I. INTRODUCTION

Urban population has seen an unimaginable growth in the modern era, and consequently it directly affects their mobilization in huge cities. According to a United Nations report, the percentage of urban population in the whole world constitutes about 56% in 2015, with a steep increase of 1.84% every year. According to the predictions, by 2050, about 64% of the developing world and 86% of the developed world will be urbanized [1]. With this astounding increase in urban population, there is a necessity for an effective system to combat with one of the unprecedented challenges, which is traffic congestion in big cities.

In the traditional traffic management system, ineffective traffic lights with predefined timers are used, along with manual control by police officers. Without taking an account of real-time traffic data for consideration, it can happen that a "green light" is granted to an empty lane while a lot of cars are lined up at a "red light" on the other lanes because the same time interval of green lights are granted to every lane.

The proposed system consists of a circuit embedded in each vehicle in commutation. The users can interact with the system either through wired or wireless connection of their smartphone with mounted board. This system uses Radio Frequency Identification (RFID) which plays a vital role in the research paradigm of Internet of Things (IoT). Instead of using GPS (Global Positioning system), this system uses a more efficient LPS (Local Positioning System) for locating a vehicle with the help of localized workstations situated at optimal points. Data analysis involves the implementation of big data analytics with clustered workstations constituting a regional computing unit, which maximizes the throughput. The basic functionalities of the proposed system include:

1. Smart Traffic Light Control System that works dynamically based on the concentration of vehicles in a specific region.
2. Parking Space Identification and Allotment System, with the placement of spatial sensors in parking lots that communicate the availability of spaces with the regional workstations.
3. Anti-theft System that automatically retrieves the location of a stolen vehicle and automatically disrupts the functioning that vehicle.

Big data analytics is used for processing the terabytes of data received from the vehicles. The proposed system is quite advantageous with the use of local workstations consisting regional processing units that receive data from each vehicle pertaining to a specific regional radius. These data are then reallocated based on volume of incoming data using Hadoop MapReduce tool. This drastically reduces data traffic which would occur when a single centralized control unit is used for analyzing the data from each vehicle. Hierarchical clustering and density based clustering techniques are used to process the data. Frequent pattern mining is used in order to derive results to enable efficient traffic management. Moreover, the efficiency of system is maximized with the use of LPS (Local Positioning System) instead of (Global Positioning System) as the time complexity is reduced.

Apart from this, the proposed system also determines the various attributes of transportation along a road using supervised learning. The various attributes analysed are the standard of roads, overall traffic flow, the average speed of vehicles in a particular road, travel path of a vehicle.

II. LITERATURE REVIEW

A. Analysis of Traffic Congestion in Cities

For the last few decades, the human population in the urban regions all over the world has drastically grown, surpassing the corresponding value in the rural areas. Figure 1 represents a graph depicting the world trends in world population from 1950-2050, with the abscissa representing the years and the ordinate representing the population in millions. As per 2016, the overall urban population accounts for 4.1 billion out of the 7.4 billion of the entire world population. According to predictions, the urban population would account 6 billion out of a total 9 billion [2]. Thus, the urban population has seen a huge upsurge from 1950 and is expected to skyrocket in the coming years.

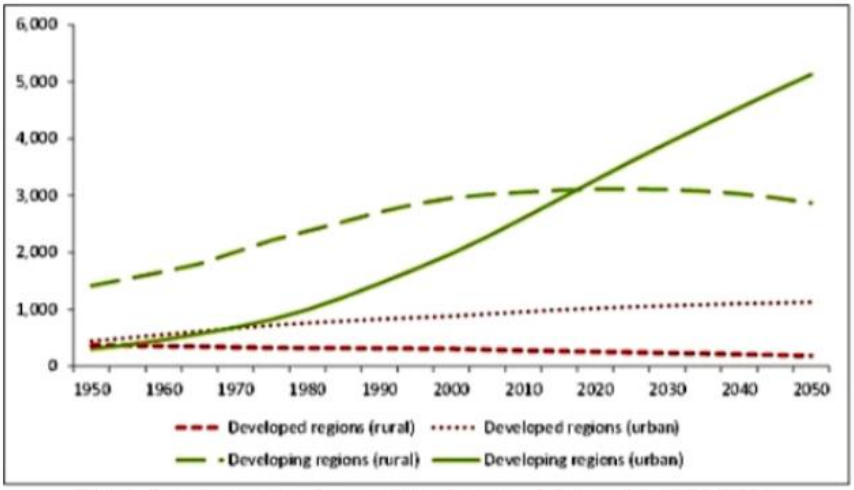


Fig. 1. Population in urban and rural areas of the world, 1950-2050.

The demand for transportation too increases with the rise in population and hence the number of vehicles have also risen. It is essential that the population growth is complemented by the infrastructural development to enhance urban mobiilty. But in the current scenario, the conventional traffic management system is ineffiecient with the use of traffic lights nad manual police control. This in turn leads to heavy traffic problems and jams and as a result increases vehicle emissions and consumes our precious time. Brussels, the capital city of Belgium, stands at the top of the list of most traffic affected cities in the world, as reported by Forbes [3].

The increase in numbers of cars on streets is not the only reason for the traffic problems to appear; there is lack of planning to deal with this amount of cars. For example, it is very common for someone to wait a traffic light to become green even if there is no car in the street. This might intensify traffic jam, which would require manual control by policemen.

The problem of traffic jams in cities is not an easy problem to address. It is necessary to improve public transportation system and provide people with means to be less dependent on their cars. In addition, it is also necessary to stimulate other possibilities of transportation, to improve safety on the streets, to increase people security and walkability through the streets. This substitute would hardly be welcomed, which necessitates the need for an efficient traffic management system, overcoming the problems of the existing system.

B. Challenges faced by the Transportation Industry

With the ever increasing vehicles on road, the transportation industry faces the following unprecedented challenges.

1. Transportation industry can't meet the rapid growth of data both in volume and variety. The data of transportation industry have rich sources, diverse types, and new sets of data are produced continually. Dynamic data generated by various sensors, such as induction coil at bayonet point, infrared detector, microwave detector, ultrasonic detector, laser detector, video detector, and so on, are of huge volumes. The data are generated by GPS vehicle location tracking system and other mobile device search equipments yearly have raised over the threshold level. The amount of data generated by the transportation industry in a city per month has exceeded TB (terabyte) level, are developing from PB (Petabyte) to EB (Exabyte) levels. A massive data storage space and equipment is required and it must have fault tolerance and stability.

2. Smart Traffic Management System in Amsterdam

The Amsterdam city is one of the busiest regions of the Netherlands. Three road managers are active; the municipality of Amsterdam, the province of North Holland and the national government. They all try to optimize the traffic flow within their management area, but the measures they take sometimes conflict with each other. To improve traffic flow cooperation between all parties was necessary. The city of Amsterdam now uses TrafficLink's SCM system that is connected to the traffic system of the national government. Both centres can see what is going on, on just one screen. This is how the centres can jointly and automatically manage traffic within the region. Since the initiative for regional cooperation and the implementation of intelligent traffic management, the percentage of vehicle loss hours in the Amsterdam area has dropped by 10 percent. Amsterdam's smart traffic management system can easily be prepared for connection with in-car and navigation equipment. This way, also in the future, Amsterdam can have its own modern digital road manager, which helps optimizing traffic flow within the whole region [5].

D. Internet of Things in Traffic Management

IoT is an emerging research paradigm and apparently the discovery of its body of knowledge is still in an infancy stage. So, the exact definition, architecture, scope, and standard are still not concretely defined [6]. However, most scholars agree on the idea of expanding and interpreting the pioneering conceptual definition of Kevin Ashton who defined IoT as "a standardized way for computer to understand the real world". With a key feature to create a smart environment together with quick response to support certain decisions and/or operations of human, IoT-based systems have been proposed in several applications such as supporting disabilities,

managing diabetes therapy, building smart home, improving safety in mining operations, and using IoT for an intelligent relationship and is sometimes interchangeably used with a Ubiquitous Computing (or Ubicomp) [8]. The slightly different is that Ubicomp does not necessary require the internet connection to enable communication among objects. It is used rather in a broader sense for defining smart ecology than IoT. For example, applying sensor technology to capture data to make a certain response is Ubicomp but not IoT. Internet of things, to put in simpler words, is the interaction between things, which, according to our case, is the communication between a vehicle, a centralized processing unit and traffic lights. The following equation describes IoT in a simplified manner [7].

$$\text{IoT} = \text{Physical Entity} + \text{Controller, Sensor, Actuators} + \text{Internet}$$

Internet of Things can be compared to the communication among humans, with respect to the equation above. For instance, humans have biological sensors such as ears, eyes, skin, taste buds, etc. to perceive what is happening in their surroundings. Human use body parts to make and receive sounds and require a name to enable others to get an attention. Communication can only succeed when the communication medium exists, e.g. telephony network, mobile network, air, etc. In order to let things carry the property similar to human's, sensors need to be attached to physical objects being considered. Objects, both sender and receiver, must have names, and digital communication is required.

This paper focuses on the communication between vehicles, processing unit and traffic lights. In this case, each vehicle acts as eye, which transmits traffic data. The communication is established by a socket programming over Wi-Fi connection, so ports are like 'mouth' and 'ear' of sender and receiver consecutively. Vehicles 'name' each other by calling RFID Reader name. One of the contributions of IoT to human operations in this scenario is that it can replace human in doing tedious exhaustive work. For example, traffic data can be collected and sent out persistently, which is nearly impossible when manual communication is applied.

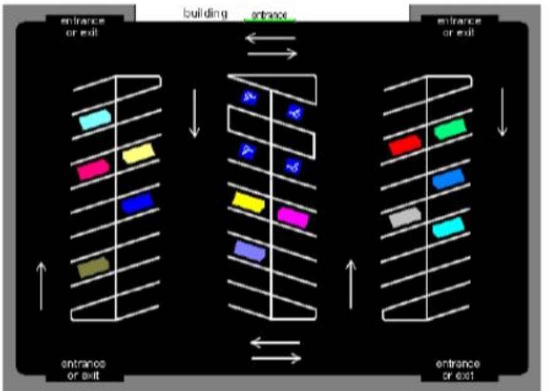
The above block diagram represents the flow control of the proposed system. The system consists of two main components: the vehicle mounted board and the regional computing unit. The vehicle mounted board has transmitter and receiver which sends and receives radio waves. Regional computing units are placed in workstations that receive data from vehicles and transmit signals to each vehicle pertaining to a specified range of area using RFID [9], [10]. The power supply to the embedded unit is provided by using electrical energy from the vehicle's battery. The end users can interact with the system by wired or wireless connection of their smartphones with embedded system using GSM module. In wired means, users can connect their smartphone with the vehicle using a dock, whereas in wireless method, the smartphone can be paired with the mounted board using Bluetooth [11]. The user interface displays the traffic intensity in the particular road along with the available parking spaces. Moreover, they are also provided with the functionality of keying in a particular location and determine the index of traffic congestion and information about parking lots.

Fig. 3. The Vehicle Mounted Board

The regional computing units send data including the vehicle's location, its velocity and other primitive attributes. Instead of Global Positioning System (GPS) the system uses Local Positioning system (LPS) to locate a vehicle on a road. In a GPS system, the location is determined by trilateration, which is the overlapping of circular regions covered by signals received from three different satellites pointing to a specific location. While in a LPS system, local beacons like radio broadcast stations are used to determine the location of a vehicle. LPS system has an edge over GPS system as the former has much lesser time complexity compared to the later. The location of each vehicle determined using LPS is transmitted to the regional computing unit, which in turn shares the collected information with the other regional units. Data analysis is performed using MapReduce to determine the net traffic flow in a particular region and control traffic lights. This helps in estimating the net manoeuvrability rate of a road and other parameters [12].

Parking Identification and Reservation System

Spatial sensors are set in the parking lots of the city. These sensors detect empty parking spaces and send information to the regional workstation in closest proximity. These workstations transmit the available parking spaces to embedded system in the vehicle. The drivers can parallely determine the parking areas during their travel. They could allot and book parking spaces using their docked or paired smartphones.



2. Big data can improve the efficiency of transportation industry.

Transportation industry, involving many aspects of work, need to handle massive amounts of data. It has more uncontrolled models of application and a great deal of equipments. If a little accident occurs, the entire system will run into an inefficient state. On implementing big data technology, the information system can process the data and discover the accidents in good time, automatically handle it, or report to the management staff and ask them to make decisions. Big data has a good predictive ability, can reduce the probability of false alarming and ignorance of traffic congestion incidents. Traffic guidance is an important part of intelligent transportation systems. By publishing guidance information for travelers, it can indicate traffic conditions of downstream road, allow travelers to choose the right travel path and improve traffic situation in the city. In the aspect of improving transport efficiency, improving the capacity of the road network, adjusting traffic demand, big data technology has obvious advantages.

3. Big data can improve the safety level of traffic.

The real-time processing capabilities of big data can accurately probe traffic accidents, its predictive ability can effectively predict the occurrence of traffic congestion. Using radio wave detection systems, video surveillance systems and mobile detection system, one can build an effective security model to improve the safety of vehicles. When security lapses occur and if emergency rescue is needed, with its comprehensive processing, decision-making and rapid response capability, big data can greatly improve the ability of emergency rescue, and reduce casualties and property losses.

B. Architecture of the Data Analysis Layer in Proposed System

Fig.5. Architecture of the Data Analysis Layer

The proposed consists of regional computing units in localized workstations. These workstations receive continuous signals from the vehicles pertaining to that particular region. The data received is analyzed using big data analytics implemented in Hadoop [13].

Hadoop is an existing big data analytics tool used for process scheduling and for analysing huge volumes of data. It uses Map reduce techniques in order to process the data. The data received onto the local processing unit in centralized and reallocated based on the volume of incoming data in the particular region [14]. This helps in dampening the time complexity involved in processing the data which is evident in the usage of distributed systems. The incoming data is first mapped using the map function and then it is delegated to the various linked systems and reduced to the information which is brought back together. The reduced data is then distributed across various regional workstations depending on the locations. The results are straightaway reflected into the traffic light control system accordingly [15].

We use various techniques such as Hierarchical Clustering and Density based Clustering in order to process the data. We then use frequent pattern mining in order to arrive at conclusions regarding efficient traffic management.

Hierarchical clustered Analysis (HCA) is one of the cluster analysis techniques, used to analyze data by building grouped clusters based on rank. Aglomerative and divisive analysis are strategies involved in hierarchical clustering. A dendrogram is used to represent this form of clustering, where map and reduce functions are computed in a greedy manner.

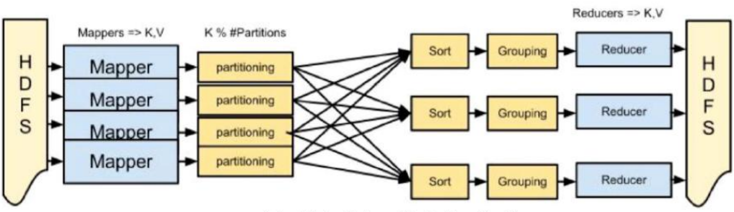


Fig. 6. The Hadoop MapReduce Pipeline

C. Supervised Learning to analyze attributes of traffic management

Supervised learning techniques are used for constantly improving the system's information processing. The training data sets are given in order to assess the various attributes of the traffic management [16].

The list of attributes to be analyzed is the following, for which the training data sets have been fed

C. Supervised Learning to analyze attributes of traffic management

Supervised learning techniques are used for constantly improving the system's information processing. The training data sets are given in order to assess the various attributes of the traffic management [16].

The list of attributes to be analyzed is the following, for which the training data sets have been fed previously for proper judgment.

1. Number of vehicles per square unit area.
2. Degree of manoeuvrability of the road
3. Friction with the road
4. Engine capacity
5. Driver Behaviour

In supervised learning algorithms, the basis of classification is training the classifiers using some training data [17]. Based on the training methods and the characteristics of the classifiers, there are five major algorithms which are used to classify and optimize the data. They are

1. k- nearest neighbour
2. Decision trees
3. Naïve Bayes
4. Logistic regression
5. Support vector machines

Apart from that, in order to boost the classifying properties such as accuracy and time complexity certain meta algorithms are used. One example of such algorithm is Ada-boost meta algorithm.

Support vector machines

Support vector machines algorithm is an algorithm which is used to define the classifier in the form of a hyper plane. A hyper plane is a classifier which is used to separate three dimensional data. This hyper plane requires intensive training from all ranges of training data in order for it to classify data with utmost accuracy. Hence, the proposed system implements this algorithm over the others [18].

AdaBoost is a Meta algorithm. Meta algorithms are ways of optimising the classifiers by using one or more than one supervised learning algorithm. AdaBoost combines the features of several algorithms in order to accurately define the classifiers.

Bagging and Boosting are the two phases used in the AdaBoost meta algorithm. Bagging, also called as Bootstrap aggregating, is a technique where the data is taken from the original dataset 'n' times to make 'n' new datasets. . The datasets are the same size as the original. Each dataset is built by randomly selecting an example from the original with replacement. Boosting is a technique similar to bagging. In boosting and bagging, the same type of classifier is usually used. But in boosting, the different classifiers are trained sequentially. Each new classifier is trained based on the performance of those already trained sets. Boosting makes new classifiers focus on data that was previously misclassified by previous classifiers. Boosting is different from bagging because the output is calculated from a weighted sum of all classifiers. The weights aren't equal as in bagging but are based on how successful the classifier was in the previous iteration.

D. Implications of Data Analysis

The results drawn from data analysis performed in Hadoop can be used to determine attributes related to traffic control on a road.

1. Computation of average velocity of vehicles on a road

The efficiency of traffic on a road can be determined using the average speed of distinct types of vehicles travelling along that path. Average velocity is the velocity of each vehicle is measured over a region

and not at a specific point on a road, as the later would represent the average velocity at that point instead of the speed along the entire road. The average velocity can be computed by using the following equation:

$$\bar{v} = \frac{n \times s}{\sum_{i=1}^{n} (t_{end} - t_{start})}$$

where

s → the distance between the two points on the road along which the average velocity is computed.
$\mathbf{t_{start}}$ → time at which the vehicle has entered the region of road.
$\mathbf{t_{end}}$ → time at which the vehicle has exited the region of road.
n → the number of vehicles passing through the region of road at that time interval.

One aspect which must be focused upon is that the vehicles which enter midway into the region of road must be discarded from consideration, as it would lead to an inaccurate computation of the average velocity.

2. Estimation of the overall traffic flow

The proposed system enables the calculation of the traffic flow in a particular region for a specific period of time. The Hadoop MapReduce presents a more efficient way to statistically analyze this attribute. The map() function has the vehicle_ID and location as the key values. The time interval for which the vehicle is travelling in a specific region is used as value in reduce() function.

The output value of the <key, value> pair of the map() function is <key, one>. The key includes the vehicleID, location and time interval and the value in the pair is one. The net traffic flow in a direction between two time intervals can be computed using the reduce() function. The output <key, value> pair for the reduce() function is <vehicleID_location_time, count>.

3. Determining the travel path of a vehicle

The use Hadoop MapReduce model automates the process of determining the travel path of vehicle, which had to be done manually currently. Travel path of a vehicle plays a key role in public security investigation for determining ant incredulous activities done in a period of time. The proposed system can determine this efficiently by committing the vehicleID into the database when querying the travel path based on

V. CONCLUSION

This paper discusses the incomprehensible growth of urban population around the globe and its effect on traffic control in big cities. The shortcomings of the traditional traffic management system are presented along with the smart traffic control mechanisms employed in a few cities. An efficient system is proposed with the implementation by big data analytics and RFID, supported by IoT. The architecture and functionalities of the proposed system are described along with supervised learning used to determine the attributes of traffic management. This advanced system, once implemented would drastically reduce traffic congestion in big cities and improve the security of vehicles.

Acknowledgements

We would like to express our heartfelt gratitude to Dr. Ilango Krishnamurthi, Dean, Computing Sciences, Sri Krishna College of Engineering and Technology, who provided insight and expertise that greatly assisted us in the research work. We are also grateful to the faculty members of the Department of Computer Science and Engineering for their valuable suggestions and comments that greatly improved the manuscript.

REFERENCES

[1] UN News, "Urban and rural areas 2015", New York - USA. Available in:<http://www.un.org/en/development/desa/population/publications/pdf/urbanization/urbanization-wallchart.pdf>.

[2] The Fraunhofer FOKUS Institute, "Urban life: Open-air computers", The Economist October. 2012. Available in: http://www.economist.com/news/special-report/21564998-cities-are-turning-vast-data-factories-open-air-computers>.

[3] Forbes List of most traffic affected cities in the world, 2015. Available in: < http://www.forbes.com/pictures/ehmk45kigm/1-brussels-belgium-2/>