A New Approach for Road Networks - A Vehicle XML Device collaboration with Big Data

Sayed Mohsin Reza
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: smrezaiit@gmail.com

Md. Mahfujur Rahman
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: mrrajuiit@gmail.com

Shamim Al Mamun
Institute of Information Technology
Jahangirnagar University
Savar, Dhaka.
Email: sam@juniv.edu

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.

Index Terms—Traffic, Big Data, Congestion, Intensity, XML device, Road Networks, travel times etc.

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 [1].

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 [2]. Big Data refers that -Traditional enterprise data, Machine generated or sensor data, Social data etc.

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.

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.

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 [3].

- 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 [1]. 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 [4].

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 [5].

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 [6].

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

Fig. 1. Entire Scenario of the System

will summarize the accumulating data to traffic organizers data. This data will be promoted to the other area users but not these area users [7].

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.
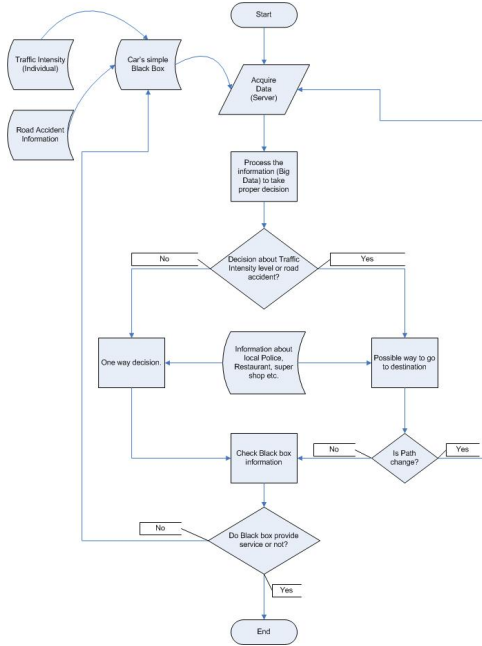

Fig. 2. Data Flow Diagram

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.
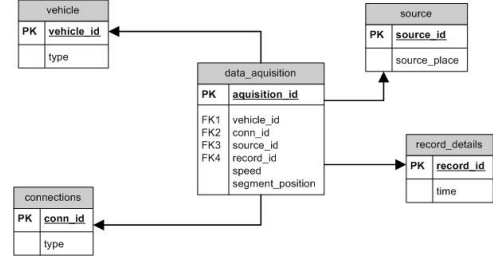

Fig. 3. Data Schema

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
- 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
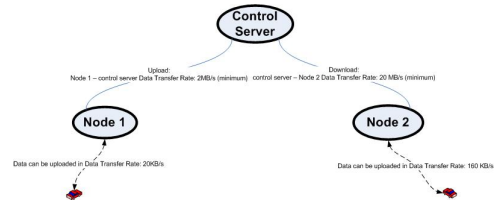

Fig. 4. Data Transfer Rate

some feasibility study. Here, some terms are needed to sign in. Data Transfer rate (DTR) is much more important to time 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.

$$DT1 + DC1 + DP + DC2 + DT2 < t \qquad (1)$$

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 [8].

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. 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 [9].

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.

C. 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.
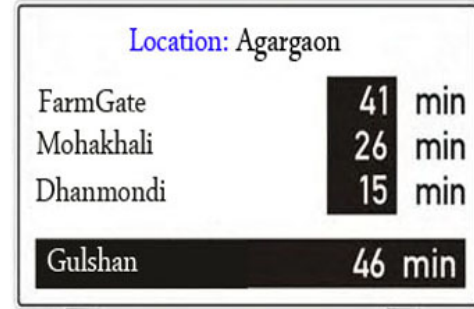

Fig. 5. Probable Output of XML Device in Vehicle

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.

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
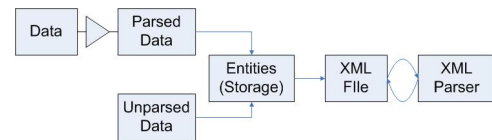

Fig. 6. Process of Parsing XML

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 [10].

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 [11].

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 [12].

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. [11]

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). [12]

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

References

[1] Detection and dissipation of road traffic congestion using vehicular communication-http://ieeexplore.ieee.org/xpl/freeabs_all.jsp? arnumber=409762&abstractAccess=no&userType=inst
[2] The Awesome Ways Big Data Is Used Today To Change Our World
[3] An Oracle White Paper, June 2013: Oracle: Big Data for the Enterprise
[4] Travel-Time Estimation in Road Networks Using GPS Data, Anders Forum Jensen and Troels Villy Larsen.
[5] Data Acquisition - http://www.ni.com/data-acquisition/what-is/
[6] Feasibility of Collecting Traffic Safety Data from Law Enforcement Agencies, U.S. Department of Transportation, National Highway Traffic Safety Administration, April 2011.
[7] Road traffic congestion detection through co-operative Vehicle-to-Vehicle communications-http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5735780&abstractAccess=no&userType=inst
[8] IEEE 802.11 - http://en.wikipedia.org/wiki/IEEE_802.11
[9] Bandwidth (signal processing)- http://en.wikipedia.org/wiki/Bandwidth_%28signal_processing%29
[10] Data Science - http://searchcio.techtarget.com/definition/data-science
[11] The Hadoop Distributed File System by Robert Chansler, Hairong Kuang, Sanjay Radia, Konstantin Shvachko, and Suresh Srinivas
[12] Hadoop Distributed File System (HDFS) Architectural Documentation Use Cases - http://itm-vm.shidler.hawaii.edu/HDFS/ArchDocUseCases.html
[13] Statistical Data Analysis - https://reference.wolfram.com/mathematica/guide/Statistics.html
[14] The WEKA Data Mining Software: An Update by Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten