Study of Hadoop-based Traffic Management System

Madhavi Vaidya,
Asst. Professor
Dept of Computer Sc., Vivekanand College,
Mumbai, India

Shriniwas Deshpande, Ph.D.
Associate Professor, Head of PG Dept of
Computer Science & Technology, DCPE,
HVPM, Amravati, India

ABSTRACT
Traffic congestion on road networks is characterized by slower speeds, longer trip times and increased vehicular queuing. We all are observing that vehicle volume is increasing day by day exponentially, but in comparison with it, the road infrastructure is not. It leads to ever increasing traffic congestion. Different technologies are used to detect traffic congestion and making congestion management more efficient. For that, lots of techniques have been found out. In this paper, the author has studied the Radio Frequency Identification technique (RFID) which is used for controlling traffic congestion. The data created from it has to be processed and then used by the different traffic control systems on the road. Processing of the data may create a problem due to large data volumes. Hence author has suggested to use the Hadoop Architecture to address this problem. By using Map Reduce framework, the data can be available at many sites as Map Reduce architecture uses parallel processing paradigm at its core.

The goal is to implement a system that would trace the travel time of individual cars as they pass the roadside readers, create an average trip time and then making that information available to the different toll centers. For this, the author has suggested using Hadoop architecture and by using Map Reduce Framework the data can be processed at different centers in a distributed manner.

General Terms
Processing, tag, transmitter, security

Keywords
RFID, Hadoop, Map, Reduce, Traffic

1. INTRODUCTION
RFID is in use all around us. RFID is a generic term that is used to describe a system that transmits the identity (in the form of a unique serial number) of an object or person wirelessly, using radio waves. It is grouped under the broad category of automatic identification technologies. [1]

This paper is structured as follows: in section II, an introduction to Radio Frequency Identification is given. In section III, the Use OF RFID for Traffic Control System is discussed. In section IV, the plan of Indian Ministry has been stated for using RFID technique to detect the vehicle moving on the road. Section V describes the HDFS architecture. Section VI suggests how Hadoop Architecture with the help of Map Reduce Framework can be useful to process the data which is generated from RFID for tracking vehicles. Finally conclusions are drawn and future work is outlined.

2. INTRODUCTION TO RFID
A radio-frequency identification system uses tags, or labels attached to the objects to be identified. Two-way radio transmitter-receivers called interrogators or readers send a signal to the tag and read its response. The readers generally transmit their observations to a computer system running RFID software or RFID middleware.

The tag's information is stored electronically in a non-volatile memory. The RFID tag includes a small RF transmitter and receiver. An RFID reader transmits an encoded radio signal to interrogate the tag. The tag receives the message and responds with its identification information. This may be only a unique tag serial number, or may be product-related information such as a stock number, lot or batch number, production date, or other specific information.

We have read that the RFID tag is connected to the 10% of cattle in India, [2] which is the largest country to have number of cows and bullocks in the whole world. In addition, RFID is increasingly used with biometric technologies for security. RFID devices into two classes: active and passive. Active tags require a power source—they're either connected to a powered infrastructure or use energy stored in an integrated battery. In the latter case, a tag's lifetime is limited by the stored energy, balanced against the number of read operations the device must undergo [3].

2.1 Components of RFID
A basic RFID system consists of three components:

- An antenna or coil
- A transceiver (with decoder)
- The antenna emits radio signals to activate the tag and to read and write data to it.
- A transponder (RF tag) electronically programmed with unique information

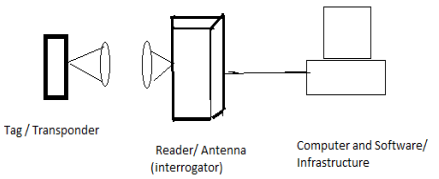

Fig 1: RFID System

- The reader emits radio waves in ranges of anywhere from one inch to 100 feet or more, depending upon its power output and the radio frequency used. When an RFID tag passes through the

electromagnetic zone, it detects the reader's activation signal.

- The reader decodes the data encoded in the tag's integrated circuit (silicon chip) and the data is passed to the host computer for processing.

2.2 Working of RFID

The RFID tag can be affixed to an object and used to track and manage inventory, assets, people, etc. in the case of this paper this can be affixed to cars. The tag can be read if passed near a reader, even if it is covered by the object or not visible. The tag can be read inside a case, carton, box or other container, and unlike barcodes, RFID tags can be read hundreds at a time.

The purpose of an RFID system is to enable data to be transmitted by a portable device, called a tag, which is read by an RFID reader and processed according to the needs of a particular application. The data transmitted by the tag may provide identification or location information, or specifics about the number of the car, etc.

RFID quickly gained attention because of its ability to track moving objects. A typical RFID tag consists of a microchip attached to a radio antenna mounted on a substrate. The chip can store as much as 2 kilobytes of data. To retrieve the data stored on an RFID tag, a reader is needed. A typical reader is a device that has one or more antennas that emit radio waves and receive signals back from the tag. The reader then passes the information in digital form to a computer system.

As a vehicle passes by a reader, it tracks the vehicle through the RFID tag attached to the vehicle and retrieves its electronic product code (EPC) data. The EPC primarily consists of the vehicular identification number (VIN). The VIN is an industry standard and each automobile has a unique VIN.

3. USE OF RFID FOR TRAFFIC CONTROL SYSTEM

Vehicular traffic control at road crossings has always been a matter of concern for administrations in our city. Efforts have been taken up to design efficient automated systems to solve this problem. Most of the present day systems use predetermined timing circuits to operate traffic signals, which are not very efficient because they do not operate according to the current volume of traffic at the crossing. It is often seen in today's automated traffic control systems that vehicles have to wait at a road crossing even though there is little or no traffic in the other direction. There are other problems as well, such as ambulances getting caught up by a red traffic signal and wasting valuable time. Congestion is often translated into lost time, missed opportunities, lost worker productivity, delivery delay, and a general increased cost.

Traffic congestion is not only recurring but nonrecurring which is caused by traffic incidents like damaged vehicles, crashes, work zones, weather and special events) in nature. To manage nonrecurring congestion, some sensor-based systems were suggested for improvement over fixed timing controlled ones. [4]

To solve these problems of congestion and speed control and to add further enhancements to the complex problem of vehicular traffic control at road crossings, Hadoop Based Traffic Control System is proposed using radio frequency identification (RFID) and store the processed data from the same the Hadoop architecture (Refer Figure 2) is used.

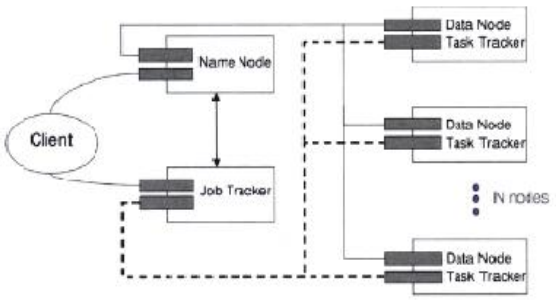

Fig 2 : Hadoop Architecture

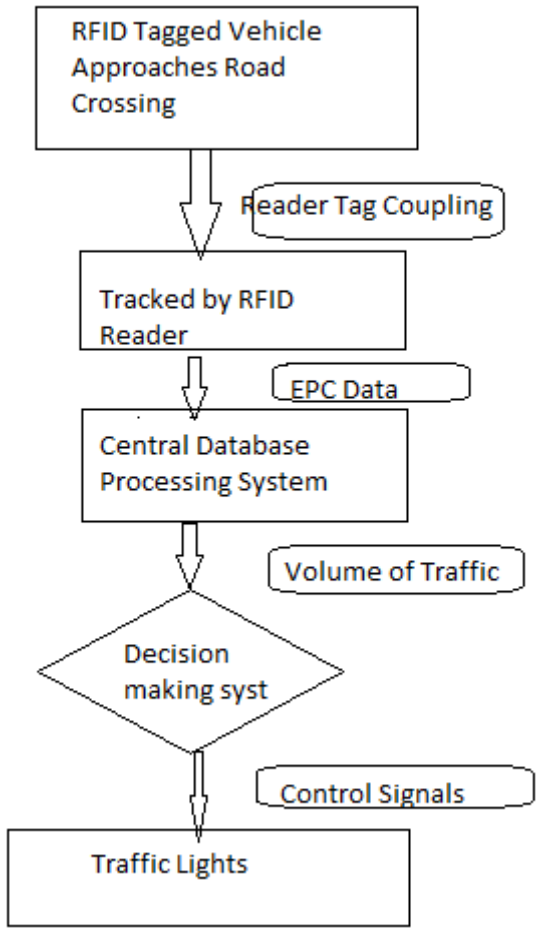

Fig 3 : A flowchart of the system

3.1 Scope of RFID based System for Vehicle Tracking System

Using ActiveRFID and Wireless Networking Technology for Automatic Vehicle Identification with the following objectives:

- Automatic Detection of Congestion of vehicles in Real-Time

- Automatic Detection of Vehicles moving towards congested area and automatic messaging for redirecting the (selected) Vehicles for Congestion Avoidance
- A high vehicle density in a particular lane causes a traffic signal in that particular direction to remain open for larger duration thus adaptively controlling the signal
- Automatic Billing of Core Area Charges / Toll Charges by identifying vehicles within Core Area / Toll Area
- Automatic Billing of Penalty Charges for selected Traffic Rule Violation (speed limit violation, entry into congested area in spite of re-routing messaging) [5]

4. THE PLAN OF INDIAN MINISTRY

The ministry of heavy industries is considering a proposal to make it mandatory to fit RFID-enabled devices in the cars manufactured in India. It is believed that RFID tags would help in traffic management as traffic violations by motorists could be tracked and all violations identified. Also, motorists would get charged automatically as soon as they enter a toll area.

The ministry of urban development has already discussed a similar agenda with many states. The ministry has proposed a core area charge for different cities to reduce traffic congestion in the city's nerve centre.

- Active RFID devices can act as tags or Readers or Routers
- Active tags can even be integrated with sensors to form a standard wireless sensor node

4.1 Detection and Management of traffic congestion

- The system will use roadside RFID readers to collect signals from active RFID attached to the vehicle units already installed in the customer vehicles.
- The RFID antenna will be coupled to the PC and once the vehicle comes across the toll booth centre, its tag is captured and this is passed on to the PC for processing.
- The goal is to implement a system that would trace the travel time of individual cars as they pass the roadside readers, compute an average trip time and then use a rule-based system to decide whether the area is congested.
- If congestion is detected, the system would control traffic signals / generate automatic re-routing messages to selected approaching vehicles (Refer Figure 3)

4.2 Advantages of RFID

- No Direct Line of Sight required for identification & tracking.
- Enables very specific detection of vehicles.
- Simultaneous multiple detection of vehicles are possible using RFID.
- No performance degradation during harsh weather.
- Eliminates manual record keeping, thereby increasing accuracy [6]

4.3 Automatic detection of speed limit violation

- The same technique is used to calculate the speed of a motorist and to detect if he violates the prescribed speed limit.

- On detection of such violation, a warning message is sent to the audio and display interface of the motorist and penalty can be calculated in the server and then billed monthly to the vehicle owner.

5. USE OF HADOOP ARCHITECTURE

Hadoop [7] provides a distributed file system and a framework for the analysis and transformation of very large data sets using the MapReduce [8] paradigm. An important characteristic of Hadoop is the partitioning of data and computation across many (thousands) of hosts, and executing application computations in parallel close to their data.

HDFS, the Hadoop Distributed File System, is a distributed file system designed to hold very large amounts of data (hundreds of terabytes or even petabytes), and provide high-throughput access to this information. Files are stored in a redundant fashion across multiple machines to ensure their durability to failure and high availability to very parallelized applications. Hadoop Map-Reduce is a software framework that process large amount of data, in parallel, in a fault tolerant manner.

It is highly efficient and scalable in nature, and is used to process huge datasets. HDFS has master/slave architecture. An HDFS cluster typically consists of a single Name Node, a master server that manages the file system namespace and regulates access to files by clients, secondary name nodes and at least 3 data nodes.

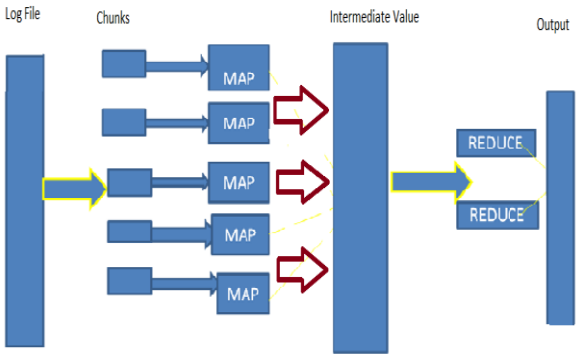

Fig 4 : Map Reduce Framework

HDFS is the cornerstone of the distributed computing, which saves all data files of the Hadoop cluster. Primary Name Node is the file system's administrator, in charge of managing the naming space, storing the file system's descriptive metadata, as well as the users' access to documents. Secondary Name Node is the backup of Name Node. Actual data storage is in each Data Node, which divides the local disk into many blocks to store data. Data Node needs to report to Name Node periodically. Files are also chunked for better throughput.

In addition, there are a number of Data Nodes, usually three nodes in the cluster, which manages storage attached to the nodes that they run on. Map-Reduce programming consists of writing two functions, a map function, and a reduce function. The map function takes a key, value pair and outputs a list of

intermediate values with the key. The map function is written in such a way that multiple map functions can be executed at once, so it's the part of the program that divides up tasks. The reduce function then takes the output of the map functions, and does some process on them, usually combining values, to generate the desired result in an output file. The Map Reduce thus uses a divide-and-conquer approach to parallelize computations closely following the framework and nomenclature proposed by Dean and Gemawatt.[8]

6. IMPLEMENTATION OF HADOOP ARCHITECTURE FOR RFID TRAFFIC CONTROL SYSTEM

In this section, the elaboration given on reading the tag information which contains the Vehicle Identification Number and location and distance information on Map/Reduce framework. Map/Reduce framework works on <key, value> format[9](Refer Figure 4). It works on the basis of parallel method when data can be processed in chunks from all nodes in parallel manner. While the vehicle passes by, it goes through a toll centre, it will collect all the information of the vehicle, the data will be taken by the client machine on that centre. Hadoop is a potential solution to the problem. As Hadoop works on the record format, hence entire database is stored in a text format which is converted by the reader and uploaded on HDFS. Here the key is nothing but the tag which may provide identification or location information, or specifics about the number of the car and value to check whether it is present in the database or if not present then newly can be added. Mappers take its input from HDFS and reducer stores the final result on HDFS. [10]

6.1 Distributed File System of Hadoop

HDFS is distributed in nature, which saves all data files of the Hadoop cluster. The HDFS structure is in Master/Slave mode. The system includes one NameNode, one Secondary NameNode, and many DataNodes. NameNode is the file system's administrator, in charge of managing and storing the file system's descriptive metadata coming from the DataNode, as well as the users' access to documents. Actual data storage is in each DataNode, which needs to report to NameNode periodically.

The HDFS is applied to the design the vehicle tracking system by RFID.

The author is suggesting to assign the Data Node as the Toll Booth Machine and Name Node as the Server Machine where all data is kept which is getting transferred from time to time from the Toll Centers. Reader unit of RFID will be placed at toll-booth and along the motorable roads around the core area which will detect each individual vehicle uniquely within its zone by capturing their device identification number along with VIN and will keep records of the time during which the vehicle was seen by those readers within its reading zone. This information will be sent to a central server. Accordingly the central server will calculate the charges and raise bills against the vehicle ids. The system for automatic messaging can also be kept as Map/Reduce framework works on <key, value> format. Here, the key will be taken as device id and the VIN; the value is whether the charges paid by the owner of the vehicle.

This technique is used when any of the situations occurs from the following-

- If the vehicle found in this area with no charges paid then that information will be stored. This status will be sent to the Name Node by the Data Node. The authorized person at the central place where Name Node is kept can take action against the owner of the vehicle.
- Calculate the speed of a motorist and to detect if he violates the prescribed speed limit.
- If congestion is detected, the system would control traffic signals can generate automatic re-routing messages by using Map Reduce to selected approaching vehicles.

The author recommends having the working of the vehicle tracking system in the following manner in the future. She has suggested a plan to work mainly in three steps for the RFID based toll booths by using Hadoop Distributed File System along with Map Reduce Framework

1. When the toll booth will detect the vehicle coming towards it, with the help of RFID, the detector will accept its information and will store it on the machine where toll booth data is stored from time to time. Here, the author is suggesting to use the Map Reduce Framework on Hadoop Architecture. The machine where all information of the vehicle is stored can be treated as a Data Node. In case of Hadoop, the Data Node is nothing but the client which will pass all the information to the server. The server is the Primary Name Node, where all the information by the data node will be passed to store the Master copy of the information on the Primary Name Node.

2. When authority at the central place needs to get information, request is sent to NameNode. NameNode divides the file into blocks, allocates these blocks to DataNode. Every time RFID device will make connection to data node, will pass the information to it.

3. HDFS uses copy storage policy. Generally, every block will have three copies at the same time. NameNode maintains a working DataNode list through receiving periodic reports from DataNode. If NameNode does not receive report within a period from some DataNode, it regards the DataNode as failed. Then NameNode copies and stores the blocks on the DataNode to another DataNode. It keeps every block in system having three copies anytime to ensure system's reliability.

6.2 Map Reduce Framework on Master Client for Vehicle Tracking System

The client in this system has got all the details of the RFID as well as the vehicle number and the amount details which are already paid by the owner of the vehicle. This is the map task which is executed by worker in the Master Worker framework of Map Reduce Framework of Hadoop Architecture. These details are submitted to the server machine where all related data is stored from all the vehicles which have passed through the toll centre. By using Map Reduce framework the server assigns the map task to all clients. The server collects all the results and consolidates the intermediate result by assigning the reduce task to the clients. Number of reduce workers depends on the size of the intermediate result. After collecting all reduce results from clients, server keeps the data with it. The master can assign the either map or reduce task to the workers.

7. CONCLUSIONS

The author believes that the task will be easier for the India Government officials to track the vehicle when it passes through the Toll centers. This system has not been implemented yet, once it gets implemented by making use of RFID, the data such as device id, VIN and other owner's information can be processed by Map Reduce operations of Hadoop architecture. The data will be easily available and can be processed in a parallelized manner at many places by this architecture. The suggested technique will give the faster results in terms of performance for tracking the vehicle. The author is planning to propose the study of the Cloud computing paradigm in order to improve the scalability and boost the performance of RFID systems by using Hadoop.

REFERENCES

[1]http://www.livemint.com/2009/10/18222403/Insurers-use-radio-tags-to-cov.html

[2]http://www.devx.com/enterprise/Article/32251/1763/page/1

[3] Intel Research "An Introduction to RFID Techonolgy " Roy Want

[4]Anuran Chattaraj, Saumya Bansal, and Anirudhha Chandra "AN INTELLIGENT TRAFFIC CONTROL SYSTEM USING RFID" 0278-6648/09 IEEE

[5]Prof Somprakash Bandyopadhyay "Traffic Congestion Management Using RFID & Wireless Technologies" INDIAN INSTITUTE OF MANAGEMENT CALCUTTA

[6]Orizin Technologies: Vehicle Tracking Solution using RFID

[7]Apache Hadoop, http://hadoop.apache.org/

[8]Jeffrey Dean, Sanjay Ghemawat, "*MapReduce*: *simplified data processing on large clusters"*, Communications of the ACM, v.51 n.1,January 2008

[9]http://hadoop.apache.org/common/docs/currentlhdfs design.html,2009.

[10]Sandhya S V, Sanjay H A, Netravathi S J, Sowmyashree M V, Yogeshwari R N "Fault–Tolerant Master-Workers framework for MapReduce Applications" IEEE Computer Society 2009