - Low Traffic Jam: **Average Speed** is between **90%** of Road Speed to 100% of Road Speed.
  - Medium Traffic Jam: **Average Speed** is between **70%** of Road Speed to 90% of Road Speed.
  - High Traffic Jam: **Average Speed** is less than **70%** of Road Speed.

Algorithm 2 Jam Detection

VIII. RESULT AND DISCUSSION

Data cleansing algorithms were applied on collected data through different phases by applying a certain algorithm on each phase. At the end of these phases, the cleansed data was available for analysis and review.

Within the research we were able to solve the complexities in the data collected from telecommunication operators and cleansing them through the automated process of data cleansing which covered different algorithms to control data at different stages. This provided the ability to benefit from this valuable information that is available historically without using it in such scenarios.

We have analyzed the traffic load within Jamal Abdulnasser Street and Al-Dakhleya Circle for one month and collected the readings at different time periods which explain the load of vehicles utilizing this street.

IX. CONCLUSION

The research was focused on gathering information through the usage of the already available infrastructure within all roads through the telecommunication towers installed. Through the research, we were able to benefit from the algorithms we have followed in cleansing the data which made it available for review and analysis. Those algorithms showed success in resolving the problems within the data collected from those sources. Through those algorithms, unmanageable huge data that was collected became simple and available for analysis.

REFERENCES

[1] McLin and John L, "A regional traffic control system using emerging intelligent transportation systems (ITS) technology", ProQuest Dissertations and Theses, December 2008.

[2] Guillaume Leduc, "Road Traffic Data: Collection Methods and Applications", Working Papers on Energy, JRC 47967 – 2008.

[3] Monahan and Torin. "War Rooms" of the Street: Surveillance Practices in Transportation Control Centers", The Communication Review 10 (4), 2007.

[4] World Road Association, Technical Committee on Network Operations (C16), Road Network Operations Handbook, V116, 2002.

[5] KristiaTorp and Harry Lahrmann, "Floating Car Data For Traffic Monitoring", Aalborg University, Denmark, 2001.

[6] Tarnoff , Philip John, Bullock, Darcy M, Young, Stanley E, et al., "Continuing Evolution of Travel Time Data Information Collection and Processing", Transportation Research Board Annual Meeting, 2009.

[7] Ulrich Fastenrath, "Floating Car Data on a Larger Scale", Report VerkehrsdatenmbH, 2007.

[8] Nele Dedene and Eric Kenis, "Cellular Floating Car Data trial in Flanders", Case study, Aug 2004-Nov 2005.
[9] T. Ottosen and J. Vomlel. All roads lead to rome - new search methods for optimal triangulations. In Proceedings of the Fifth European Workshop on Probabilistic Graphical Models, pages 201–208, June 14, 2012.
[10] Wu, Jie and Dai, F. "Efficient Broadcasting with Guaranteed Coverage in Mobile Ad-hoc Networks", IEEE Transactions on Mobile Computing, Vol.4, No.3, 2005, pp. 259-270.
[11] Al-Kasasbeh Basil M., Muzhir S. Al-Ani, Rafa E.Al-Qutaish, Khalid T. Al-Sarayreh, "Mea Measuring the Optimal Transmission Power of GSM Cellular Network", A Case Study', Communications of the IBIMA, Vol. 4, pp. 216-221, 2008.
[12] Samfat D., Devernay V., Bonnet C., 1995, "A GSM simulation platform for intrusion detection", In:Proceedings of the IEEE international conference on communications (ICC'95), vol. 2. pp. 766-770.
[13] I.V. Kashtanov and O.D. Arhelyuk, "Architectural Solutions Of Testing Network For Mobile Operators", International Conference on Antenna Theory and Techniques, October, 2009, Lviv, Ukraine pp. 225-227.
[14] Vipin Jain, Ashlesh Sharma, Road Traffic Congestion in the Developing World,In Proceeding ACM DEV '12 Proceedings of the 2nd ACM Symposium on Computing for Development, Article No. 11 ,ACM New York, NY, USA ©2012.
[15] Stefano Messelodi et al., "Intelligent extended floating car data collection", 2008.
[16] Erlangung des Doktorgrades, "Cooperative Traffic Information Systems Based on Peer-to-Peer Networks", March 2011.
[17] Peter Fabián, "The End of Congestion: Developing a Large Scale Floating Car Data System", April 2004.
[18] Joachim Wahle, "Information in Intelligent Transportation Systems", Jan 2002.
[19] "Dubai's Road and transport authority", http://www.salik.gov.ae , last modified: 5/07/2012.

Technical Biography

Muzhir Shaban Al-Ani has received Ph. D. in Computer & Communication Engineering Technology, ETSII, Valladolid University, Spain, 1994. Assistant of Dean at Al-Anbar Technical Institute (1985). Head of Electrical Department at Al-Anbar Technical Institute, Iraq (1985-1988), Head of Computer and Software Engineering Department at Al-Mustansiriya University, Iraq (1997-2001), Dean of Computer Science (CS) & Information System (IS) faculty at University of Technology, Iraq (2001-2003). He joined in 15 September 2003 Electrical and Computer Engineering Department, College of Engineering, Applied Science University, Amman, Jordan, as Associated Professor. He joined in 15 September 2005 Management Information System Department, Amman Arab University, Amman, Jordan, as Associated Professor, then he joined computer science department in 15 September 2008 at the same university. He joined computer science department, Al-Anbar University in 18 August 2009 until now.

Shireen Y. Khazneh finalizing MSc. In Computer Science, Amman Arab University, Amman Jordan. Received her BSc. in Computer Science, Petra University, 2000. She has joined Petra University, Amman, Jordan since 2003 as an assistant instructor within the Management Information System Department.