A Innovative Approach forSecure Traffic Data Analytics using Hadoop

Bagadi Siddarth[1]

[1] Lecturer in CSE Dept, JNTU-Hyderabad ,Telangana, India.

Abstract: Nowadaystraffic 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 richsources, 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. Dynamicdata 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 locationtracking system and other mobile deviceseach year the amount of data generated by the transportation industryin a cityhas exceeded TB level, are developing from PB level to EB.A massive data storage space and equipmentare required and it must have fault tolerance and stability.

One of the maximum useful applications of the structureshired in visitors control is the stepped forward ability to controlthe road community site visitors. Tracking strategies are used to seizethe location and/or location of the large variety of cellularitems. With the help of that tracked data, evaluation andprediction 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 accidentsin that network.The subject of huge data for resolving the above queriesaffords a brand new technical approach. Big data carried out to roadvisitors evaluation has the following benefits:

1. Traffic control mechanism that uses huge information generationcan handle sizable amounts of complex and diverse facts. Big data have resolved 3 principal troubles: statistics storage data evaluationand data control. Hadoop is instinctive with the capability tohandle large amounts of information wherein facts is segmented and isstored on different nodes. A big task is split into smallduties and is processed in a MapReduce version. At the

equaltime, the device balance and fault tolerance are crucial.

2. Big data can improve the efficiency of transportationenterprise in large part. Transportation enterprise, related to manyfactors of work, need to address big amounts of data everyday, desires more controlled mode of utility and has a superbdeal of system. In the issue of improving deliveryefficiency, enhancing the threshold capability of the roadnetwork, adjusting site visitors demands, large records technology hasobvious blessings.

II. RELATED WORKS

A smart metropolis framework for intelligent traffic device usingVANET turned into proposed in [2]. VANETs offercommunique between vehicles themselves and amongautomobiles and street aspect units. VANETs help improves themobility of vehicles on the road and develops a more secure andstate-of-the-art town. Also, easier communique facilityamong motors is furnished using VANETs. But theissue with VANET is that a selected hardware needsto be established on each car. Installation of suchhardware on -wheeler could be a hard assignment.An infrared primarily based clever traffic gadget turned into proposedin [3]. In this IR primarily based machine, the IR transmitter firstproduces a 38KHz service wave that's then modulatedat a lower frequency to ship facts. At the receiver end, theprocessor converts this energetic-low sign into a widespreadTTL stage signal. The IR based totally gadget proves to be cost-effective due to the fact all of the microcontrollers used are cheap. Thetransmitter and receiver are powered by using batteries and thus consume much less strength. Also, the setup method isclean since the transmitter and receiver may be establishedeverywhere on the street. But the quandary with IR based totallygadget is that it requires the receivers and transmitters tobe in direct line of sight of each different. This reduces theamount of pliability in motion. Another disadvantage isthat excessive depth or fluorescent lighting purpose interferencein the infrared signal. Also, massive areas require multipleemitter panels

and satisfactory of these emitter panels varieswith corporation.

In [4] fuzzy logic is used to optimize the traffic lighttiming at a Diphasic'sremoted intersection. Here fuzzy logic controllers are used – one is to optimize thesignal and different controller is used to extend the inexperiencedphase of a lane in an intersection. The sensors used tocollect input records are video cameras that are the region ofincoming and outgoing lines. The controller then makes use ofthe statistics gathered through those sensors to make the top of the line decisions and decrease the aim characteristic. Thisdevice showed incredible upgrades in site visitors manipulate ascompared to constant time controllers the use of exclusive visitorssituations inclusive of sure, unsure and random data.

Although the paper states development in velocity and highprecision but the use of fuzzy gadget proves transparent most effectivefor easy structures.

A heterogeneous network combining RFID and WSN isproposed in [5] to improve the performance of roads. In [6]the paper focused on the basic framework of the sensible cityTraffic Management System Based on Cloud Computingand Internet of Things. The Internet of factors (IOT) is aform of pc network that's on the basis of theInternet. It uses Radio Frequency Identification (RFID)and wi-fi facts verbal exchange technology to assemblea community which covers the whole lot. The hassle withRFID systems is they may be without difficulty disrupted. SinceRFID systems make use of the electromagnetic spectrumthey're surprisingly easy to jam the usage of energy on the rightfrequency. Also, RFID has security, privacy, and ethicstrouble. RFID tags can be read without the consumer'sexpertise. There is an ability vulnerability to present dayRFID software program if used together with a backend database.

In [7] an Intelligent go road traffic management systemhas proposed the use of lengthy variety photoelectric sensors. Thevisitor's management department chooses the suitabledistance to install these sensors, in order to revealthe transferring cars.

This information is then sent to the traffic manipulatecabinet wherein software is set up for you to calculate therelative weight of every avenue. Based on the relative weightcalculated the machine will allow the overcrowded avenue tohave the larger period of the sign. Also, this device is designedin 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 thatit calls for the deployment of photoelectric sensors. Theprice of retaining these sensors is large given that they may bedeployed in rugged outside situations and are liable toharm.

In [8] a vision based clever visitors management gadgetis proposed. The proposed gadget uses picture framesacquired thru cameras established on roads. The first stepis to perform car detection. Also, backgroundsubtraction and other morphological operations have beenused to boom the performance of vehicle detection. The regionof Interest-based method is used to achieve an accuratevehicle depend. The proposed machine shows correct actual timeoverall performance, but this machine faces positive boundariesconsisting 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 oftraffic data. Based on the analysis, prediction of traffic is madeto show at what time it will be high and low in a day for aparticular area. Prediction is also made for which month willhave most number of accidents. Authentication is providedusing signature, hence user security is ensured. Encryptionalgorithm 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 amaster server device and is

answerable for handling thecomplete document gadget. It additionally presents an get admission to to the record deviceas requested with the aid of the clients. Each cluster additionally consists ofa couple of facts nodes, and each statistics node, in turn, consists ofblocks of split records. Data distribution to those records nodes ismaintained via the master node. The predominant operations of recordsnodes include record examine and write.Hadoop adopts the idea of rack recognition, which in particularfacilitates in storing information into a rack and locating its region in thecluster. This way that a Hadoop Administrator can reallydefine what chunks of records should be saved on a specificrack. This manner, it is feasible to mitigate the information loss if a rackfails entirely. Network performance is likewise advanced due to the factstatistics replication is completed on numerous racks of different machines.

MapReduce:
MapReduce is a software program framework which turned into added viaGoogle to perform parallel processing on big datasets. Thisbig statistics set is sent over a massive quantity of machinesfound in a cluster. For brief get right of access to, every machine computesand shops the facts locally, this, in turn, contributes todispensed parallel processing. Such a computation entails two parts Map and Reduce. In the Map section, data nodestake uncooked enter information and produce intermediate facts primarily based on theform of computation and then that records is stored locally. InReduce phase, intermediate outputs from map phase are fetchedwith the aid of the nodes after which it's miles combined to derive very last output thatis saved in HDFS. Name node with its previous knowledge of the data distribution attempts to assign the venture to a specific nodebased at the locality of information. Developers can write custom mapand decrease features appropriate to the application and theMapReduce characteristic then looks after dispensing andparallelizing tasks across a rack on commodity hardware within thecluster below.
Inter-machine communication is controlled via theframework, consequently programmers simplest

should awareness on real map-reduce capabilities. Hadoop makes use of this framework to investigatehuge datasets disbursed over HDFS due to its unbeatablefault tolerance, reliability, dispensed and parallel computingfunctions.

Architecture design is a diagram which represents the basicstructure of the entire project. It includes the variouscomponents that are a part of the project and how thecomponents are connected. It also shows the actions performedby each component. Fig.1 shows the architecture designwhere the user needs to first sign in to communicate with theserver. The user is provided with options to generate a datasetand also to manually insert an entry into the dataset.
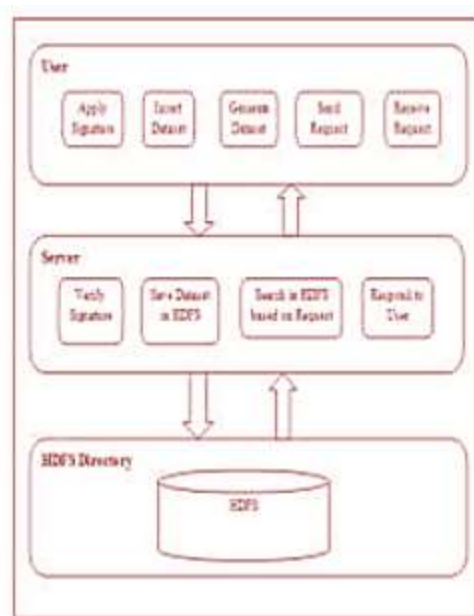

Fig.1 System Architecture

The other operations of server include storing thedataset into HDFS and fetching the matching data from thedirectory as requested by the user. It does so by sending therequest to HDFS directory where the Map and Reduce phasetakes place according to the input request data.The HDFS directory, when finishes the reduce phase, sends theresult back to the server where the data is encrypted and sentacross network. Once the user's signature is verified,decryption takes place and the requested

information is sent tothe 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
$E0 = g * roE$
$E1 = h+(h1*roE)$
$E2 = h+(h2*roE)$
$ACOM (A*(a2^{rom}) \bmod n).\bmod n$
$s = (Eprime+ke)* roM$
$BCOM (B *(w^{} roR \bmod l) \bmod l$
$t = Eprime * roR$
$V0 = g*muE$
$V1 = (g * mux+(h1 * muE)$
$V2 = (g*mux))+(h2* muE)$
$Vmpk= (((a1^{mux} \bmod n)*(a2^{mus} \bmod n)).\bmod(n))*(ACOM^{\sim mueprime} \bmod n)).\bmod(n))$
$Vrev= ((w^{mut} \bmod l))*(BCOM^{\sim mueprime} \bmod l)))\bmod lE =E0+E1+E2$
$V = V0+V1+V2$
reste = ACOM+BCOM+V+Vmpk+Vrev
Set c = f(E + reste + message)
Construct the following numbers
$taux = c * (x+mux)$
$taus = c * (s+mus)$
$taut = c *(t+mut)$
$tauePrime = c *(Eprime+mueprime)$
$tauE = (c*(roE+muE)) \bmod(o)$
Return
σ=(E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)
End

Signature Verification Algorithm

Input: System parameters and signature σ = (E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)
Output: True or False
Compute the following values

```
taue = (c*( expKe+tauePrime);
tauEG = g *tauE
a0a1 = (a0^c mod n)*(a1 ^ taux mod n))mod n
a2A= (a2^taus modn)*(ACOM^ ~taue mod n)mod n
Vmpk = (a0a1 * a2A)mod n
Bw= ((b^c mod l)*(w^taut mod l))mod l
Vrev = (bw*(BCOM^~tauePrime  mod l)) mod l
E = E0+ E1+E2
V = V0+V1+V2
reste = ACOM+ BCOM+ V + Vmpk+ Vrev
if c = f(E + reste + 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 authenticatedand only upon signature verification the requested info isfurnished to the users. The proposed machine is designed forcozy evaluation and giving predictions approximately the traffic driftdetails in an afternoon and variety of accidents taking place in line with the month.Overall, that data which the user needs to retrieve may beeffectively fetched using the Big Data techniques inclusive ofHadoop and MapReduce technology, by way of overcoming all thelimitations that had been encountered by the prevailing system.At present, the application operates offline, within thedestiny adjustments that make it work online have to be integrated.

REFERENCES

[1] "Traffic Prediction and Analysis using a Big Data andVisualisation Approach" - Declan McHugh, Department ofComputer Science, Institute of Technology Blanchardstown,March 10, 2015

[2] "Traffic Accident Analysis Using Machine Learning Paradigms"- Miao Chong, Ajith Abraham and MarcinPaprzycki, ComputerScience Department, Oklahoma State University, USA,December 20, 2004

[2] Ganesh S. Khekare, Apeksha V. Sakhare, "A Smart CityFramework for Intelligent TrafficSystem UsingVANET,"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 TrafficSystem,"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 aDiphasic'sIsolated

intersection. Journal of Intelligent & FuzzySystems.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 CommunicationTechnologies in Agriculture Engineering, Chengdu, 2010, pp. 227-230.

[6] Xi Yu, Fuquan Sun, Xu Cheng (2012,Aug). Intelligent UrbanTraffic Management System Based on Cloud Computing andInternet of Things. Presented at 2012 International Conference onComputer 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 ComputerTechnology and Development, Cairo, 2010, pp. 27-30.

[8] Muhammad Hassam Malhi, Muhammad Hassan Aslam, FaisalSaeed, OwaisJaved, Muhammad Fraz, "Vision Based IntelligentTraffic Management System," Frontiers of InformationTechnology, Islamabad, 2011, pp. 137-141.ZoranZivkovic,"ImprovedAdaptiveGaussianMix tureModelforBackground Subtraction," Proceedings of the 17th InternationalConference on Pattern Recognition,2004, pp. 28-31

[9] Pakize S, Gandomi A (2014), "Comparative Study ofClassification Algorithms Based On MapReduce Model",International Journal of Innovative Research in AdvancedEngineering (IJIRAE), 1(7), 251-254.

[10] Bhagattjee B (2014), "Emergence and Taxonomy of Big Data asa Service".

[11] Ashish Vijay, PriyankaTrikha, KapilMadhur, "A New Variantof RSA Digital Signature", International Journal of AdvancedResearch in Computer Science

and Software Engineering,Volume 2, Issue 10,October 2012, ISSN: 2277 128X.

[12] Gang Zeng, "Application of Big Data in Intelligent TrafficSystem", 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.