TRAFFIC CONGESTION DETECTION USING DATA MINING IN VANET

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEM

Submitted By
Mohd Anas

2K16/ISY/07

2016-2018

Under the Supervision of:

Ms Anamika Chauhan
(Assistant Professor)

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
JULY, 2018

CANDIDATE'S DECLARATION

I, **Mohd Anas** (**2K16/ISY/07)** student of M.Tech **(Information System),** hereby declare that The project Dissertation titled **"TRAFFIC CONGESTION DETECTION USING DATA MINING IN VANET"** which is submitted by me to the **Department of Information Technology**, Delhi Technological University, in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or some other title or recognition.

Place: Delhi **Mohd Anas**

Date: **2K16/ISY/07**

CERTIFICATE

This is to certify that project dissertation entitled "**TRAFFIC CONGESTION DETECTION USING DATA MINING IN VANET**" submitted by **Mohd Anas (Roll No. 2K16/ISY/07)** Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master in Technology (Information System), is a record of a project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full time for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Ms Anamika Chauhan

Assistant Professor

Date:

(Department of Information Technology)

Delhi Technological University

ACKNOWLEDGEMENT

I am very thankful to **Ms Anamika Chauhan** (Assistant Professor, Information Technology Dept.) and all the faculty members of the Information Technology Dept. of DTU. They all provided us with immense support and guidance for the project.

I would also like to express my gratitude to the university for providing us with the laboratories, infrastructure, testing facilities and environment which allowed us to work without any obstructions.

I would also like to appreciate the support provided to us by our lab assistants, seniors and our peer group who aided us with all the knowledge they had regarding various topics.

MOHD ANAS

2K16/ISY/07

M.Tech (Information System)

Department of Information Technology

Delhi Technological University, Delhi

ABSTRACT

Information Technology, in the past few years has progressed to a level where it is impossible for an aspect of life to not be touched by it. It is being used to the advantage of humanity in solving difficult engineering problems and improving the quality of life. Vehicular traffic is one such area where modern technology has advanced to a phase where the ideas of interconnecting the vehicles on the road are being experimented with in different countries. These networks of interconnected vehicles are commonly known as Vehicular Ad-hoc Networks or VANETs for short. VANETs provide a platform for the implementation of road safety procedures and access to various features of internet like multimedia, emails, etc. VANETs make use of the on-board communication abilities of a vehicle and pre-installed roadside Units to achieve this goal.

One of the most interesting areas of research is the analysis of road traffic. This includes vehicle path tracking, path prediction, intelligent vehicles, congestion detection and many more. Most of the research that has been done to detect traffic congestion used vehicular ad-hoc network (VANET) but of late data mining approach has been applied. Though most of the proposed work has successfully detected traffic congestion, it is complex to come up with an effective mechanism that incorporates detection, control and prediction of recurrent and non-recurrent traffic congestions all in one system.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Vehicular Ad-hoc networks (VANETs)

Today, the safety of drivers and passengers on road remains a major concern as we are witnessing a galloping pace of vehicular activity on roads. To respond to that first hand, Vehicular Ad-hoc networks (VANETs) are designed and put across as systems where vehicles like self-aware entities would become conscientious of their locality and would cater the driver with necessary inputs to safeguard themselves. It serves at tracing crucial traffic events like overtakes and sudden braking by conveying the adjoining vehicles of the condition. It also caters to facilitating aid to the drivers through maps, thus fostering comfort and infotainment by rendering AV content [1].

As a potential requisite, vehicular networking today has streamlined the concerned workforce to develop and modulate applications based on vehicular traffic. A far cry from the tried and tested lab experiments, these applications however, facilitate the process of Intelligent Transport Systems which further aids in overseeing the traffic flow. It also caters to the drivers and help them in keeping a check on their road activity, imparting safety while also deploying infotainment tools for those who travel along [2].

Falling on these lines are automated toll collection system and emergency warning system. Actions are being taken across the globe by concerned workforce in mapping efforts to regulate a systematic module equipped with guiding and communication particulars (principles and protocols) for its effortless realization into the system [3].

Introducing itself on the technological spectrum as a fresh newcomer, Vehicular Ad-Hoc Network (VANET) as a concept aims to integrate Local Area Networks (WLAN), ad hoc networks and cellular networks to employ intelligent vehicle to vehicle communication along with efficient management and safety at the highway traffic realm. What sets VANETs apart from their counterparts on the basis of various critical factors like compound network systems, node movement and many others [2].

Vehicles subjected to VANETs can move at high speeds, running through an on-demand network which is very dynamic and infrastructure-less. VANETs as one of its most critical application, securely transmits datagrams between multiple locations.

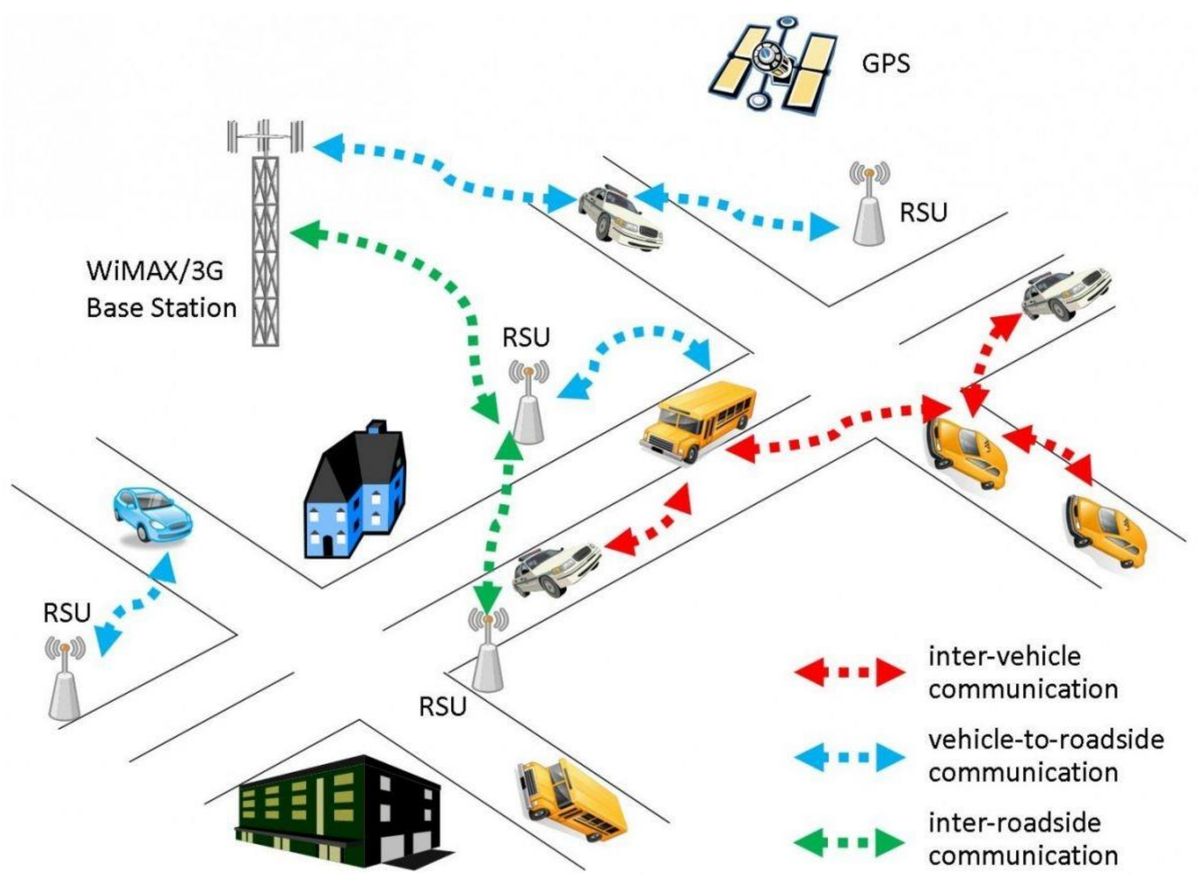

Figure 1.1-AsampleVANETscenario

The design, execution and implementation of Vehicle Ad-Hoc Networks are done in coherence with the integration potential of contemporary technologies in wireless networking. This is kept in line to deliver encompassing connectivity at times when users are on the road, who often are connected to their personal domains like homes and workplaces, segueing into inter-vehicle communication. This makes VANETs to be acknowledged as Vehicle-to-Vehicle (V2V) communications or Inter-vehicle Communications (IVC). A mélange of applications comes across through ITS, which includes traffic flow control, collision prevention, cooperative monitoring of traffic, neighboring information services, blind crossing, and alternative route suggestion in real time. Catering to ever-encompassing connectivity through internet into mobile vehicle nodes allows users to make the most of it like downloading files such as music, etc., play movies, send emails, or play online video games, serves as one of the many applications of VANETs [4].

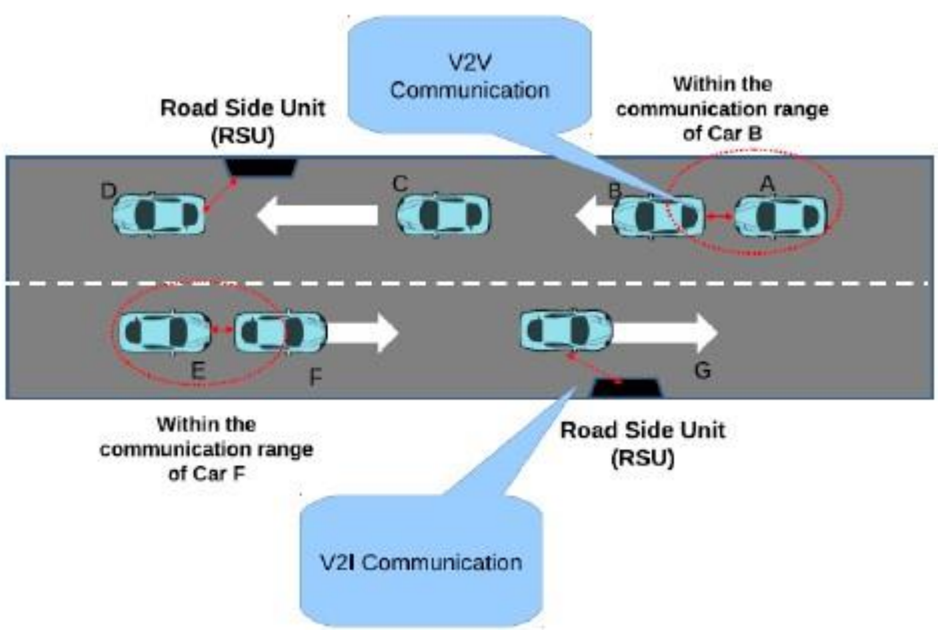

Figure 1.2-CreatinganAd-hocNetworkusingVehicles

As geographical road typology and traffic conditions hinders the distribution and mobility pattern of the vehicles, thus, serving to deliver routing decisions are digital maps and geographical locations of vehicles. However, there might occur various challenges like fluctuating vehicular mobility gives way to unreliable connectivity which furthers culminates into imbalanced service quality. Thus, routing protocols for VANET should be competent enough to tackle these challenges [3]

As often seen in such protocols, mostly the position-based routing protocols are most befitting for VANET. In these cases, accurate knowledge pertinent to network and traffic conditions are of utmost use serving to circumvent the selection of disadvantageous routes. Many studies have come to accept the idea of dissecting a segment of the road into cells of fixed-size leading to selecting a vehicle closest to the center of the cell as the traffic information aggregator in the respective cell [5]. The distribution of vehicle, however, is mostly random and thus, not all of the selected information aggregators be put near to the corresponding cell centers. This culminates into the discontinued involvement of few vehicles in the process of collecting the traffic information owing to the inadequate transmission coverage of the appointed information aggregators. Thus, the accumulated information may not always characterize the actual conditions of the activity on the road

1.2 Applications of VANETs

VANETs expedite a broad realm of utilities – from a simple single hop information distribution of, for example, cooperative awareness messages to the very complex multi hop distribution of packets spread across very large distances. Many applications relevant to the MANETs (i.e., Mobile Ad-hoc Networks) can be put across VANETs, but the actual information often differs. Besides incoherent movements, the mobility patterns of vehicles often act to be of an ordered manner. A large segment of the vehicle can be restricted in their relevant range of motion, to be pushed to tread on a particular road. Thereby, the communication interchanges with the roadside units being characterized to be logically to the point [6].

Most of the vehicular networking applications can be broadly classified into the following three major categories:

1. Active road safety

2. Traffic management

3. Infotainment

1.2.1 Active Road Safety

The key objective for positioning the active road safety applications is to decrease the risk of traffic mishaps and thereby the deaths and injuries of the driver and the passengers in the vehicles [7]. A substantial percentage of roadside mishaps that happen every year across the globe are based on the intersection, the head or rear-end and lateral vehicle collisions. These incidents could be prevented if the drivers are cautioned with relevant information on time prior to the mishap. This can take place through active road safety applications. Correspondence of information between roadside units and vehicles is crucial in these applications which can be applied to foresee collisions, and thus, prevent any setback on time [3]. This knowledge often depicts vehicle location, intersection location, velocity and distance heading. Additionally, the knowledge exchange across roadside units and the vehicles can be served to evaluate the dangerous points and accident-prone areas falling on the roads like potholes or slippery segments [8]. Below are some of the uses of such applications:

1. **Lane change assistance:** The probability of lateral collisions involving the lane changing vehicles with may have a blind spot for larger vehicles, for example, a truck, is decreased, thereby reducing the risk of collision.

2. **Intersection collision warning:** The roadside units or the neighborhood vehicles collect the information regarding the risk pertaining to the lateral collisions for those vehicles that are approaching the junction on the road, and then the information is transmitted to the approaching vehicles thereby decreasing the probability of lateral collision.

3. **Overtaking vehicle warning:** There is also a risk of vehicle collision when a vehicle is trying to overtake another vehicle on the highway. For example, Vehicle A is trying to overtake Vehicle B, and meanwhile another vehicle, let's say Vehicle C is already trying to overtake Vehicle B. This can lead to a collision between Vehicle A and Vehicle C. To prevent such an outcome, Vehicle C can inform Vehicle A to stop the overtaking process.

4**. Head on collision warning:** When two vehicles are travelling in the opposite directions, then there is a risk of a head-on collision between the two vehicles, especially on narrower roads. This situation can be prevented by sending an early warning message to the involved vehicles on time so that the drivers know about the approaching vehicle from the opposite direction.

5. **Rear-end collision warning:** The maximum risk of occurrence of a rear-end collisions is in situations such as a road curvature or a slowdown (for example, hilly areas or sharp curves). In such situations, a vehicle's driver can be informed about the probable danger of a rear-end collision.

6. **Cooperative forward collision warning:** The vehicles can cooperate among themselves to detect the risk of a forward collision incident on time. Therefore, such an accident can then be prevented by assisting the driver.

7. **Emergency vehicle warning:** There can be situations of emergency which require the vehicles on the road to free up an emergency lane for vehicles such as police cars, fire brigades, or ambulances. This can be achieved by using the vehicular networks to inform the vehicles on the road to free up the emergency lane. Such an information can be broadcasted in the entire region by the roadside units and the vehicles.

8. **Pre-crash Sensing/Warning:** There can be some cases where the accident is unavoidable and is certain to happen. To limit the damage from the accident, equipment such as airbags, bumpers, actuators, etc. can be used in an optimum manner. To facilitate their proper use, the information about the predicted crash, the position and the movement of the vehicle needs to be actively exchanges between the vehicles and the roadside units.

9. **Cooperative merging assistance:** At the junctions, the vehicles can use the roadside units and cooperate with each other to avoid collisions while involving in a junction merging maneuver.

10. **Emergency electronic brake lights:** There can be situations when a vehicle on the highway has to hard brake. This may lead to a collision with a vehicle coming from behind. To prevent such an accident, the hard-braking vehicle can inform the other vehicles and roadside units.

11. **Wrong way driving warning:** There is a risk of collision in a situation when a vehicle has detected that it is moving in the wrong way. In such a situation the concerned vehicle can inform its circumstances to the roadside units and other vehicles on the road segment.

12. **Stationary vehicle warning:** There can be situations in which a vehicle may become stationary in the middle of the road owing to a breakdown, or an accident. To prevent causing an accident with the other vehicles, the stationary vehicle can inform the nearby roadside units and vehicles of its situation.

13**. Traffic condition warning:** While moving on the highway if a vehicle comes across some rapid traffic situation which may be of concern, then it can transmit such information to the neighboring vehicles and the roadside units, so as to prevent any mishaps on the road.

14. **Signal violation warning:** There can be a situation in which a roadside unit detects a violation of the traffic light signal. Information regarding such a violation can be then broadcasted by that roadside unit to the neighboring units and vehicles to implement the traffic safety rules.

15. **Collision risk warning:** In a situation where a roadside unit comes to know of a probable collision between vehicles that are without the ability to communicate with each other, the roadside unit can transmit the information to all its neighboring vehicles to warn them of the danger.

16**. Hazardous location notification:** If a roadside unit or a vehicle happens to come across a hazardous location on the highway, for example, an ongoing construction work, or an obstacle on the highway, or a slippery segment of the road, then it can warn the neighborhood vehicles and roadside units about such conditions, and thus prevent a possible accident.

17. **Control Loss Warning:** There can be a situation in which it is required to enable the vehicle's driver to compose a control-loss message and broadcast it to the neighborhood vehicles and roadside units. Such a message can inform the nearby vehicles of the event and they can then ascertain the importance of the event and warn the drivers, if required.

1.2.2 Traffic Management

The primary focus of the application connected to traffic management and efficient traffic flow is the enhancement of the vehicular traffic movements, traffic assistance and traffic coordination along with supplying the updated neighborhood environment data, local maps and vital data packets which are time bound and/or geographic region bound.

Cooperative navigation and Speed management are two standard categories of this type of utility [8].

1. Cooperative navigation: These applications intend the improvement of the traffic efficiency through management of the vehicular navigation on the highway by facilitating cooperation between vehicles and roadside units (i.e., V2I) and through cooperation among vehicles (i.e., V2V). Platooning (allowing vehicles to become electronically coupled "road trains" by closely following a leading vehicle which provides acceleration and steering information), cooperative adaptive cruise control, and traffic information and recommended itinerary are common examples.

2. Speed management: The primary purpose of the applications relevant to speed management is to aid the driver in management of vehicle's speed for unfateful driving and also avoiding unnecessary halts along the highway. Regulatory speed limit notifications is one of the most common example.

1.2.3 Infotainment

Infotainment applications are known as the type of applications which are centered on forwarding media which caters a mélange of information and entertainment.

1. Cooperative local services: The main focus of such application is obtaining infotainment from locally based services, for example notifications about the points of interest in the neighborhood [8].

2. Global Internet services: Such applications are mainly focused on the data that can be acquired from the Internet providers. Some of the most common examples are, music streaming, backseat video game playing, movie streaming, etc. [8].

1.3 VANET Requirements

VANETs require some preliminary networking technologies for seamless operation of the applications. There are some most common vehicular networking applications requirements which are crucial for deployment in vehicular networks [9]:

1**. Radio communication capabilities:** Vehicular networking needs potentials such as radio frequency channels, single hop radio communication range, radio communication channel robustness, available bit rate and bandwidth.

2. **Network communication capabilities:** A vehicular network should be potential enough of different modes of discourse such as unicasting, broadcasting, multicasting, geocasting (broadcasting only within a specified geographic region). Additionally, the network should be equipped with applications for congestion control, data buffering, message priority, IPv4 and IPv6 looking after support, channel and connectivity management, management of mobility relevant to the alterations in the point of attachment to the Internet.

3. **Vehicle absolute positioning capabilities:** A vehicular network can utilize facilities such as Global Positioning System (GPS), Global Navigation Satellite System (GNSS), etc. to step-up its operation potentials.

4**. Vehicle communication security capabilities:** Provisions should be provided in vehicular network for revering anonymity and privacy of the users, confidentiality and integrity of data, authenticity of received data, and protection from the external security threats.

Chapter 2

LITERATURE SURVEY

2.1 Network Architectures

MANETs (Mobile Ad-hoc Networks) mostly do not dependent on a particular infrastructure for communication between nodes and transmission across data packets. Bringing forward the same foundational concept, VANETs (Vehicular Ad-hoc Networks) puts it into a magnetic realm of vehicular transportation [2].

VANET architecture can be classified into three broad categories:

1. Pure Cellular or WLAN architecture

2. Pure Ad-hoc architecture

3. Hybrid architecture

2.1.1 Pure Cellular or WLAN Architecture

VANETs takes use of WLAN access points and fixed cellular gateways at traffic junctions to connect to Internet, gather information about traffic or for routing requirements, presenting a pure cellular or WLAN vehicular network architecture. They can integrate into both WLAN as well as cellular networks culminating into an architecture that makes use of a 3G connection with no availability of WLAN access point.

This however brings to perspective the impracticality of a pure cellular-based vehicular network due to burgeoning costs of communication between base stations and vehicles, also the large number of handoff incidents taking place at base station, brings into thought the high mobility of vehicles [4].

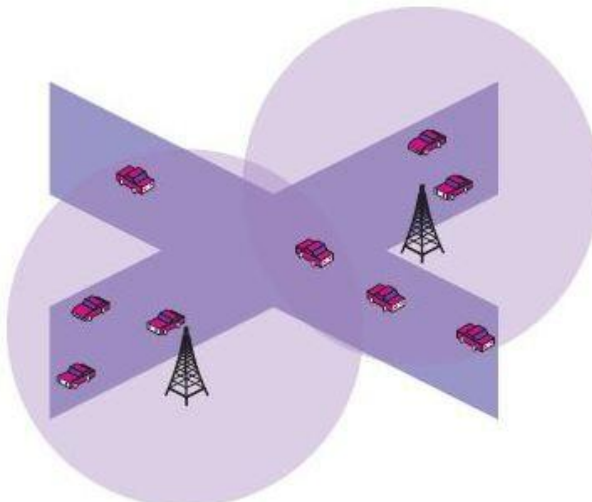

Figure 2.1-Pure WLAN or Cellular architecture

2.1.2 Pure Ad-hoc Architecture

Through these cases, all the roadside wireless devices and adjoining vehicles can pave way to mobile ad hoc network to streamline inter-vehicle communications, and further put into use applications like crossing uninterrupted eschewed by any light control (blind crossing). Stationary gateways encompassing roadsides can forward connectivity to the mobile vehicles but culminates to be impractical in aspects of costs involved for the entire infrastructure [4].

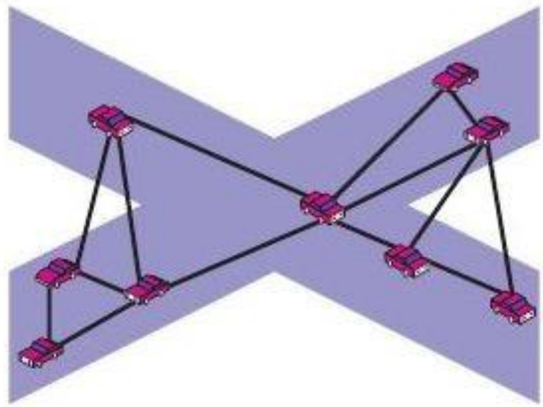

Figure 2.2-PureAd-hocarchitecture

2.1.3 Hybrid Architecture

A mélange of the above two architectures culminates into taking the form of hybrid architecture. It makes use of nodes which presents both cellular as well as WLAN capabilities. Here, these nodes (i.e., vehicles) are put to serve as gateways and routers so that they can be communicated using multi-hop links by only those vehicles that have WLAN capability [10].

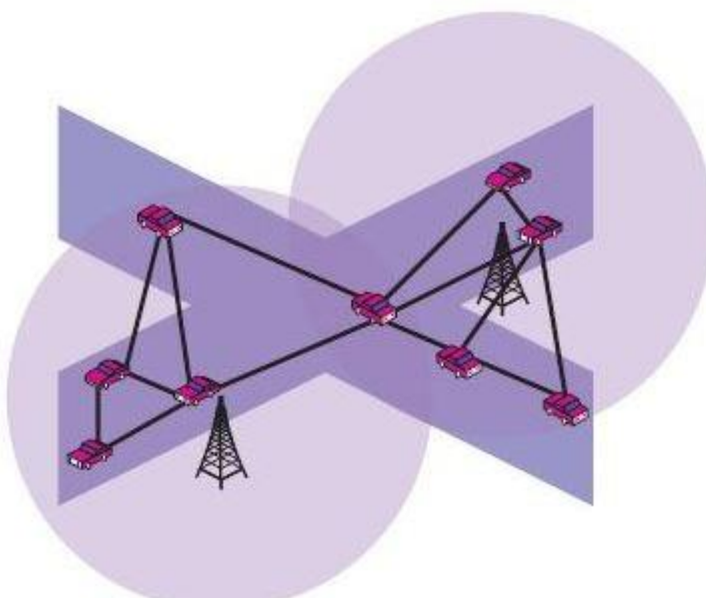

Figure 2.3-Hybridarchitecture

2.2 Network Characteristics

Vehicular Ad-hoc Networks are composed of vehicles having radio communication capability which presents itself as mobile nodes and also as routers for the other nodes in the network segment neighborhood.

In spite of the striking coherence with other ad hoc networks like low bandwidth, short range of radio transmission, self-management, and self-organization, Vehicular Ad-hoc Networks can be separated from other contemporaries on the following features [2]:

2.2.1 Highly dynamic topology

Due to the high mobility of the nodes (i.e., vehicles), the topology of vehicular networks changes constantly. For example, assuming that each vehicle has a transmission range of 250 meters for wireless communications, thus, there is a communication link between two vehicles if they are less than 250 meters apart from each other. But, considering the case in which the two vehicles with the speed of 25 meters per second start to move in opposite directions from each other. In such a case the communication link will only last for a mere 10 seconds in the best case.

2.2.2 Frequently disconnected network

Take the instance where the vehicle density dips tremendously in a segment of the highway. In this scenario, there is a high probability that the network may get disconnected in that segment of the highway. This leads to infrequent connectivity of vehicular networks. Some applications need constant access to the Internet where they do not work in bad network connectivity. Rectifying this issue brings to the fore pre-deployment of various access points or relay nodes along the highway to maintain the connectivity. However, it could accelerate the cost of vehicular networking.

2.2.3 Sufficient energy and storage

As the nodes in the vehicular networks are vehicles, thus they possess sufficient battery power, processing capabilities, and storage capabilities in comparison with most other ad-hoc network implementations which may incorporate handheld devices or small sensors equipped with very limited resources. For example, Wireless Sensor Networks (WSNs) have very limited processing capabilities and battery power. This however, is not seen in Vehicular Ad-hoc Networks.

2.2.4 Geographical type of communication

Many other relevant applications use standard concepts of networking like unicasting, broadcasting, and multicasting. As end points, it culminates either into ID of the node or group. Whereas in Vehicular Networking, a whole new type of communication is rendered to look after the various geographic locations. The messages to be sent across can be time bound or region bound.

2.2.5 Mobility modelling and prediction

Nodes in the vehicular networks are mostly barred in mobility patterns by the predefined streets, roads, or highways, thus, catering the velocity of the vehicle and the geographical map, the future location of the vehicle can be approximated to a certain degree depending upon the traffic conditions. And due to the highly dynamic topology and mobile node movement, prediction and mobility models puts across a very crucial part in design of the network protocols for vehicular networks.

2.2.6 Various communications environments

Vehicular networking is often used in two communication realms. First runs through the highway traffic layout where neighborhood of the node is very seamless and unchallenging, and most node movement is witnessed in single dimension. To establish communication links in such environments therefore, is realized very easily. The second aspect falls on an urban layout where neighborhood of the nodes is very convoluted because of the city's typology. Here, roads could be set apart by diverse obstacles like trees, flyovers, buildings etc. This leads to various challenges in keeping track of the communication channels between vehicles as weak signals often results in disconnection of networks.

CHAPTER 3

PROBLEM FORMULATION

Road traffic is a significant problem in Urban societies. People waste lot of time and huge amount of fuel everyday as their vehicles are stuck in traffic.

There are lot of applications through which vehicles can detect the traffic scenario and analyses the issue and make decision, it can be detected where the traffic is high and make decision accordingly. Vehicles communicate to other vehicles (V2V) and they can also communicate to infrastructure as well (V2I). **Vehicular ad-hoc networks (VANET**) is a type of **mobile ad-hoc networks (MANET)** which can be used for communications between nearby vehicles and nearby fixed gadgets.

Areas of high traffic density and low speed traffic can be tracked detected using Congestion detection algorithms. Each vehicle captures and disseminates information from other vehicles in the network to calculate data such as Lat, long positions and velocity at which the vehicle is moving and submit the data. Many trials have taken place to implement congestion detection in VANETs. VANETs is used to make the smart traffic and it is designed in such a way that it avoids the traffic to fall into the congestion and aware the driver to make the alternative route is just one of its many applications. Heavy traffic congestions have very huge effects on the market and the life of the people and their surroundings as we waste a lot of time in congestion rather than doing some useful work which can contribute to environment, and all these can be very vastly reduced by the serious impact of traffic congestion detection system

Data Mining algorithms can be used for analyzing the traffic scenario in the present with information gathered by vehicles using **V2V communications** and then processing it irrespective of the devices present (such as antennas, satellites, etc.).

3.1 Problem Statement

There are many factors that lead to traffic congestion like road constructions, office hour rush or bottle-necks at critical routes and some cannot be predictable like crashing the vehicle, nature imbalance and uncertainty. As the drivers are not aware of the traffic after few minutes ultimately, they become a part of it increasing its severity.

If a driver would know the traffic situation ahead he/she can try to seek alternate routes saving time and fuel. In cases of incidents like an accident or temporary disruptions, if many drivers would already know the situation they are going towards, congestions can be curb and It'll only affect the vehicles which are present there at particular point only.

There are cases where driver behavior leads to the traffic congestions and unavailability of current traffic. Present relief systems, such as helicopter traffic information require a lot of resources and are not available/practical in remote locations, it is very useful since it provides the whole network scenario like how long the traffic is, and exact location from where it's starts and ends and at what speed it is moving. So as to provide the knowledge of the traffic system to drivers the system must:

a) Find out the location, boundaries, acuteness and estimated time.

b) Provide accurate info to drivers stuck in the traffic and also to others who are going near it.

These are the main factors for a successful traffic congestion system. Observer can identify a congestion, a person in a plane, have to look at vehicles which are a long distance away from each other, and outside of each other's line of sight. A clear picture of the congestion can be created from a place where there is a good view of the whole situation. For drivers already stuck in the traffic to create its own scene of a traffic they have to integrate with the vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication.

Once the drivers get the clear picture of the traffic scenario ahead, they can take suitable steps to avoid the traffic by taking alternate routes, informing others somehow related to the traffic in some ways.

CHAPTER 4

PROPOSED WORK

4.1 Traffic Congestion Detection

Areas of high traffic density and low speed are found out by Congestion algorithms. These algorithms are designed for this purpose only. Information is decimated by each of the vehicle that it obtains from its own design and the nearby objects. The knowledge from the different nodes present in the vicinity is also accessed.

Among various other applications of VANET, Congestion detection is one. Its application is not in the department of automated driving but rather it is used as a tool for the delivery of information to the driver which helps him/her take decisions accordingly in order to avoid heavy traffic. The Collision avoidance applications [11] [12], which are not discussed in this part make use of similar kind of GPS positioning information for the purpose of alerting the drivers about possible dangerous situations like a road accident, uncertain vehicle or an hazard, the response time needs to be very less in these kinds of applications. There are numerous applications that can be created by VANETs [13].

4.2 Proposed Solution

A very essential feature of these congestions is their slow relative movement and very high vehicle density. We will be using the same to our advantage while we will be designing our system for detecting congestion.

4.2.1 Distributed vs. Centralized

There are two ways on which we can rely as we develop a congestion system. One would be to depend on one central network and data gathering infrastructure. The other way information can be collected, analyzed and disseminated is that we can use chip installed in the vehicles. The most efficient approach towards traffic congestion would be a distributed approach as they have more reliable and major investment in infrastructure is not required since the costs will be divided among various drivers, therefore this work is specifically focused on finding distributed solutions.

4.2.2 Reliability

If we put these devices in each of the vehicle, it will in turn increase the reliability of the entire network since the probability of a number of nodes failing all at the same time is highly unlikely in comparison to failing of one central component. The redundancy factor will also come into play and will marginally increase the reliability since the vehicles which are close to one another will be capturing and will be transmitting the very same information virtually.

4.2.3 Ease of Deployment and Cost Effectiveness

Gradually the deployment of a distributed network can take place; vehicle factories can install the devices while manufacturing the vehicle or these devices sold in the market the same way the GPS navigation systems are sold. Every user will be paying for its parts and its maintenance and will be benefitting directly from its functionality which will make the equation of cost-effectiveness very transparent. With the growth in the number of users, the effectiveness of the system will grow furthermore and the system will become more and more useful and will in turn have a bigger appeal to the new users. However, in case of a centralized system, large investments is required in the infrastructure even before the network can start to provide any sort of service.

4.2.4 Comparison with Previous Work

Table 1 shows a comparison between the most relevant related works.

Table 1. Data Mining Approaches

Table 2. VANET Approaches

Table I. Data mining approaches

Author	Detection	Control	Prediction
[3] Lin *et.al*	Yes	No	Yes
[4] Amelia *et.al*	Yes	No	Yes
[5] Giannotti *et.al*	Yes	No	No
[7] Wang *et.al*	Yes	No	No
[8] Gupta *et.al*	Yes	No	Yes
[9] Florido *et.al*	Yes	No	Yes
[10] Yang *et.al*	Yes	No	No

Table II. VANET approaches

Author	Detection	Control	Prediction
[11] Jain *et.al*	Yes	Yes	No
[12] Ramon *et.al*	Yes	No	No
[13]Milojevic *et.al*	Yes	No	No
[14] Roy *et.al*	Yes	Yes	No

4.3 Components of a Distributed Congestion Detection System

4.3.1 Information Gathering

Information gathering from the environment is the most common step in detecting the congestion. There is a requirement for each vehicle to be equipped with one or more devices which are very much capable of gathering data like the current location of the vehicle and the speed at which the vehicle is moving. This essential data constitutes the fundamental building blocks of the system and it changes rapidly.

4.3.2 Information Sharing

Numbers are the major source of strength in any distribution system. Sharing of information between the vehicles is of extreme importance. Wireless transmission and receiving of information to other Vehicles and from other vehicles in the network is necessary by the creation of a vehicular ad-hoc network (VANET). In order to achieve that, all the vehicles need to be equipped with a device which allows the vehicles to perform this function. The job of congestion detection algorithm is to choose what data has to be processed and what factors needs to keep in mind for analyzing the current scenario.

4.3.3 Information Consolidation

The information gathered by the vicinity of the vehicle and the information transfer from distanced vehicles needs to be combined for the purpose of eliminating the inconsistencies and to control the amount of information that requires processing.

4.3.4 Knowledge Creation

After the consolidation of information, the information will then be converted into knowledge. This knowledge would be about the congestion in the traffic. And it is required for the detecting the congestion as a final step. This knowledge needs to be transferred to the driver with the help of some visual interface which will show the related congestion on the device.

4.3.5 Knowledge Propagation

The information will be disseminated in order to assist the vehicles far away from the traffic congestion zones who are clueless about the congestion ahead and they are able to benefit from it. This knowledge is going to be much compact and more meaningful than Brew information. As soon as the device is able to determine the existence of congestion at particular point, this knowledge would be disseminated to nodes far away as this knowledge would prove to be of more utility than raw data. Also, the network's workload would have a much lesser impact as this version of knowledge would be a condensed one of a larger amount of data.

4.3.6 Knowledge Consolidation

The knowledge which would be received from different nodes would be constantly combined and will be re-broadcasted in order to reflect the change in traffic in real time. The job of a congestion detection algorithm is to give present traffic alert and reliable information, so as to get that, constant verification and constant re-evaluation of the information received is extremely necessary.

4.4 Other Considerations for Congestion Detection Systems

4.4.1 Bandwidth Utilization

The practicality of a Detecting congestion lies in the smart use of the bandwidth, transmission of lot of knowledge without any concern for the bandwidth limitations will overload the system and will make it difficult to operate. Information should be smartly consolidated and the irrelevant as well as out of date data should be filtered. This is of the utmost importance. In environments such as VANET transferring of information of V2V it becomes very essential to provide means in order to reduce the redundant communications without it significantly having an impact on the network's effectiveness.

4.4.2 Vehicle Participation

Now a day, hardware's are equipped in the vehicles which would be exactly the same in each of the vehicles allowing them to gather the knowledge and share it among the different real-world technologies should be designed in a way that it is able to work and perform despite of low numbers of vehicles which are present on the network to contribute to the detecting congestion. There may be other factors like of heterogeneous networks in which nodes have other different capabilities. It can be kept of the future scope of the project.

4.4.3 Privacy

Subject of loss of privacy would be a major point of concern after new cooperative networks systems such as VANETs are deployed. No would like to give away the knowledge of where their vehicle is at all the time. We need to be very much aware of such issues while developing congestion detection algorithms over VANETs. Mechanisms should be included in order to curb or avoid such issues till the step that it can be acceptable by the user. Most of the thesis work on congestion detection in VANETs ignores this very essential factor and they make use of unique vehicle IDs that have been persisted over time which can detect the location of the vehicle by tracking the node., for this particular reason. In this research the proposed solution doesn't required the unique id.

Removal of unique vehicle IDs surely ensures privacy at the application level, however, there is a risk of compromise of privacy at lower levels. By the use of Radio transmission, we can track the location of the vehicle.

4.4.4 Security

A legitimate concern after privacy would surely be security for real world systems. Prevention of sabotaging and abusing the system by users is paramount. In recent years, many applications have been developed to provide secure vehicle to vehicle communication [14] [15] [16].

CHAPTER 5

CONCEPT OUTLINE

5.1 Data Mining in VANET

With the rapid modernization of vehicles today, it would not be long before they come equipped with a wireless contrast. It can enable the structuring of vehicular networks, commonly referred as VANETs, as befitting models of mobile ad hoc networks which brings across cars to be the mobile nodes.

VANETs (Vehicular Ad hoc Networks) are broadly mobile wireless ad hoc networks and will assist a crucial part in deploying public safety communications and commercial practices. The intent of mobilization of these VANETs span the nuances of the accelerating dynamics of high speed car.

Continuing to be the de rigueur are Position based routing protocols due to its progression and reach with GPS devices. Often, a common question referencing VANETs gives way to splitting of the corridor high speed vehicle nuances, further treading to a severed step and then towards a low throughput and high overhead costs.

VANET in the last year has taken a magnetic position of research and deliberation; one of the imperative idea is deploying data mining techniques to retrieve functional patterns form VANET, which subsumes classification, clustering, association among others.

Here, through this paper we bring across applications of VANET which employs data mining techniques.

5.1.1 Introduction

Vehicular Ad Hoc Networks (VANET) serves as one of the examples of ad-hoc network applications. The amalgamation of ad-hoc networking with contemporary wireless network is done in order to aid short range communication across vehicles on the road.

Intended to serve utmost safety and comfort of the passengers on road, a program for intelligent wireless services is created. These comes across through projects like CAR-2-CAR communication consortium, Car Link and Fleet net.

As we can witness today, devices like Global Positioning System (GPS) which facilitates the power to oversee the setting of the vehicle, its direction and speed in motion comes equipped in the vehicular assembly.

Any vehicle can serve to be the node of the network and begin getting and relaying messages when it comes positioned with wireless communication devices like Onboard Unit (OBU) with transmission signal range between 100 to 300 metres. In the cases where the vehicles miss out the signal range, it is when other vehicles join in.

Road side units enable the vehicles to connect to external networks like internet, RSUs comes segregated and employed rarely to be junction points across vehicular nodes and nodes of central authority (CA) which regulates the network.

The proposed architectures of VANETs are classified as the following three categories.

1) **The Wireless Wide Area Network** (WWAN) has the access points of the cellular gateways fixed which enables direct correspondence across vehicles and access points. These points due to their costly installation makes them extremely unfeasible.

2) **The Hybrid Wireless Architecture deploys** HWAN access points at specific junctures whereas ad hoc communication caters to access and communication across access points.

3) **Ad Hoc V2V Communication** does not need any fixed access points to build communication for the vehicles. Comes installed with wireless network card, swift positioning of an ad-hoc network could be executed for every vehicle.

Vehicular Ad hoc Networks (VANETs) serve to enable wireless communication across vehicles on road. This subsumes the road side wireless sensors allowing the communication of relevant knowledge which takes care of safety and planning for dynamic routing as well as fostering mobile sensing and entertainment for the user.

The present challenge has been of the limited range of wireless sensors on the vehicles that spans close quarters to a few hundred meters while the aspects of the traffic across vehicular urban environment mostly alters.

Besides this, VANET routing protocols bear witness to challenges like unstructured roads, the setbacks in the sizes of potential junctures in specific positions, sharp edged curves on roads, unkempt slopes and other hindrances like large sized structures, trees, sign boards etc.

Because it is not feasible to utilize enormous cost on restructuring the roads in the contemporary scenario, a routing protocol serving the purpose of long distance data communication along one to one transfers pertinent for VANETs is crucial to be realized.

5.1.2 Applications of VANETs

Vehicular network applications span from road safety applications aligned to the vehicle or its driver towards entertainment options and commercial uses for passengers, making way for a large number of co-operating technologies.

The central vision for vehicular networks subsumes real time & safety applications catered for passengers and drivers, fostering safety for the former and providing imperative tools to think of the best route.

These applications intend to curtail accidents and enhance traffic conditions by facilitating passengers and drivers with helpful information which includes road sign alarms, in-place traffic view and collision warning.

• **Safety applications:** Road safety could be enhanced by VANETs through sharing the imperative knowledge across vehicles on road, weather situations and accidents.

Vehicles moving along the range of the accident will receive warning messages by the vehicles that encompass any accident. This would bring knowledge of the accident to the notice of more drivers on road, before them reaching the spot.

This would result in appropriate actions like decelerating the speed, put to use electric break lights, or in case of landslides, turn conflict or road curve, accidents can be avoided through sharing warning messages at the right time to vehicles on road.

• **Convenience application**s: If the adjacent vehicles on road decelerate their space, it signals road congestion, thus the traffic management vehicles spread this knowledge to rest of the vehicles falling on the network, to change their way. Through this application, there are techniques available that aid the vehicles to hunt an available space in the parking lots and helps collecting toll with no need for long waiting. It is also imperative in this application to keep away from collision domains, specifically when at an intersection.

• **Commercial applications:** The broad objective is to enhance passenger's comfort, safety and efficiency on road, in times of traffic. This may subsume the nearest POI settlement, present traffic and weather conditions along with interactive communication.

This could be applied to all varieties of applications which may put to use on top of TCP/IP stack. For example, instant messaging and online games.

Some other application is retrieving information from commercial vehicles and infrastructure on road's side for their actions, propaganda radio. Shopping malls, hotels, gas station and fast food restaurants can put their gates for use of transferring fixed marketing data to potential customers that pass by.

5.2 Data Mining Techniques for VANETs

5.2.1 Association Techniques

Association Techniques Association Rules were used to extract information relevant to certain events classified on minimum value of service and support catered in the underpinning criteria in how data mining becomes a delicate process. These rules could be used to find common features among juveniles as they appear within the same realm of presented dataset.

The application of Data Mining Techniques in Vehicular Ad Hoc Network and the discernment of false vehicles under data mining techniques makes it important to ensure the dependence and feasibility across safety applications with VANETs as it depends on the soundness of contacts across vehicles.

In the process of detecting faulty vehicles, especially those which are not connected to the district persuader to Association Rules, this process proposed a strategic mechanism referred as VANET Association Rule (VARM). It has to oversee all information about every neighboring vehicle for retrieving rules and temporal relations across vehicles that take part in area of transmission.

This approach is unlike other and one of a kind which unveils defects through generation of Association Rules over vehicular activities. Classified on the present vehicular status knowledge and the retrieved rules, a strategic mechanism akin to the pattern employed to gauge relationship across current events in information system that gives way to accidents and loss of lives.

Taking use of this mechanism, it can present drivers with the most feasible and efficient processes which should be noticed to prevent accidents or to decrease the effect of accidents. Data mining method to retrieve models presents the vehicular crossing movement whilst the journey from source to culminating destination.

This style module for movements acts as a subset of a prolific data mining method called sequential models. The pioneering objectives of this system includes applications of rules across movement models to envision the path of a certain vehicle. This envisioning is very helpful in the following situations:

(1) when determining the shortest path a vehicle can get; (2) envisioning the easiest and most often taken routes by criminals, making way for police to find them on time, and (3) aiding to find alternative roads for vehicles in situations of congestions, accidents etc.

5.2.2 Clustering Techniques

Vehicle ad networks are one of a kind models of MANETs (Mobile Ad hoc Networks) where the mobile nodes are modes of movement with features of both random as well as deterministic. Deterministic features of the vehicular movements subsume following speed of vehicular front and driving kept within speed limits. Whereas the random aspects subsume changing lane, passing over other vehicles, radical change of speed due to accident or appearance in heavy traffic and excess speed. Ample motivation lies in clustered network due to immense difficulties within VANETs. These however brings many problems happened due to routing across intensive as well as rapid movement environment, with problems across invisible terminals. Excess crowding and the challenge of invisible terminals could be prevented through clustering the network. Ahead, for VANETs to deal with delayed information as well as safety messages, Quality of Service is needed and included within network. To handle Qos requirements, Clustering techniques are useful. The pivotal use of this system is to facilitate a process for securing the wasted fuel and time spent in movement through traffic jams and the spread of the discerning pattern of traffic jams. The main difference seen between this system and that of others is the facilitation of more accurate real time relevant to jam information and ease and cost of implementation in all kinds of roads.

In order to efficiently find and share the traffic patterns and due to the high mobility of the VANETs, this proposed system employs circulated techniques alongside high performance communication. The system employs Global Positioning System (GPS) along with peer-peer wireless communication medium (801.11 or 801.15).

Each vehicle presents its map classified on discerning the speed of nearby vehicles treading on the chosen roads and all vehicles afterwards that exchanges speed maps with one another. Through the network, any vehicle can get speed map for each road even when these roads are not being used by it.

The proposed system presents the traffic as set of clusters in slow traffic, each node on the network oversees a summary statistic of these clusters. The node communicates in times when recorded speed is kept outside with variance in expected speed in the road segment. Nodes transfer and receive the statistics by epidemic correspondence where each node calculates high level cluster employing the local statistics and the ones encompassing the network.

3.3 Classification Techniques

With this technique of data mining a model is created for presenting a group of pre-conceived categories across a group of tuples referred as training group. It is designed to find a group of unknown creatures according to few characteristics they exhibit. This technique requires the input which is to be entered by user and means it is a supervised technique. The input becomes a training group that is used to create and train the classified model. Aiding the security required for diverse applications in VANETs, the authors intends to present an approach in this realm. To extract this output, the authors uses classification mining techniques in evaluating a wide range of applications in VANETs, its classification based on safety necessities (this includes high level of verification requirement, return potential attack), and presents safety for every class falling in the application. Through the use of this model, a new way of implementation can be addressed under the classification of safety procedures and appropriate safety actions to be applied. VANET misinterpret that exploits many sources of knowledge accessible in a VANET background empowering on-board unit to pick difference between threat alerts and lawful alerts.

The sources of information were previously individual instrument aiding to find misconduct, here are the five sources below:

(1) Cryptographic Authentication: for checking the digital signature if it's included with the message or not for confirmation reason;

(2) Source Location: for checking the sender location to make sure that he can send the message to other vehicles or not;

(3) Local Sensor: for checking the ability of the local sensors to support the alert;

(4) Infrastructure Validation: for checking the ability of the roadside unit (RSU) to support the message and

(5) Sender Reputation: for checking if any previous messages from the sender were harmful messages.

Chapter 6

IMPLEMENTATION

6.1 Obtaining Data Set

The trajectory data containing parameters such as latitude, longitude and timestamp is taken. The dataset should exhibit congestion events. The Data used is taken from the UCI site.

6.2 Pre-processing

Pre-processing is performed on the trajectory data so that errors and missing values are eliminated or minimized.

6.3 Cluster Formation

The Clusters formation is done to identify vehicles within the same space and travelling at the same time.

6.4 Calculating Distance

For each cluster and given the timestamp of each trajectory point, calculate the distances between two trajectory points (using Haversine formula, and determine its speed using the distance/time).

6.5 Congestion Detection

The detection of congestion will be done by comparing the cluster speeds obtained, to a speed threshold that is assigned based on the road capacity.

6.6 Controlling Congestion

The Congestion can be controlled by providing the awareness to the vehicle by showing Normal/Heavy Congestion message. .

Chapter 7

RESULT ANALYSIS

7.1 Traffic Congestion Detection

1) Main window

The main Window consist of Data set which must be uploaded from the system.

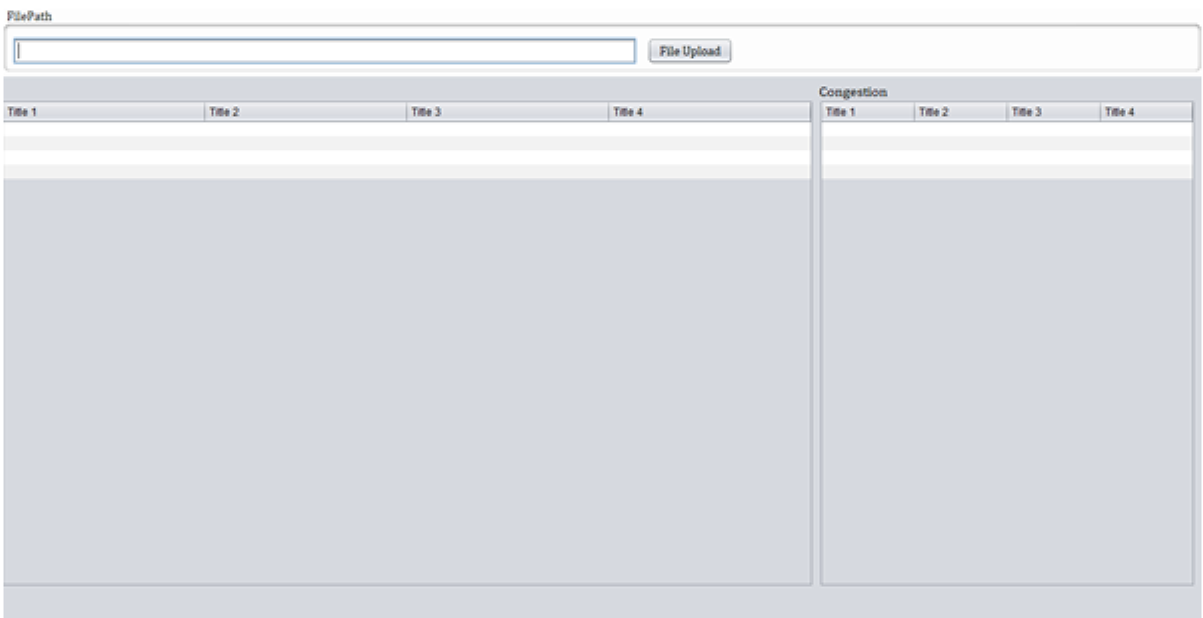

Figure 7.1-Main Window

2) Uploading the Data set

After uploading the data set, the Clustering Tab will be enabled The Data Set will consist of various fields like S.No, Lat, Long, Time and Timestamp.

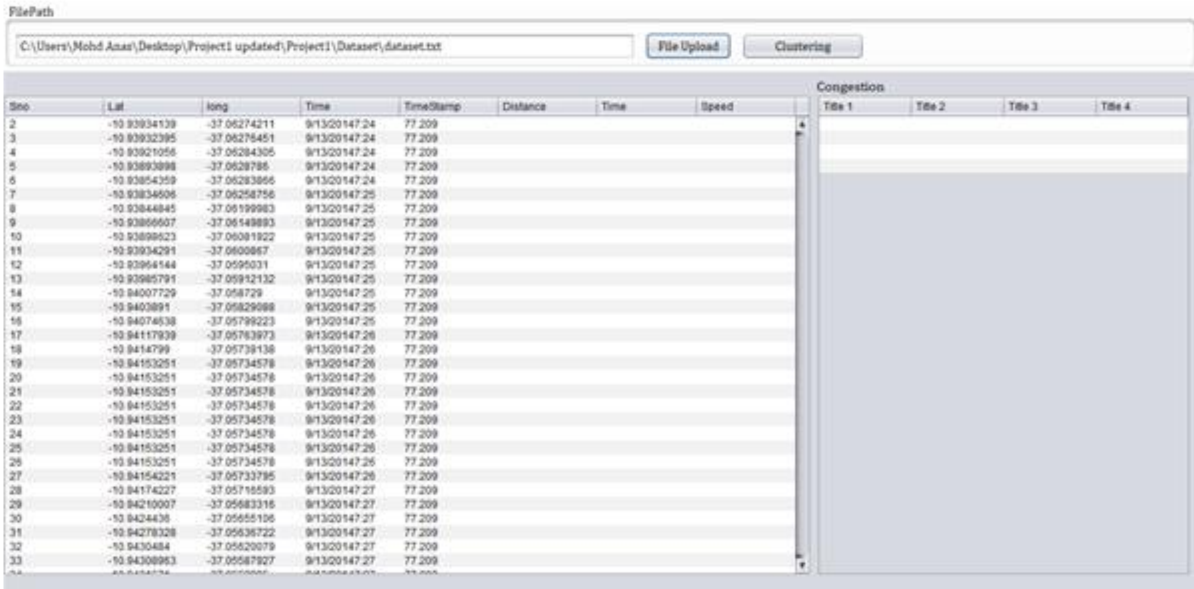

Figure7.2-Data Set

3) Clustering window

After uploading the data, the clustering Tab will be activated and after clicking on the tab the new window will be generated as shown in fig.

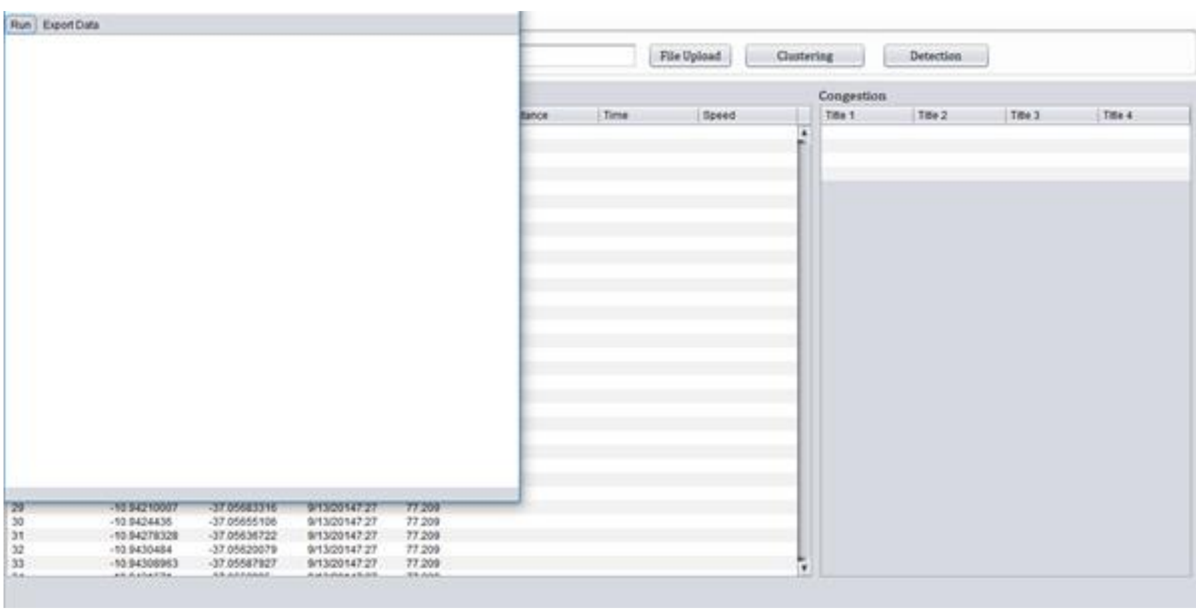

Figure 7.3-Clustering Window

4) Cluster Formation

In Clustering Window, after clicking the run button the clusters will be formed.

The clusters are formed by k means algorithm

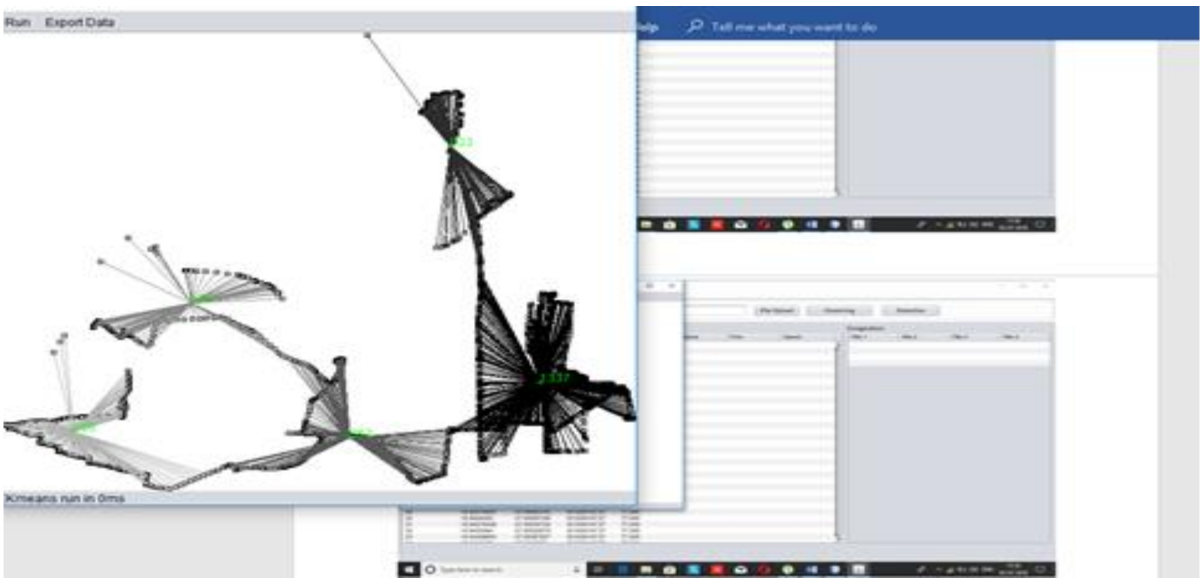

Figure 7.4-Cluster Formation

5) Detection Window

Based on the clustering, the detection is bone by comparing the average speed of the clusters to the threshold speed. If the speed is less than the threshold speed, it detects the congestion.

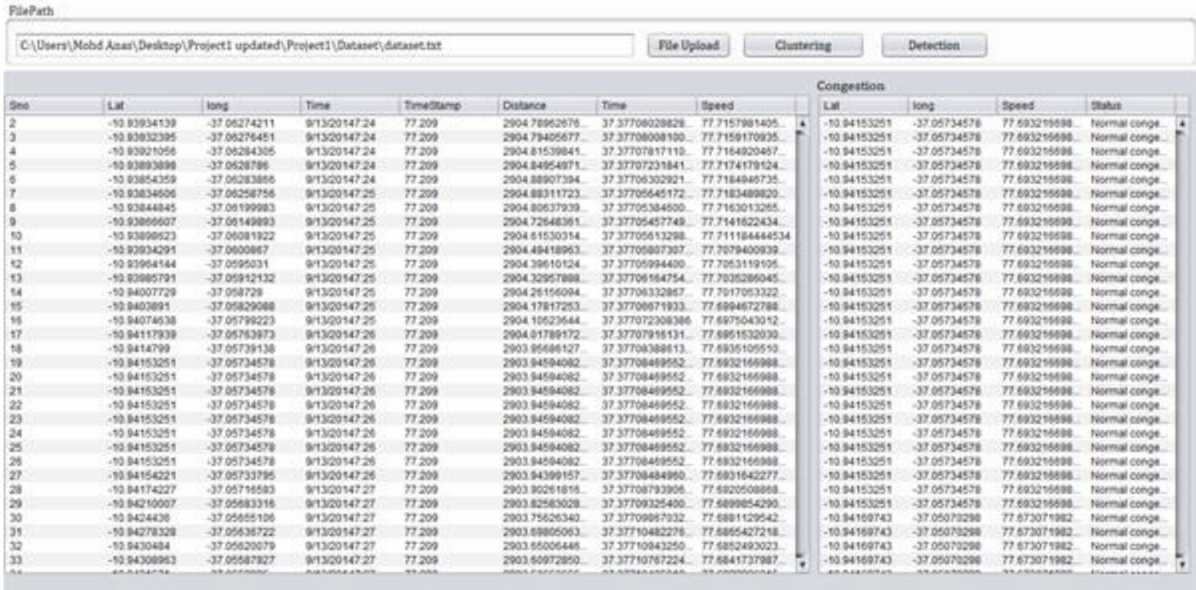

Figure 7.5-Detecting Congestion

6) Displaying message to the vehicles

Based on the speed of the vehicle, the message is shown to the vehicles if they increase the speed to a certain level since it will lead them into the congestion. So the normal/Heavy congestion will be alerted to the vehicles so as to avoid them falling into the congestion

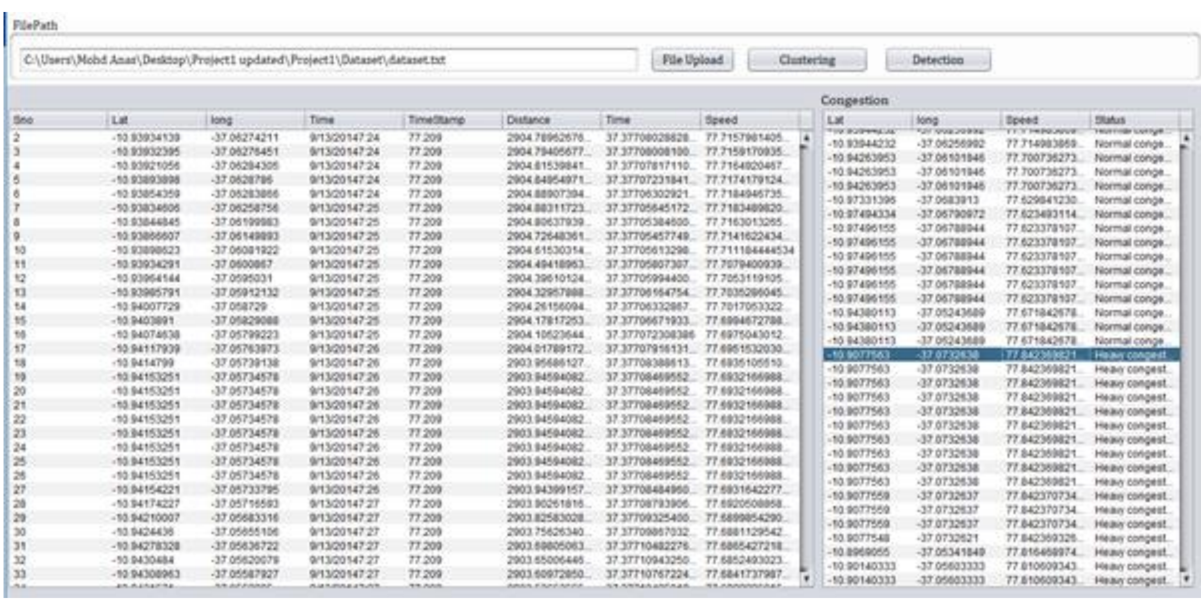

Figure 7.6-Displaying message

Chapter 8

CONCLUSION AND FUTURE SCOPE

8.1 Conclusion

We have use the Data Mining strategy in the VANET to detect the Congestion in the road traffic. VANET have the Feature of communication of Vehicle to Vehicle and Vehicle to Infrastructure. Data Mining is focused on the Detecting the Congestion based on the information Received from GPS by applying its Mining Techniques. Data Mining use the Historical Data to perform the Prediction and process the information to make it more useful to understand. It can Detect the various Patterns and Trends in the Data which is useful for Detecting the congestion in VANET.

8.2 Future Scope

We have detected the Congestion on the highways, roads scenario in day to day life. for future work the same can be designed for complex road structures, though the algorithm presented in this work. Moreover, simulation can be done to generate the complex road structure and apply the technique.

We can also add the IOT Feature in the VANET to avoid the Congestion which Required the Hardware installation.

REFERENCES

[1] Wai Chen, Ratul K. Guha, Taek Jin Kwon, John Lee and Irene Y. Hsu (2008), A Survey and Challenges in Routing and Data Dissemination in Vehicular Ad-hoc Networks, IEEE International Conference on Vehicular Electronics and Safety.

[2] Fan Li and Yu Wang (2007), Routing in Vehicular Ad Hoc Networks: A Survey, IEEE VEHICULAR TECHNOLOGY MAGAZINE.

[3] Georgios Karagiannis, Onur Altintas, Eylem Ekici, Geert Heijenk, Boangoat Jarupan, Kenneth Lin, and Timothy Weil (2011), Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions, IEEE COMMUNICATIONS SURVEYS & TUTORIALS.

[4] Sabih ur Rehman, M. Arif Khan, Tanveer A. Zia, Lihong Zheng (2013), Vehicular Ad-Hoc Networks (VANETs) - An Overview and Challenges, Journal of Wireless Networking and Communications.

[5] C.R. Lin and M. Gerla (1997), Adaptive clustering for mobile wireless networks, IEEE Journal of Selected Areas in Communications, vol. 15, no. 7, pp. 1265–1275, 1997.

[6] Mayank Dixit, Rajesh Kumar, Anil Kumar Sagar (2016), VANET: Architectures, research issues, routing protocols, and its applications, 2016 International Conference on Computing, Communication and Automation (ICCCA).

[7] M. L. Sichitiu and M. Kihl (2008), Inter-Vehicle Communication Systems: A Survey, IEEE Commun. Surveys Tutorials, vol. 10, no. 2, pp. 88–105, 2nd Quarter 2008.

[8] ETSITR102638, Intelligent Transport System (ITS); Vehicular Communications;
Basic Set of Applications; Definition, ETSI Std. ETSI ITS Specification TR 102 638 version 1.1.1, June 2009.

[9] C2C-CC, Car to Car Communication Consortium Manifesto: Overview of the C2C-CC System, Car to Car Communication Consortium, Tech. Rep. Version 1.1, 2007.

[10] V. Namboodiri, M. Agarwal, and L. Gao (2004), A study on the feasibility of mobile gateways for vehicular ad-hoc networks, in Proceedings of the First International Workshop on Vehicular Ad Hoc Networks, pp. 66–75, 2004.

[11] Rawashdeh, Z.Y. and Mahmud, S.M. Intersection Collision Avoidance System Architecture. *5th IEEE Consumer Communications and Networking Conference.* 2008, pp. 493 - 494.

[12] Mak, T.K., K. P. Laberteaux and R. Sengupta. A Multi-Channel VANET Providing Concurrent Safety and Commercial Service. *Proceedings of the 2nd ACM international workshop on Vehicular adhoc networks.* January 2005.

[13] Ian Cassias, Andrew L. Kun. *Vehicle Telematics: A Literature Review.* s.l. : Technical Report ECE.P54.2007.9, 2007.

[14]Kargl, Frank, Ma, Zhendong and Schoch, Elamar. Security Engineering for VANETs. *Ulm University, Institute of Media Informatics.* 2006.

[15] Hubaux, J., Capkun, S. and Luo, j. The Security and Privacy of Smart Vehicles. *IEEE Security and Privacy Magazine.* May-June 2004, pp. 49-55.

[16] Raya, M., Papadimitratos, P. and Hubaux, J.-P. Securing Vehicular Cmmunications. *IEEE Wireless Communications Magazine.* Special Issue on Inter-Vehicular Communications, 2006.

[16] E. Chung, "Classification of traffic pattern," In Proc. of the 11th World Congress on ITS, pp. 687-694, 2003.

[17] W. Weijermars and Eric Van Berkum, "Analyzing highway flow patterns using cluster analysis," In Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, pp. 308-313, 2005.

[18] Lin Xu, Yang Yue and Qingquan Li, "Identifying Urban Traffic Congestion Pattern from Historical Floating Car Data," In Procedings of Social and Behavioral Sciences, Vol. 96, pp. 2084– 2095, 2013.

[19] A. Amelia and P. Saptawati, "Detection of potential traffic jam based on traffic characteristic data analysis," In International Conference on Data and Software Engineering (ICODSE), pp.1-5, 2014.

[20] R. Ong, F. Pinelli, R. Trasarti, M. Nanni, C. Renso, S. Rinzivillo and F. Giannotti, "Traffic Jams Detection Using Flock Mining," In European Conference (ECML PKDD-11), pp. 650-653, 2011.

[21] M. Benkert, J. Gudmundsson, F. Hübner and T. Wolle, "Reporting flock patterns," In 14th Annual European Symposium, pp. 111-125, 2006.

[22] Z. Wang, Min Lu and X. Yuan, J. Zhang and Van de Wetering H, "Visual Traffic Jam Analysis Based on Trajectory Data," In IEEE Transactions on Visualization and Computer Graphics, Vol. 19, No. 12, pp. 2159 – 2168, 2013.

[23] A. Gupta, Netaji Subhas, S. Choudhary and S. Paul, "DTC: A framework to Detect Traffic Congestion by mining versatile GPS data," In 1st International Conference Proceedings of Emerging Trends and Applications in Computer Science (ICETACS), pp. 97 – 103, 2013.

[24] E Florido, O Castaño, A Troncoso and F Martínez-Álvarez, "Data Mining for Predicting Traffic Congestion and Its Application to Spanish Data," In 10th International

Conference on Soft Computing Models in Industrial and Environmental Applications Volume 368 of the series Advances in Intelligent Systems and Computing, pp. 341-351. 2015 .