

# A New Approach for Secure Traffic Data Analytics using Hadoop

<sup>1</sup> A.Pramod Kumar, <sup>2</sup> B.Sateesh

<sup>1</sup> Assistant Professor, Matrusri Engineering College, Hyderabad, T.S, India

<sup>2</sup> Assistant Professor, Chaitanya Bharathi Institute of Technology, Hyderabad, T.S, India

**Abstract:** Nowadays traffic congestion is an extreme difficulty. Traffic congestion is most predominant in metro towns. There are distinct reasons for traffic congestion such as growing population, rising earning main to greater automobiles on the road, the inadequate capability of roads to address traffic and so on. A cluster of garage devices is wanted to save such big amounts of data and also a parallel computing version for studying the ones huge inputs of data. Hadoop is one such framework that gives the reliable cluster of storage facility, which stores huge statistics in an allotted way using a unique report machine, known as Hadoop Distributed File System and presents efficient parallel processing feature through MapReduce framework. Using Map Reduce the filtered traffic data may be fetched without problems, to offer quit user with traffic analysis and giving beneficial predictions.

**Keywords-** HDFS: Hadoop Distributed File System, Map Reduce

## I. INTRODUCTION

The data of transportation industry have rich sources, diverse types, and new data is produced continually. customer information of rail way, road traffic, aviation industry, public transit, are recorded ,and tens of billionstravel records are generatedevery year. Operating data generated by transport companies, such as the data generated by the courier companies. 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, and the data are generated by GPS vehicle location tracking system and other mobile deviceseach year the amount of data generated by the transportation industry in a cityhas exceeded TB level, are developing from PB level to EB.A massive data storage space and equipment are required and it must have fault tolerance and stability.

One of the maximum useful applications of the structures hired in visitors control is the stepped forward ability to control the road community site visitors. Tracking strategies are used to seize the location and/or location of the large variety of cellular items. With the help of that tracked data, evaluation and prediction of traffic density in a given network is superior.

This renders precious data for controlling traffic waft, prediction of congestion and reducing the quantity of accidents in that network. The subject of huge data for resolving the above queries affords a brand new technical approach. Big data carried out to road visitors evaluation has the following benefits:

1. Traffic control mechanism that uses huge information generation can handle sizable amounts of complex and diverse facts. Big data have resolved 3 principal troubles: statistics storage data evaluation and data control. Hadoop is instinctive with the capability to handle large amounts of information wherein facts is segmented and is stored on different nodes. A big task is split into small duties and is processed in a MapReduce version. At the equal time, the device balance and fault tolerance are crucial.

2. Big data can improve the efficiency of transportation enterprise in large part. Transportation enterprise, related to many factors of work, need to address big amounts of data every day, desires more controlled mode of utility and has a superb deal of system. In the issue of improving delivery efficiency, enhancing the threshold capability of the road network, adjusting site visitors demands, large records technology has obvious blessings.

## II. RELATED WORKS

A smart metropolis framework for intelligent traffic device using VANET turned into proposed in [2].

VANETs offer communicate between vehicles themselves and among automobiles and street aspect units. VANETs help improves the mobility of vehicles on the road and develops a more secure and state-of-the-art town. Also, easier communicate facility among motors is furnished using VANETs. But the issue with VANET is that a selected hardware needs to be established on each car. Installation of such hardware on -wheeler could be a hard assignment. An infrared primarily based clever traffic gadget turned into proposed in [3]. In this IR primarily based machine, the IR transmitter first produces a 38KHz service wave that's then modulated at a lower frequency to ship facts. At the receiver end, the processor converts this energetic-low sign into a widespread TTL stage signal. The IR based totally gadget proves to be cost-effective due to the fact all of the microcontrollers used are cheap. The transmitter and receiver are powered by using batteries and thus consume much less strength. Also, the setup method is clean since the transmitter and receiver may be established everywhere on the street. But the quandary with IR based totally gadget is that it requires the receivers and transmitters to be in direct line of sight of each different. This reduces the amount of pliability in motion. Another disadvantage is that excessive depth or fluorescent lighting purpose interference in the infrared signal. Also, massive areas require multiple emitter panels and satisfactory of these emitter panels varies with corporation.

In [4] fuzzy logic is used to optimize the traffic light timing at a Diphasic's remoted intersection. Here fuzzy logic controllers are used – one is to optimize the signal and different controller is used to extend the inexperienced phase of a lane in an intersection. The sensors used to collect input records are video cameras that are the region of incoming and outgoing lines. The controller then makes use of the statistics gathered through those sensors to make the top of the line decisions and decrease the aim characteristic. This device showed incredible upgrades in site visitors manipulate as compared to constant time controllers the use of exclusive visitors situations inclusive of sure, unsure and random data.

Although the paper states development in velocity and high precision but the use of fuzzy gadget proves transparent most effective for easy structures.

A heterogeneous network combining RFID and WSN is proposed in [5] to improve the performance of roads. In [6] the paper focused on the basic framework of the sensible city Traffic Management System Based on Cloud Computing and Internet of Things. The Internet of factors (IOT) is a form of pc network that's on the basis of the Internet. It uses Radio Frequency Identification (RFID) and wi-fi facts verbal exchange technology to assemble a community which covers the whole lot. The hassle with RFID systems is they may be without difficulty disrupted. Since RFID systems make use of the electromagnetic spectrum they're surprisingly easy to jam the usage of energy on the right frequency. Also, RFID has security, privacy, and ethics trouble. RFID tags can be read without the consumer's expertise. There is an ability vulnerability to present day RFID software program if used together with a backend database.

In [7] an Intelligent go road traffic management system has proposed the use of lengthy variety photoelectric sensors. The visitor's management department chooses the suitable distance to install these sensors, in order to reveal the transferring cars. This information is then sent to the traffic manipulate cabinet wherein software is set up for you to calculate the relative weight of every avenue. Based on the relative weight calculated the machine will allow the overcrowded avenue to have the larger period of the sign. Also, this device is designed in this kind of way that it may take care of emergency situations (along with passing of ambulance, ministries and other VIPs) with the aid of opening entire paths for these motors to bypass first. In this way, fluency in the traffic is guaranteed. The gadget designed by means of Salama [7] faces challenge when you consider that it calls for the deployment of photoelectric sensors. The price of retaining these sensors is large given that they may be deployed in rugged outside situations and are liable to harm.

In [8] a vision based clever visitors management gadget is proposed. The proposed gadget uses picture frames acquired thru cameras established on roads. The first step is to perform car detection. Also, background subtraction and other morphological operations have been used to boom the performance of vehicle detection. The region of Interest-based

method is used to achieve an accurate vehicle depend. The proposed machine shows correct actual time overall performance, but this machine faces positive boundaries consisting of occlusion and shadow overlapping.

### **III. THE PROPOSED APPROACH**

A dataset which consists of information about traffic is created. Map Reduce is used for processing to get the secure analysis of traffic data. Based on the analysis, prediction of traffic is made to show at what time it will be high and low in a day for a particular area. Prediction is also made for which month will have most number of accidents. Authentication is provided using signature, hence user security is ensured. Encryption algorithm is used to protect data, as it is sent across network.

#### **Hadoop Distributed File System – HDFS:**

A cluster in HDFS consists of a unmarried name node which is a master server device and is answerable for handling the complete document gadget. It additionally presents an get admission to to the record device as requested with the aid of the clients. Each cluster additionally consists of a couple of facts nodes, and each statistics node, in turn, consists of blocks of split records. Data distribution to those records nodes is maintained via the master node. The predominant operations of records nodes include record examine and write. Hadoop adopts the idea of rack recognition, which in particular facilitates in storing information into a rack and locating its region in the cluster. This way that a Hadoop Administrator can really define what chunks of records should be saved on a specific rack. This manner, it is feasible to mitigate the information loss if a rack fails entirely. Network performance is likewise advanced due to the fact statistics replication is completed on numerous racks of different machines.

#### **MapReduce:**

MapReduce is a software program framework which turned into added via Google to perform parallel processing on big datasets. This big statistics set is sent over a massive quantity of machines found in a cluster. For brief get right of access to, every machine computes and shops the facts locally, this, in turn,

contributes to dispensed parallel processing. Such a computation entails two parts Map and Reduce. In the Map section, data nodes take uncooked enter information and produce intermediate facts primarily based on the form of computation and then that records is stored locally. In Reduce phase, intermediate outputs from map phase are fetched with the aid of the nodes after which it's miles combined to derive very last output that is saved in HDFS. Name node with its previous knowledge of the data distribution attempts to assign the venture to a specific node based at the locality of information. Developers can write custom map and decrease features appropriate to the application and the MapReduce characteristic then looks after dispensing and parallelizing tasks across a rack on commodity hardware within the cluster below.

Inter-machine communication is controlled via the framework, consequently programmers simplest should awareness on real map-reduce capabilities. Hadoop makes use of this framework to investigate huge datasets disbursed over HDFS due to its unbeatable fault tolerance, reliability, dispensed and parallel computing functions.

Architecture design is a diagram which represents the basic structure of the entire project. It includes the various components that are a part of the project and how the components are connected. It also shows the actions performed by each component. Fig.1 shows the architecture design where the user needs to first sign in to communicate with the server. The user is provided with options to generate a dataset and also to manually insert an entry into the dataset.

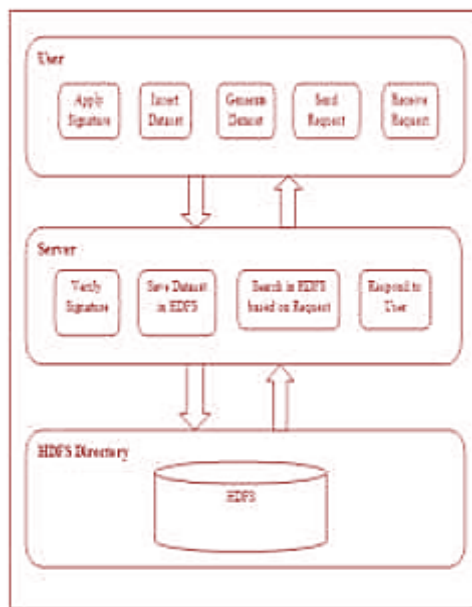


Fig.1 System Architecture

The other operations of server include storing the dataset into HDFS and fetching the matching data from the directory as requested by the user. It does so by sending the request to HDFS directory where the Map and Reduce phase takes place according to the input request data. The HDFS directory, when finishes the reduce phase, sends the result back to the server where the data is encrypted and sent across network. Once the user's signature is verified, decryption takes place and the requested information is sent to the user. The user on receiving the result of requested query, can analyze the data easily.

#### Signature Generation Algorithm

**Input:** Public key (A,B,h) , system parameters , message M

**Output:** Generate a valid group signature on M  
Select random numbers a, roM, roR, mus,mux, mueprime, mut, muE

Computes the following values

$$E0 = g * roE$$

$$E1 = h + (h1 * roE)$$

$$E2 = h + (h2 * roE)$$

$$ACOM = (A * (a2^{rom} \bmod n)) \bmod n$$

$$s = (Eprime + ke) * roM$$

$$BCOM = (B * (w^{roR} \bmod l)) \bmod l$$

$$t = Eprime * roR$$

$$V0 = g * muE$$

$$V1 = (g * mux + (h1 * muE))$$

$$V2 = (g * mux) + (h2 * muE)$$

$$Vmpk = (((a1^{mux} \bmod n) * (a2^{mus} \bmod n)) \bmod n) * (ACOM^{~mueprime} \bmod n) \bmod n)$$

$$Vrev = ((w^{mut} \bmod l) * (BCOM^{~mueprime} \bmod l)) \bmod l$$

$$V = V0 + V1 + V2$$

$$reste = ACOM + BCOM + V + Vmpk + Vrev$$

$$\text{Set } c = f(E + reste + \text{message})$$

Construct the following numbers

$$\text{taux} = c * (x + mux)$$

$$\text{taus} = c * (s + mus)$$

$$\text{taut} = c * (t + mut)$$

$$\text{tauePrime} = c * (Eprime + mueprime)$$

$$\text{tauE} = (c * (roE + muE)) \bmod o$$

Return

$$\sigma = (E0, E1, E2, ACOM, BCOM, c, \text{taux}, \text{taus}, \text{tauePrime}, \text{taut}, \text{tauE})$$

End

#### Signature Verification Algorithm

**Input:** System parameters and signature  $\sigma =$

(E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)

**Output:** True or False

Compute the following values

$$\text{taue} = (c * (\exp K_e + \text{tauePrime}))$$

$$\text{tauEG} = g * \text{tauE}$$

$$a0a1 = (a0^c \bmod n) * (a1^{\text{taux}} \bmod n) \bmod n$$

$$a2A = (a2^{\text{taus}} \bmod n) * (ACOM^{~\text{taue}} \bmod n) \bmod n$$

$$Vmpk = (a0a1 * a2A) \bmod n$$

$$Bw = ((b^c \bmod l) * (w^{\text{taut}} \bmod l)) \bmod l$$

$$Vrev = (bw * (BCOM^{~\text{tauePrime}} \bmod l)) \bmod l$$

$$E = E0 + E1 + E2$$

$$V = V0 + V1 + V2$$

$$\text{reste} = ACOM + BCOM + V + Vmpk + Vrev$$

$$\text{if } c = f(E + \text{reste} + \text{message})$$

Return True

else

Return False

End

#### Algorithm used for Busy/Idle Traffic Prediction

**Input:** Traffic details from MapReduce

**Output:** Time when the traffic is low, average, high

**Step 1:** Filter traffic between 6 to 10

**Step 2:** Find highest number of vehicles

For all traffic details

{

**Find:** number of vehicle / highest number of vehicles

```

}
Step 3: Sort average
Step 4: Display time when traffic is less
if (val<0.5)
{
Display as best case
}
else if (val<0.8)
{
Display as average case
}
else
{
Display as bad case
}
Step 5: Stop

```

#### IV. CONCLUSION

Each user is authenticated and only upon signature verification the requested info is furnished to the users. The proposed machine is designed for cozy evaluation and giving predictions approximately the traffic drift details in an afternoon and variety of accidents taking place in line with the month. Overall, that data which the user needs to retrieve may be effectively fetched using the Big Data techniques inclusive of Hadoop and MapReduce technology, by way of overcoming all the limitations that had been encountered by the prevailing system. At present, the application operates offline, within the destiny adjustments that make it work online have to be integrated.

#### REFERENCES

- [1] "Traffic Prediction and Analysis using a Big Data and Visualisation Approach" - Declan McHugh, Department of Computer Science, Institute of Technology Blanchardstown, March 10, 2015
- [2] "Traffic Accident Analysis Using Machine Learning Paradigms" - Miao Chong, Ajith Abraham and Marcin Paprzycki, Computer Science Department, Oklahoma State University, USA, December 20, 2004
- [2] Ganesh S. Khekare, Apeksha V. Sakhare, "A Smart City Framework for Intelligent TrafficSystem Using VANET,"International Multi-Conference onAutomation, Computing, Communication, Control and Compressed Sensing, Kottayam, 2013, pp. 302-305.
- [3] Sikder Sunbeam Islam, KowshikDey, Mohammed RafiqulIslam , Mohammad KurshedAlam, "An Infrared Based Intelligent Traffic System,"International Conference on Infonnatics, Electronics & Vision, Dhaka, 2012, pp. 57-60.
- [4] Hamid SaadatTalabandHadisehMohammadkhani, (2017,July). Design optimization traffic light timing usingthe fuzzy logic at a Diphasic'sIsolated intersection. Journal of Intelligent & Fuzzy Systems.27(4), pp.1609–1620.
- [5] LejiangGuo, Wei Fang, GuoshiWang, Long sheng Zheng, "Intelligent Traffic Management System Base on WSN and RFID," International Conference on Computer and Communication Technologies in Agriculture Engineering, Chengdu, 2010, pp. 227-230.
- [6] Xi Yu, Fuquan Sun, Xu Cheng (2012,Aug). Intelligent Urban Traffic Management System Based on Cloud Computing and Internet of Things. Presented at 2012 International Conference on Computer Science and Service System.
- [7] Ahmed S. Salama, Bahaa K. Saleh, Mohamad M. Eassa, "Intelligent Cross Road Traffic Management System (ICRTMS),"2nd International Conference on Computer Technology and Development, Cairo, 2010, pp. 27-30.
- [8] Muhammad Hassam Malhi, Muhammad Hassan Aslam, Faisal Saeed, OwaisJaved, Muhammad Fraz, "Vision Based Intelligent Traffic Management System," Frontiers of Information Technology, Islamabad, 2011, pp. 137- 141. Zoran Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction," Proceedings of the 17th International Conference on Pattern Recognition, 2004, pp. 28-31

[9] Pakize S, Gandomi A (2014), “Comparative Study of Classification Algorithms Based On MapReduce Model”, International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1(7), 251-254.

[10] Bhagattjee B (2014), “Emergence and Taxonomy of Big Data as a Service”.

[11] Ashish Vijay, Priyanka Trikha, Kapil Madhur, “A New Variant of RSA Digital Signature”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012, ISSN: 2277 128X.

[12] Gang Zeng, “Application of Big Data in Intelligent Traffic System”, IOSR Journal of Computer Engineering (IOSR-JCE), eISSN:2278-0661, p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. VI(Jan –Feb. 2015), PP 01-04.