A New Approach of Big Data collaboration for Road Traffic Networks considering Path Loss Analysis in context of Bangladesh

Sayed Mohsin Reza, *Member, IEEE,* Md. Mahfujur Rahman, *Member, IEEE,*
M Manjur Mahmud, and Shamim Al Mamun, *Member, IEEE*

***Abstract*—Large amount of money is lost to manipulate traffic system or road congestion worldwide every year. By providing the travel time for a specific road, traffic congestion can be minimized. Many approaches has been taken such Automatic Vehicle Identification, Loop Detectors etc. But this methods are costly. In this paper, A system is proposed that provides traffic intensity level information to the user according to recent traffic data analysis. To minimize the traffic congestion, the paper proposal is to use big data concept in this arena. This proposal develops a structure of a simple XML device which is installed on a vehicle to trace and provide information of traffic intensity. It can estimate travel times in a road network accurately. Further, according to this system anyone can develop application for business process development and increase information transfer to the local user.**

***Keywords*—Traffic, Big Data, Congestion, Intensity, XML device, Road Networks, travel times.**

I. Introduction

EVERY day, time is wasted in traffic congestion. This problem is suffered by millions of people. This traffic congestion wastes millions of money of every government. Every country government propose budget every year to solve this type of traffic mobbing. Escalation of traffic volume, every government is on dilemma. On the basis of this need, new dynamic travel time system with traffic intensity information should be developed k[1][2].

Big data is castoff to increase many aspects of municipalities and countries. For example, it allows municipalities to adjust road traffic movements based on real time traffic flow information. Several cities are currently conducting big data analytics with the aim of turning themselves into Keen Cities, where the transportation infrastructure and utility routes are all combined.

Further, Analysis of Big Data refers as road traffic data turn into valuable resources. It can increase efficiencies of country's competitive advantage. For this, Data Science will help us to produce valuable information from raw data [3]. Big Data refers that -Traditional enterprise data, Machine generated or sensor data, Social data etc.

Sayed Mohsin Reza, Md. Mahfujur Rahman and Shamim Al Mamun are with Institute of Information Technology(IIT), Jahangirnagar University, Savar, Dhaka, Bangladesh.

M Manjur Mahmud is with Datasoft Systems Bangladesh Limited.

Manuscript received May 1, 2014.

In this proposal, machine generated data will be included. The system manipulates these machine-generated data and use the observations per road segment and estimates a travel time based on the average speed.[2]

II. Utilization of Big Data in Society

Analyzed and distilled big data with traditional enterprise data can enhance productivity, competitive stronger position and can innovate much better significant impact on society.

At first, social media sites wouldnt be present without big data. By capturing available data of a user or a member of social media, business developers can model their business and modified experience on the web. Additionally, in healthcare facilities, long term in-home monitoring device to measure progress and vital signs can be useful to ensure patient health. This type of service assistances the patient to reduce appointment to the doctor and hospital admittance. [1]

III. Big Data Platform

Now a day, IT infrastructure is developing in such a way that produces unique requirements considering big data platforms components. To develop the paper proposal, the system should go through some big data requirements [4].

- Big data acquisition
- Big data organizing process
- Big data Analysis
- Solution prediction

IV. Road Traffic Intensity Parameter using Big Data

There are several surveys happens to find out the parameter in the concept of Big Data. In simple way, we can assume that the volume of data (Road traffic data) is one of the parameters. In fact, it is not only the parameters that controls Big Data terms. There are Volume, Velocity, Variety and Value.

Each year, data size is mounting 40 percent per year as saying by McKinsey Global Institute researchers [2]. As their say, larger quantities of data are produced by Machine Generated data. In Big data terms, Traditional data and non-traditional data formats are on contrast position.

Among these concepts, we face the challenge that which data is valuable and that valuable data should be used for analysis road traffic summarize data to ensure road traffic intensity and can potentially help to improve the traffic situation in large cities.

V. Road Traffic Data [Big Data] Infrastructure

The continual growth of road traffic increases the need for solutions allowing monitoring and controlling traffic helping to increase road safety, improve traffic flow, and to protect the environment. So, a system which provides traffic management solutions allowing road authorities and operators to manage, monitor and maintain their roadways while giving the road users intelligent information to ensure convenient and save trip [5].

For ensure a Road Traffic Network system, we should follow some proper ways to monitor, control traffic. These are -

A. Data Acquisition

Data acquisition is the method of evaluating an electrical or somatic phenomenon such as current, voltage, pressure, temperature and sound with a computer. A Data Acquisition system consists of devices, hardware, and a computer with programmable software. Compared to traditional measurement systems, PC-based Data Acquisition systems abuse the processing power control, productivity, demonstration, and connectivity capabilities of industry-standard computers providing a more authoritative, stretchy, and cost-effective measurement clarification [6].

B. Data Organization

All data from the devices as well as manual actions of the operating staff are considered. The Traffic Management Systems enables the automated generation of traffic information. Raw traffic data and environmental data are processed to traffic information so that customers can access it via a broad spectrum of telecommunication technologies such as web services [7].

C. Data Analysis

Assembling, evaluating and processing the entire data from various devices on the roads in order to understand and monitor in real-time the prevailing traffic conditions. This analysis is the last step to provide correct information to the user.

VI. System Description

A. Introduction

In traffic system, Government of every country expends their money for organizing the traffic system. Billions of dollars are lost. People waste their times for congestion in traffic. People want to save their time but is there any way to save time? To ensure traffic intensity, we can survey road traffic data. But how the road traffic data is collected? Or in which way we will provide the users the traffic information?

Is Big Data concept applies here to ensure any possibility of road traffic Data? Data acquisition is a now challenge for us. In past, there were proposed some ways to provide this type of service.

To acquisition of road traffic data, we have to collect data using point based approach. For a short time, a control server will summarize the accumulating data to traffic organizers data.

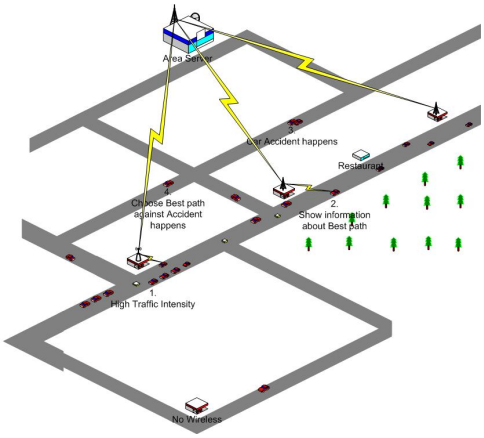

Fig. 1. Entire Scenario of the System

This data will be promoted to the other area users but not these area users [8].

For time consume in this method, how much bandwidth and transfer rate will be needed? To solve this, we have to use less data information but not whole. We can use XML format to show information to users device.

B. Algorithm to Develop this System

XML vehicle device to server data handover and server to that XML device information transfer is a continuous process to develop this system. For this, we develop an algorithm which helps us to ensure the process of this system.1

```
Data: Data Acquisition from vehicle XML device
      (simple black box) send to the Server.
Result: summarize the accumulating data to traffic
        organizers data.
Process the data known as Road Traffic Data (Big Data)
to take proper decision
while device status do
    read current vehicle information;
    if traffic intensity level is high then
        appearance possible way(if exist) to go to
        destination, time and appearance local
        information.;
        if path changes then
            go to the beginning of section.;
        else
            check black box provide service or not?;
            break;
        end
    else
        one way decision to go to destination, time and
        appearance local information;
    end
end
```

Algorithm 1: Algorithm for develop the system

C. Work flow

Road Traffic congestion system should follow a work flow diagram. As far as system data acquisition, data organization and at last data analysis is shown in this work flow diagram.

Here, a server as Google file system or Hadoop Distributed File System will work. It acquires data from Vehicle device that transmit traffic information as XML format. System Server further processes this information (Big Data or Road Traffic Data). These processing raw data will produce valuable resources for sending information to vehicle which is under around nodes.

These information contains the traffic intensity level or road accident information that helps the around user vehicle to take proper decision about his/her vehicle.[1]

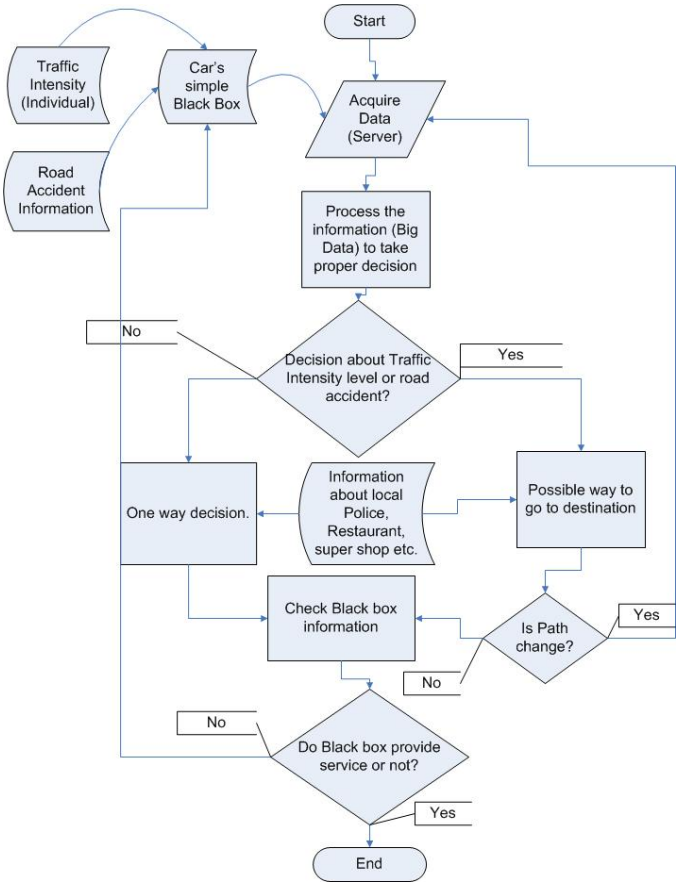

Fig. 2. Data Flow Diagram

If any vehicle changes its path then next node/station will acquire its information and send it to the server. This continuous process will produce Big data refers as Road Traffic Data.

At last, every data sync time, we send the information about device that is the device services working or not? If not then the server take an automated decision, else if it continues its continuous process.

D. Data Schema

To estimate travel times, different data source (users information) are relied on. Acquisition of data should be manipulated by a Data schema developed in the Control server. From that data, users are provided various types of information like longitude, latitude, timestamp, possible ways etc.

Here, we proposed a data schema 3 installed on the control server. This data schema will be helpful to the developer to collect several types of information which is related to the users information.

To be built a data schema, the main table data acquisition, contains data such as drivers information, connections information, source information etc. The abbreviations FK and PK are Foreign key and Primary Key respectively. By querying the main table, developer can summarize the area information which is under of Control Server.

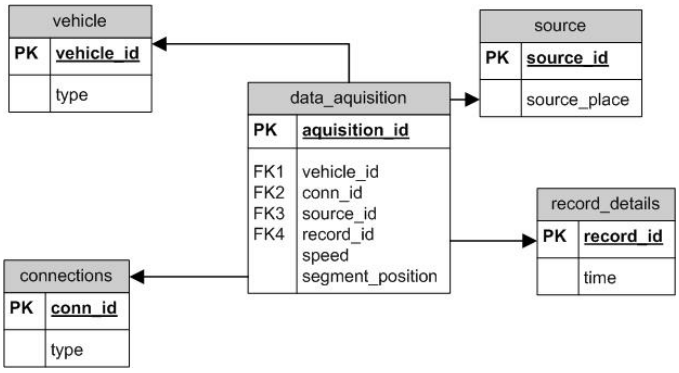

Fig. 3. Data Schema

VII. Feasibility Study

To build up the system we have to go through a feasibility study. In Active Traffic Management, there are some criteria to go through the feasibility study to provide estimated travel time and condition reports to communicate travel and traffic conditions. These are Speed Harmonization, Queue Warning, Junction Control, Hard Shoulder Running and Dynamic Rerouting.

In the system, such like that we determine some criteria related to the feasibility study.

- Data Rate Analysis
- Path Loss Manipulation
- Power Optimization
- Output Scenario

A. Data Rate Analysis

The speed with which data can be transmitted from one device to another is known as data rate. Data rates are often measured in megabits (million bits) or megabytes (million bytes) per second. These are usually abbreviated as Mbps and MBps, respectively. The system should be proposed under some feasibility study. Here, some terms are needed to sign in. Data Transfer rate (DTR) is much more important to time

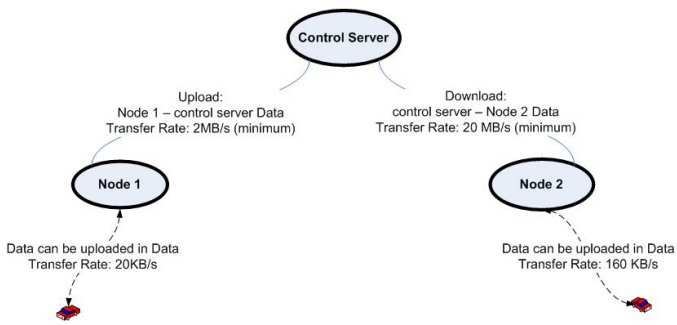

Fig. 4. Data Transfer Rate

consumption in this system.

Calculation:

Normally, a vehicle goes through Node 1 to Node 2 and the distance is d km. the vehicle takes time t s. Distance (Node 1 to Node 2) = d km At last, we can give this type of conclusion.

$$DT1 + DC1 + DP + DC2 + DT2 < t \tag{1}$$

where,

Device to node1 Data Transfer Rate = DT1
Device to node1 Data Transfer Rate = DT1
Node 1 to Control Server Data Transfer Rate = DC1
Control Server Data Process Time = DP
Control Server to node 1 Data Transfer Rate = DC2
Node 1 to device Data Transfer Rate = DT2

If it is possible then the system is feasible to development. Because if transferring data is longer before car reaching to the next station/node, then the system doesnt provide information to the users vehicle device. In the context of IEEE 802.11 technology, minimum net bit rates are 1 or 2 Mbit/s. It has some advantage like forward error correction code [9].

TABLE I. IEEE 802.11 TECHNOLOGY AND RATE

Technology	Rate
diffuse infrared operating	1 Mbit/s
frequency-hopping spread spectrum operating	1 Mbit/s or 2 Mbit/s
1 Mbit/s or 2 Mbit/s	1 Mbit/s or 2 Mbit/s

A vehicle device produce data in XML format size consists of Less than 10KB. So, we use IEEE 8-2.11 technology for simple wireless data transfer manipulation. Minimum wireless Data transfer rate - I

DT1 = 250 KB/s (25 car per second)
DT2 = 250 KB/s (25 car per second)

We proposed a High speed broadband connection between control server and Nodes. Minimum Data transfer between Nodes to control server:

DC1 = 2 MB/s (250 vehicles data)
DC2 = 2 MB/s (250 vehicles data)

Traffic data acquisition by node covers average 1 KM. So between two nodes we can easily control data acquisition. In control server, for better data manipulation and data organization we can assume some more time. But by calculating data transfer rate among the node we can say that the proposed system can easily pass the time consumption process.

B. Path Loss manipulation

Okumuras field test results construct Hata Model and projected various equations for Path loss with different types of confusion. It is well suited model for the Ultra High Frequency (UHF) band. The restrictions on Hata Model owing to range of test results from carrier frequency 150 MHz to 1500 MHz, the remoteness from the base station ranges from 1 Km to 20 Km, the height of base station antenna (hb) ranges from 30 m to 200 m and the height of mobile antenna (hm) ranges from 1 m to 10 m.[10]

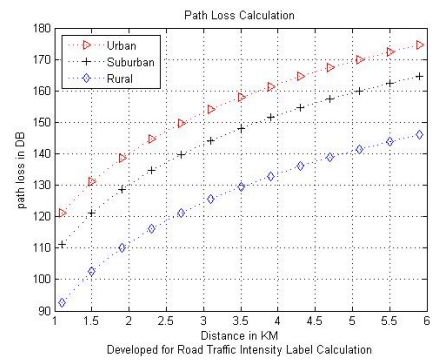

Fig. 5. Path loss Calculation using Hata Model

Observation showed that the signal strength is a function of distance and antenna height, as we can see in this work the uppermost antenna has less propagation path loss and as the distance upsurges the path loss also increases. It was created a number of representative Path loss mathematical models for each of the urban, suburban and open place environments, as illustrated in following equations, respectively. It is taken urban areas as a reference and applies correction factors as following:[11]

a) Urban areas: :

$$L_{dB} = A + B\log_{10} R - E_{1,2.3} \tag{2}$$

b) Suburban areas: :

$$L_{dB} = A + B\log_{10} R - C \tag{3}$$

c) Open areas: :

$$L_{dB} = A + B\log_{10} R - D \tag{4}$$

where,
A = $69.55 + 26.16\log_{10} f_c - 13.82\log_{10} h_b$
$B = 44.9 - 6.55\log_{10} h_b$
$C = 2(\log_{10}(f_c/28))^2 + 5.4$
$D = 4.78(\log_{10} f_c)^2 + 18.33log10f_c + 40.94$

for large cities, when $f_c \geq 300MHz$

$$E_1 = 3.2(\log_{10}(11.7554h_m))^2 - 4.97$$

for large cities, when $f_c < 300MHz$

$$E_2 = 8.29(\log_{10}(1.54h_m))^2 - 1.1$$

for medium to small cities.

$$E_3 = (1.1\log_{10} fc - 0.7)h_m - (1.56\log_{10} f_c - 0.8)$$

Here,
h_b : base station antenna height above local terrain height[m]
$f_c = f \times 10^{-6}$ carrier frequency [MHz]
h_m : mobile station antenna height[m]

1) Path Loss Effect: In this System, all vehicles act as clients while the Web Server is performances as the receiver. Vehicles produce positions under the area server and distance between vehicles and Access Points is calculated from distance equation. This project consists of a path length of 1000m with a communication range of 300m. The distance calculation is done between vehicles and Access Point by using the distance equation between two points, so that number of vehicles in communication range can be determined[12]. Other values of this system are shown in Table II.

TABLE II. PRIMARY SAMPLE DATA

Parameters (Units)	Values	Type
Road Length (m)	1000	Not Fixed
Number of vehilces	150	Not Fixed
Position of Access Point (m)	500	Not Fixed
MAC Protocol	802.11	Fixed
Communication Range (m)	300	Not Fixed
Maximum Retransmission	3	Not Fixed

In this project, simulation has been done using MATLAB. From figures 6, it demonstrations how path loss exponent marks the network act and ability. As shown in Table III, in free space, n is equal to 2 because of vehicle position in an open space to access point and in the existence of obstacle, the value of n will be higher. The value of n can be governed by on the environment. Simulation for values of n between 2 and 4 to assess the influence of transmission procedures on different situation[11]. The values castoff for this simulation matches to the different situations listed in Table III.

TABLE III. PATH LOSS EXPONENT

Situation	Path Loss Exponent (n)	Cause
Free Space	2	Vehicles in the position of very near and open space to access point
Building Obstacle	1.6 - 1.8	Vehicles position in the city area
Cars Obstacle	3-4	Vehicles Position in the Traffic area

Path loss is the deprivation in received power of an Electromagnetic signal when it broadcasts through space. Path loss is owing to numerous things such as free space path loss, bending, deflection, reflection, coupling and engagement. Path loss is be governed by on some aspects such as type of environments, space between transmitter and receiver, height and location of antennas. Also the signal from the communicating antenna may take multiple paths (multipath) to touch the acceptance side, which effects in either upturn or reduction of received signal level depending on the beneficial or damaging intervention of the multipath waves.[13]

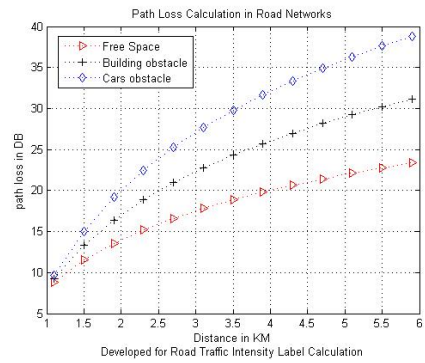

Fig. 6. Path loss for Road Network

C. Power Optimization

Wireless power transmission should be established on inductive power. Now a days wireless data communication is become popular and expectation is now on rise. So power distribution or measurement is considerable in sense.

Wireless power technology in consumer devices benefit greatly from the availability of a standard that is widely acceptable. To establish interoperability between wireless power chargers and power devices, we have to issue a Wireless power consortium [14].

The standard defines a power transmission of 5W. 110 kHz is Nominal operating frequency. Power control can be initialized by controlling amplitude or frequency. For better power control we have to calculate differences of output value of power transfer. Output may be voltage or current or power.

D. Output Scenario

In the proposed system, at first the XML device which is installed on that vehicle detects the local base station and show the base station name.

VIII. IMPLEMENTATION OF BIG DATA FOR DATA MANIPULATION

One of the challenges: data acquisitions are complete in arbitrary sense. The data which is collected from the user [vehicles] should be manipulated. After manipulation, summarize data will be transferred to the user output devices in XML format.

XML format is much easier to transfer into a device or we know XML was designed to transport and carry data. XML tags are not predefined. Self-descriptive design is the main theme of XML format.

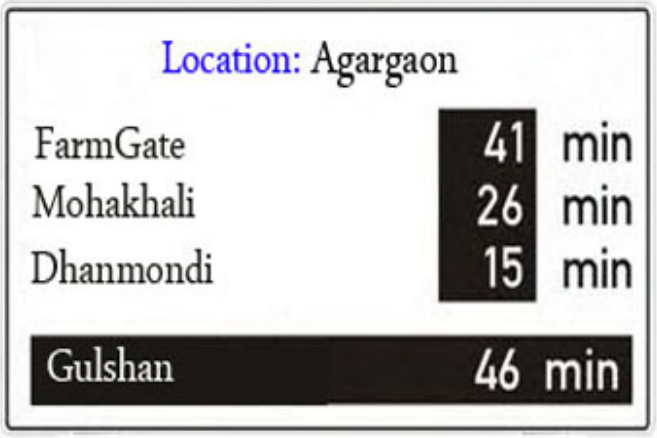

Fig. 7. Probable Output of XML Device in Vehicle

A. Parsing XML into Data Structures

This data structure (a mixture of hashes and arrays) that relates to the arrangement and content of an XML file. For example, XML device on behalf of a configuration file. XML

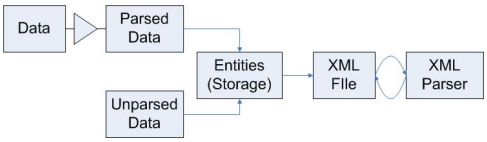

Fig. 8. Process of Parsing XML

files are created for storage elements called entities, which contain either parsed or unparsed data. Parsed data contains data.XML delivers a appliance to execute restrictions on the storage outline and logical structure.

A XML device module called an XML processor is hand-me-down to read XML files and afford access to their content and structure. It is expected that an XML processor is doing its effort in support of another module, called the server application.

IX. Tools for Data Manipulation

The proposed system produces huge data as independent computational agent like Machine generated data. It unknowingly generates the data. So we have to manipulate it to give proper credit like valuable information. To do that, we will use some of these tools for my proposed system [15].

A. Google File System [Hadoop Distributed File System]

Since the proposed system produce large data sets and we have to choose such a system that can store, stream that large data sets at higher bandwidth then a tool named Hadoop Distributed file system will be the proper answer. Hadoop provides a distributed file system to analyze and transformation of very large data sets. Its framework works as like as to develop computation among large amount of hosts [16].

In the proposed system, large data set will produce to collect traffic data which should be store then analyze that data and transform into information. In this case, large computation may be needed.

To increase the performance of application, Hadoop Distributed file system is perfect because it is patterned after UNIX file system. This file system improves the execution of program computations in parallel. These file system stores its metadata on its dedicated server and program data are stored on Data Nodes. For better reliability, the content data are located in multiple data Nodes. Every server is connected with each other using TCP based protocols [17].

In HDFS, user can access a library that exports HDFS file system interface. This system supports various operations like read, write and other operation on directory. Here, the data nodes are on different location, but user can run their application without knowing that. [16]

Traffic data acquisition and then manipulation concept also contains different location so Hadoop distributed file system will work here properly as the paper proposal. Moreover, HDFS contains some more advantage like Backup Node. It is capable of creating periodic checkpoints. This node can be viewed as Name Node (read only). [17]

To ensure security about traffic data, HDFS provides backup node for the valuable data. The data may be in read only format so its security establish as more proper format.

X. Conclusion

In this paper, proposal is developed in such a way so that server system can determine travel time approximation, this system estimate travel times for a road traffic network using a simple device which can send data and retrieve data from server. To collect traffic data, the road network, and the processing data that permits us to filter out exact Road Traffic data, a data schema has been proposed for flexible efficiencies to provide accommodations, upcoming add-ons to the travel-time system as well as vehicle and traffic facts, road circumstances.

It has been conversed how the input scenario and output scenario can be joined in preparation to analyze travel times and make available one possible approach. As a side benefit, this methodology can be used to inspect the frequencies of Road Traffic intensity level. This is used by traffic experts when analyzing road traffic data.

The principal of this system is now in a state where we can initiate to guess travel times. The following is a little list of the future work we would like to look into.

- Mobile Android Apps
- Vehicle Trace Application
- Built A Web Application to Show An Entire City Road Traffic Information

Acknowledgment

This research was supported by Institute of Information Technology, Jahangirnagar University (IIT-JU) and Datasoft Systems Bangladesh Limited.

REFERENCES

[1] Sayed Mohsin Reza, Md. Mahfujur Rahman and Shamim Al Mamun, A New Approach for Road Networks - A Vehicle XML Device collaboration with Big Data. *1st International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, April 2014.

[2] Lakas, A., Detection and dissipation of road traffic congestion using vehicular communication. *Microwave Symposium (MMS)*, Mediterrannean, 2009.

[3] Bernard Marr, The Awesome Ways Big Data Is Used Today To Change Our World. *http://www.linkedin.com/today/post/article/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world*, November 13, 2013.

[4] Jean-Pierre Dijcks, Oracle: Big Data for the Enterprise, June 2013.

[5] Anders Forum Jensen and Troels Villy Larsen, Travel-Time Estimation in Road Networks Using GPS Data. *http://projekter.aau.dk/projekter/en/studentthesis/traveltime-estimation-in-road-networks-using-gps-data%2847f7e0de-e8b5-4120-b113-e21375530d96%29.html*, June 2007.

[6] Data Acquisition. *http://www.ni.com/data-acquisition/what-is/*, December 2013.

[7] Feasibility of Collecting Traffic Safety Data from Law Enforcement Agencies, *U.S. Department of Transportation, National Highway Traffic Safety Administration*, April 2011.

[8] Bauza, R., Gozalvez, J. and Sanchez-Soriano, J., Road traffic congestion detection through cooperative Vehicle-to-Vehicle communications. *IEEE 35th Conference on Local Computer Networks (LCN)*, 2010.

[9] IEEE 802.11, Wikipedia, the free encyclopedia. *http://en.wikipedia.org/wiki/IEEE_802.11*

[10] Zia Nadir and Mohammed Bait-Suwailam, Pathloss Analysis at 900 MHz for Outdoor Environment. *International Conference on Communications on Signal Processing and Computers*, 2014.

[11] S. P. Sotiroudis, K. Siakavara, and J. N. Sahalos, A Neural Network Approach to the Prediction of the Propagation Path-loss for Mobile Communications Systems in Urban Environments. *PIERS Proceedings*, Czech Republic, 2007.

[12] Raul A. Gorcitz, Yesid Jarma, Promthe Spathis, Marcelo Dias de Amorim, Ryuji Wakikawa, John Whitbeck, Vania Conan, and Serge Fdida, Vehicular Carriers for Big Data Transfers. *IEEE Vehicular Networking Conference (VNC)*, Korea, 2012 .

[13] David Gundlegrd and Johan M Karlsson, Generating Road Traffic Information from Cellular Networks - New Possibilities in UMTS. *6th International Conference on ITS Telecommunications*, 2006.

[14] Bandwidth (signal processing), Wikipedia, the free encyclopedia. *http://en.wikipedia.org/wiki/Bandwidth_%28signal_processing%29*

[15] Margaret Rouse, Data Science. *http://searccio.techtarget.com/definition/data-science*, September 2012.

[16] Robert Chansler, Hairong Kuang, Sanjay Radia, Konstantin Shvachko and Suresh Srinivas, The Hadoop Distributed File System. *http://www.aosabook.org/en/hdfs.html*, Ch. 9, pp.121-134.

[17] Hadoop Distributed File System (HDFS) Architectural Documentation Use Cases. *http://itm-vm.shidler.hawaii.edu/HDFS/ArchDocUseCases.html*

[18] Statistical Data Analysis. *http://reference.wolfram.com/language/*

[19] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, The WEKA Data Mining Software. *ACM SIGKDD Explorations Newsletter*, Volume 11 Issue 1, June 2009.

Sayed Mohsin Reza is continuing his B.Sc.(Hons.) degree in Information Technology (IT) from Jahangirnagar University, Savar, Dhaka, Bangladesh. In his 3 years of practical experience in the Information Technology (IT) field, he completed more than 10 successful project. His research area covered Big Data, Database, Software Engineering etc. He is a Cisco Certified Network Associate(CCNA) and Certified Database Designer and implementation with Web Applications. He is a Graduate Student member of IEEE.

Md. Mahfujur Rahman is continuing his B.Sc.(Hons.) degree in Information Technology (IT) from Jahangirnagar University, Savar, Dhaka, Bangladesh. In his 3 years of practical experience in the Information Technology (IT) field, he completed more than 10 successful project. His research area covered Big Data, Database, Software Engineering etc. He is a Cisco Certified Network Associate(CCNA) and Certified Database Designer and implementation with Web Applications. He is a Graduate Student member of IEEE.

M Manjur Mahmud with his Computer Sc major brings in more than 24 years of practical experience in the Information Technology (IT) field. In years Mr. Mahmud has been successful rolling out more than 100+ projects in various fields. His main area of specialization is in Operational Excellence by managing Enterprise level project in PMI framework and also drives the Quality Assurance (QA) initiatives for the company. Mr .Mahmud is one of the early Oracle Certified Professional in the country , a Project Management Professional (PMP-USA) and Information System Audit Professional (CISA-USA). He is also a trained Technopark Professional from South Korea . He is an alumnus of Indian Institute of Management (IIM), Calcutta and University of Hong Kong (HKU) .

Shamim Al Mamun has been serving as an Assistant professor in Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka, Bangladesh. He received his M.Sc.(Eng) in IICT , B.SC in CSE from Bangladesh University of Engineering and Technology (BUET) and Jahangirnagar University respectively. His research interest includes Optical Wireless Communication and Big Data. He is the Member of IEEE Computer Society.