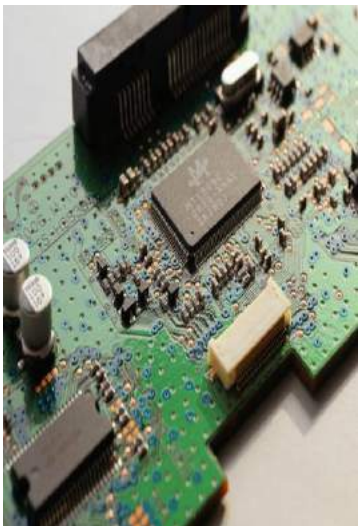
ISBN: 978-0-9957075-7-3

INTERNATIONAL JOURNAL of ENGINEERING AND APPLIED COMPUTER SCIENCE

EMPIRICAL RESEARCH PRESS

Volume: 02 | Issue: 06 | June | 2017






EMPIRICAL RESEARCH PRESS LTD.

London, United Kingdom

IJEACS

International Journal of

Engineering and Applied Computer Science

Empirical Research Press Ltd.

London, United Kingdom

Volume: 02, Issue: 06

ISBN: 978-0-9957075-7-3

www.ijeacs.com

Indexing, Hosting and Advertising

Internet Archive

Message

International Journal of Engineering and Applied Computer Science (IJEACS) is an open access, double-blind peer reviewed international journal, monthly online publishing by Empirical Research Press Ltd. Empirical Research Press is a research publishing company with name, trademark registered, incorporated in England and Wales, United Kingdom.

The scope of International Journal of Engineering and Applied Computer Science is to publish high quality research contributions, latest innovations, advance development carried out in the field of Engineering, Computer Science and Technology. The original research, review, case study, survey, new interpretation and implementation of concepts and theories, recent technological development, technical reports, empirical discussions are invited to submit for publication.

The major objectives of International Journal of Engineering and Applied Computer Science are to maintain high quality of publications and indexing with world's highly accessed and cited research and academic databases. The scope of IJEACS publications also includes special and interdisciplinary research contributions. We offer best wishes to readers, reviewers and contributors of IJEACS.

Board Members of IJEACS

Content

International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3, June 2017

Performance Assessment of Several Filters for Removing Salt and Pepper Noise, Gaussian Noise, Rayleigh Noise and Uniform Noise

Prashant Dwivedy
Department of Electrical & Electronics Engineering, National Institute of Technical Teachers' Training & Research, Bhopal, India

Anjali Potnis
Department of Electrical & Electronics Engineering, National Institute of Technical, Teachers' Training & Research, Bhopal, India

Madhuram Mishra
Department of Electrical & Electronics Engineering, National Institute of Technical Teachers' Training & Research, Bhopal, India

***Abstract*—Digital images are prone to a variety of noises. De-noising of image is a crucial fragment of image reconstruction procedure. Noise gets familiarized in the course of reception and transmission, acquisition and storage & recovery processes. Hence de-noising an image becomes a fundamental task for correcting defects produced during these processes. A complete examination of the various noises which corrupt an image is included in this paper. Elimination of noises is done using various filters. To attain noteworthy results various filters have been anticipated to eliminate these noises from Images and finally which filter is most suitable to remove a particular noise is seen using various measurement parameters.**

Keywords- Image de-noising; Image restoration techniques; Noise models; Average filter; Median filter; Poisson noise; Gaussian noise; PSNR; MS.

I. INTRODUCTION

During processing of digital images by means of digital computers de-noising or removal of noise is very essential [1]. Noise is an unwanted signal in image which gives change in visibility of any image and occurs usually due to thermal or electrical signals such as from sensors or environmental conditions. The problem at hand is removing the noise of an image while preserving its main features (edges, textures, colors, contrast, etc.) This has been widely examined over the last two decades and several types of approaches have been developed. There are two domain processes available for restoring the image, first one is spatial domain and second one is frequency domain. In the spatial domain filtering action is done by operating on the pixel of the digital image directly for restoring the image. On the other hand filtering action is completed by in frequency domain by mapping spatial domain into frequency domain of the image function by taking Fourier transform of the image function. After the filtering, in order to conclude the restored image we have to re map the image into spatial domain by taking inverse Fourier transform. Noise may be categorized as multiplicative noise for example speckle noise, substitutive noise for example salt and pepper noise and additive noise for example Gaussian noise. In this paper first image is occupied and noise to be deal is added to image to make it a noisy image and then noisy image is passed by filters. It becomes significant to de-noise the image before smearing to various applications [2].The principle approach of image de-noising is filtering. Numerous filters are used to eliminate noise such as averaging filters, median filters, mean filters etc. The image quality is measured by various performance parameters like the peak signal to noise ratio (*PSNR*) and mean square error (*MSE*) [3].

II. NOISE MODELS

Noise is an outcome of inaccuracy in image acquisition process [4] that results in pixel values that do not imitate true intensities of the actual picture. Using probability density functions we can describe a set of noise models. The most occurring noises in digital images are poisson noise, exponential noise, salt and pepper noise, Gaussian noise, multiplicative noise, Rayleigh noise, Erlang noise or Gamma noise and uniform noise. Following, these noises are discussed at stretch.

A. Salt and Pepper Noise

An image comprising salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions. It is also sometimes called Impulse Noise. This noise is usually caused by sudden and sharp disturbances in the image signal. It often presents itself as sparsely occurring black and white pixels. This type of noise can be produced by dead pixels, analog-to-digital convertor errors, bit errors in transmission, etc. If a = 0 (black) and b = 1 (white) then probability distribution is specified by

$$P(z) = \begin{cases} P_a \ for \ z = a; \\ P_b \ for \ z = b; \\ 0 \ otherwise \end{cases} \quad (1)$$

B. Gaussian Noise

Gaussian noise is statistical in nature having a probability density function equivalent to that of the normal distribution. Gaussian noise is normally a set of values taken from a zero mean Gaussian distribution [6] which is added to every pixel value. The distribution is given by the expression

$$P(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (2)$$

Where μ = mean of random variable of z, z = gray level and σ^2 = variance of z

C. Rayleigh Noise

The Rayleigh noise follows the Rayleigh distribution

$$P(z) = \begin{cases} \frac{2}{b}(z-a)e^{-\frac{(z-a)^2}{b}} & for\ z \geq a; \\ 0 & for\ z < a; \end{cases} \quad (3)$$

Rayleigh density can be used to approximate skewed image histograms [9].

D. Uniform Noise

This sort of noise generates a noise sequence and follows the uniform distribution function [11] with value ranging from a to b and is added uniformly to all the pixels of the image. The PDF of uniform noise [12] is specified by

$$p(z) = \begin{cases} \frac{1}{b-a} & if\ a \leq z \leq b \\ 0 & otherwise \end{cases} \quad (4)$$

III. Filtering Techniques

Removal of noise from the corrupted image is done by filtering. There are two types of filtering techniques [13]. First one being spatial filtering and second one is frequency filtering.

Spatial filtering is the filtering operations that are accomplished straight on the pixels of image. In Spatial Domain the filtering operation [14] is done by convolving the image pixels with the pixels of the mask. A mask is a small sub image, often of size 3 × 3 pixels. The mask size is varied according to the requirement. These include the following classes of filters

- Mean filters
- Order statistics filters
- Adaptive filters

A. Arithmetic Mean Filter

In this type of mean filter the middle pixel value of the mask is replaced with the arithmetic mean [15] of all the pixel values within the filter window. It calculates the average value of the ruined image $g(x, y)$ in the area defined by $S_{x, y}$. The assessment of the reestablished image $f(x, y)$ at any point (x, y) is

$$f(x,y) = \frac{1}{mn} \sum_{(s,t)\in S_{x,y}} g(s,t) \quad (5)$$

B. Geometric Mean Filter

The working of a geometric mean filter is same as the arithmetic mean filter; the only difference is that as a replacement for taking the arithmetic mean the geometric mean is taken. The reestablished image is given by the expression

$$f(x,y) = \left[\prod_{(s,t)\in S_{xy}} g(s,t) \right]^{\frac{1}{mn}} \quad (6)$$

C. Median Filter

Order-statistics filters are built on assembling the pixels enclosed in the mask. Median filter comes under this class of filters. Median filter exchanges the value of a pixel with the median value of the gray intensities within the filter window [17] or mask. Median filters are very effective for impulse noise.

$$f(x,y) = \underset{(s,t)\in S_{xy}}{median} \{g(s,t)\} \quad (7)$$

D. Max and Min filter

The max filter is beneficial for finding the brightest points in an image. Since pepper noise has very small values, it is condensed by this filter as an outcome of the max selection processing the sub image area $S\ x, y$.

$$f(x,y) = \underset{(s,t)\in S_{xy}}{max} \{g(s,t)\} \quad (8)$$

The min filter is valuable for finding the darkest points in an image. Also, it decreases salt noise [18] as a result of the min operation.

$$f(x,y) = \underset{(s,t)\in S_{xy}}{min} \{g(s,t)\} \quad (9)$$

E. Midpoint filter

This filter computes the midpoint between the maximum and minimum values in the area included by the filter. This filter works finest for arbitrarily distributed noise like Gaussian noise.

$$f(x,y) = \frac{1}{2}\left[\underset{(s,t)\in S_{xy}}{max} \{g(s,t)\} + \underset{(s,t)\in S_{xy}}{min} \{g(s,t)\} \right] \quad (10)$$

IV. Performance Measurement Parameters

Consider an image of dimensions M and N. If *f(x, y)* is the original image and *g(x, y)* is the distorted image then the various measurement parameters are described as follows.

A. Mean Square Error (MSE)

The MSE [19] is cumulative squared error between the compressed and the original image. It is calculated using

$$MSE = \frac{1}{MN}\sum_{0}^{M-1}\sum_{0}^{N-1}\|f(x,y) - g(x,y)\| \qquad (11)$$

B. Peak Signal to Noise Ratio (PSNR)

The PSNR is used to determine the ratio among the maximum power of a signal and power of corrupting noise. The formula of PSNR is given as

$$PSNR = 10\,log_{10}\left[\frac{M*N}{MSE}\right] \qquad (12)$$

C. Average Difference (AD)

The average difference is specified by the formula

$$|f(x,y) - g(x,y)| \qquad (13)$$

D. Maximum Difference (MD)

The maximum difference is specified by the formula

$$\max|f(x,y) - g(x,y)| \qquad (14)$$

E. Normalized Absolute Error (NAE)

The normalized error is specified by

$$y = NAE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(f(x,y) * g(x,y))}{\sum_{x=1}^{M}\sum_{y=1}^{N}(f(x,y))^2} \qquad (15)$$

F. Structural Content (SC)

SC is correlation based measure and measures the similarity between two images. It is specified by the equation

$$SC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(y(i,j))^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j))^2} \qquad (16)$$

V. Simulation Result and Analysis

Simulation has been run on Matlab using gray scale image 'lena.bmp' of size 512 x 512 as a test image shown in Fig 1

Fig. 1. Test Image Lena

Table 1 shows the different measurement parameters after applying all the filters for Rayleigh noise.

TABLE I. RAYLEIGH NOISE

FILTER	MSE	PSNR	AD	MD	NAE	SC
Arithmetic Filter	4.0900	42.0136	0.4832	100	0.0049	0.9993
Geometric Filter	5.3772	40.8252	0.6426	118	0.0065	0.9977
Harmonic Filter	8.8568	38.6581	1.0787	156	0.0109	0.9980
Contra-Harmonic Filter	252.150	24.1142	98.735	252	0.9969	160.65
Median Filter	4.4275	41.6693	0.5988	135	0.0060	0.9970
Max and Min Filter	70.7250	29.6351	9.1765	181	0.0926	1.0147
Mid-Point Filter	15.9580	36.1010	2.1084	89	0.0213	0.9975

Table 2 shows the different measurement parameters after applying all the filters for Salt and Pepper noise.

TABLE II. SALT AND PEPPER NOISE

FILTER	MSE	PSNR	AD	MD	NAE	SC
Arithmetic Filter	27.0425	33.8103	3.4363	113	0.0347	0.9977
Geometric Filter	30.5430	33.2817	8.5275	232	0.0861	0.9982
Harmonic Filter	34.0357	32.8115	9.1250	231	0.0921	0.9996
Contra-	252.125	24.1146	98.708	252	0.9966	153.33

Harmonic Filter						
Median Filter	9.2991	38.4464	1.3019	135	0.0131	0.9978
Max and Min Filter	94.1594	28.3922	17.950	229	0.1812	1.0191
Mid-Point Filter	15.9580	36.1010	2.1084	89	0.0213	0.9975

Table 3 shows the different measurement parameters after applying all the filters for Gaussian noise.

TABLE III. GAUSSIAN NOISE

FILTER	MSE	PSNR	AD	MD	NAE	SC
Arithmetic Filter	31.9247	33.0895	3.8760	113	0.0391	0.9978
Geometric Filter	252.1500	24.1142	98.925	252	0.9988	186.19
Harmonic Filter	252.1500	24.1142	98.925	252	0.9988	185.95
Contra-Harmonic Filter	252.1500	24.1142	99.046	253	NaN	NaN
Median Filter	49.0823	31.2216	4.7516	135	0.0480	1.0064
Max and Min Filter	243.519	24.2655	37.450	215	0.3781	1.0500
Mid-Point Filter	15.9580	36.1010	2.1084	89	0.0213	0.9975

Table 4 shows the different measurement parameters after applying all the filters for Uniform noise.

TABLE IV. UNIFORM NOISE

FILTER	MSE	PSNR	AD	MD	NAE	SC
Arithmetic Filter	0.2494	54.1623	0.0387	66	0.00039092	0.9998
Geometric Filter	0.0071	69.6468	0.0007476 8	26	0.0000075488	0.9997
Harmonic Filter	0.0484	61.2823	0.0054	54	0.000054459	0.9996
Contra-Harmonic Filter	251.509	24.1253	98.063	252	0.9901	99.972
Median Filter	0.0093	68.4459	0.0019	135	0.0000190 65	0.9997
Max and Min Filter	8.7633	38.7041	1.1406	160	0.0115	0.9988
Mid-Point Filter	15.9580	36.1010	2.1084	89	0.0213	0.9975

VI. CONCLUSION

On seeing the factors revealed in the overhead tables we can conclude that which filter will be best for removing respective noise. This conclusion is stated below in the Table V.

TABLE V. FILTERS TO BE CHOSEN FOR DIFFERENT NOISES

NOISES	FILTERS
Rayleigh Noise	Arithmetic Mean Filter
Salt and Pepper Noise	Median Filter
Gaussian Noise	Mid-Point Filter
Uniform Noise	Geometric Mean Filter

REFERENCES

[1] R.C. Gonzalez and R.E.Woods"Digital Image Processing.

[2] Suresh Kumar, Papendra Kumar, Manoj Gupta, Ashok Kumar Nagawat, "Performance Comparison of Median and Wiener Filter in Image De-noising", International Journal of Computer Application, Vol.12 – No.4, November 2010.

[3] C.Saravanan, R. Ponalagusamy"Gaussian Noise Estimation Technique for Gray Scale Images Using Mean Value". Journal of Theoretical and Applied Information technology. 2005-2007

[4] Pawan Patidar, Manoj Gupta and Sumit Srivastava, "Image De – Noising by Various Filters for Different Noise", International Journal of Computer Application, November 2010.

[5] M.S.Alani, Digital Image Processing using Mat lab, University Bookshop, Sharqa, URA, 2008.

[6] G. Pok, J. Liu, and A. S. Nair, "Selective Removal of Impulse Noise Based on Homogeneity Level Information," IEEE Trans. Image Processing, vol. 12, pp.85–92, Jan. 2003.

[7] E. Abreu, M. Lightstone, S. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," IEEE Trans. Image Processing, vol. 5, pp. 1012-1025, June 1996.

[8] K. S. Srinivasan and D. Ebenezer, _A new fast and efficient decision based algorithm for removal of high density impulse noises, IEEE Signal Process. Lett. vol. 14, no. 3, pp. 189–192, Mar. 2007.

[9] S. P. Awate and R. T. Whitaker, "Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering," in *Proc IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 44–51.

[10] T. Batard and M. Berthier, "Spinor Fourier transform for image processing," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 4, pp. 605–613, Aug. 2013.

[11] P. Blomgren and T. F. Chan, "Color TV: Total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process*, vol. 7, no. 3, pp. 304–309, Mar. 1998.

[12] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 60–65

[13] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," Image Process. On Line, vol. 2, pp. 175–213, Aug. 2012

[14] M. Lebrun, M. Colom, and J. M. Morel, "The noise clinic: A universal blind denoising algorithm," in *Proc. IEEE Int. Conf. Image Process*, Oct. 2014, pp. 2674–2678.

[15] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2011, pp. 2833–2840.

[16] M. Lysaker, S. Osher, and X.-C. Tai, "Noise removal using smoothed normals and surface fitting," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1345–1357, Oct. 2004.

[17] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," Multiscale Model. Simul., vol. 4, no. 2, pp. 460–489, 2005.

[18] T. Rahman, X.-C. Tai, and S. Osher, "A tv-stokes denoising algorithm," in Scale Space and Variational Methods in Computer Vision (Lecture Notes in Computer Science), vol. 4485. Berlin, Germany, Springer Verlag, 2007, pp. 473–483.

AUTHORS' PROFILE

Prashant Dwivedy received his B.Tech degree in Electronics & Communication Engineering in 2014 from Gurukula Kangri Vishwavidyalaya, Haridwar, Uttarakhand and currently pursuing M.Tech in Digital Communication Engineering from NITTTR, Bhopal. His area of interest includes Digital Image Processing, Digital Signal Processing.

Dr. Anjali Potnis is Professor at Department of Electrical & Electronics Engineering, National Institute of Technical Teachers" Training & Research Bhopal. She has got a total of 16 years of teaching experience. She has published many research papers. Her area of interest includes Digital Image Processing and Digital Signal Processing.

Madhuram Mishra received his B.E degree in Electronics & Communication Engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal and currently pursuing M.Tech in Digital Communication Engineering from NITTTR, Bhopal. His area of interest includes Digital Image Processing and Digital Communication.

International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3, June 2017

Survey on Big Data Analytics

Dhruva M.S
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

Shashikala M.K
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

***Abstract*—This paper aims to highlight distinct features of Big of information. We are living on the planet with huge varieties and tremendous volume of data Information is the new money. Data is being generated by sensors, digital images, activities in social media sites, forensic science, business informatics, research activities across various domains and many other internet based events as either source data or as cumulative to the existing data. This data increase from time to time exponentially from heterogeneous sources, techniques and technologies. The data is categories as "Big Data". The core theme focuses on linking Big Data with people in various ways. Big Data is enormous in Variety, Velocity and Volume. It may be in any of the form like structured, unstructured. The idea of Big Data analysis is to manage giant volume of data, obtain beneficial information, suggest casualties and support decision making. This survey provides comprehensive review and audit of Big Data analytics. Leading and evolving applications of Big Data analytics are discussed. Some of the techniques for efficient analysis of Big Data are also illustrated.**

Keywords- Big Data management; Big Data analytics; Big Data analyzing technique.

I. INTRODUCTION

It was in early 21st century concept of Big Data came into existence and started to evolve. It was the first time when attributes like volume, structure and speed were used for the describing the nature of data [1]. Big Data's important attribute is the volume. Data is quantified by counting the space occupied, digital transactions, statistical tables, or files but it was found more constructive to symbolize [9] Big Data with respect to time [3]. The very next is the variety of data. This happens, as data come from variety of sources like census, blogs, logs, streams, issue of nationalized identification cards, research data, partial structured data from business-to-business processes, satellite images. The last attribute is the velocity that refers to speed of applying the analytics and processing the data.

Big Data is vast in majority and complex data. Dissimilarity, storage and transport [6], privacy and security, and complexity problems with Big Data impede the progress at all stages of that can create value from data. There are various sources of Big Data, for example: Opinion polls, audio-visual files, scientific data, various database tables, email attachments etc. Big Data has great importance in fields like research, public sector services, healthcare services, web/social, manufacturing, artificial intelligence, education and cyber-physical models. Big data have priority in every sector in the global economy [2]. It was calculated that by 2005, practically all arena in the economy will have 200 terabytes of minimum data stored per company having more than 1,000 employee [11]. Big data forward to enlarge rapidly, driven by mutation and modification in elementary technologies [15]. Conventional data administration and analysis system substantially depend on Relational database management system (RDBMS) [17].

There major aspects in which RDBMS and Big Data differs are:

1) RDBMS is limited to structured data, but big data supports various data processing architectures [18].

2) RDBMS provides an insight to a problem at the small level, big data offer better view and efficient operations on metadata and unstructured data [18].

At the point when does examination turn out to be Big Data Analytics? The size that characterizes Big Data has developed. In 1975 participants of the primary VLDB (Very substantial databases) gatherings stressed over dealing with the Millions of information focuses found in US statistics Information. Huge Data Analytics is the course of classifying bulk datasets to the variety of data type i.e. indirect relations, digitized documents, consumer priorities and other useful details. The examination can prompt productive showcasing, better nature of administrations [1]. Huge Data examination venture are promptly rising as the honorable answer for perceive business and innovation slants that are irritating conventional information administration strategies. Examination finds necessity and conceivable arrangements. With enormous information investigation, the associations are attempting to recognize leave surveys, new business actualities and patterns. This paper incorporates writing overview of Big Data examination in segment 2. Segment 3 contains foundation and information types of Big Data. Segment 4 contains Big Data investigation in detail and area 5 contains systems to break down enormous information and segment 6 finishes up the paper [8].

II. LITERATURE SURVEY

Over last many years, there are many researchers and scholars completed their work successfully on Big Data. Many articles have been published in the various journals and magazines (For example Forbes, Harvard Business review, Optimize, The Wall street journal). The Government of India has implement enormous [12] techniques of Big Data to determine the feedback of Indian electorate to government plans and policies. The Obama Administration has announced that, it would invest the 200 million dollars on big data research plan in March 2012[13].

Reports of International Data Corporation predicts that global Data from 2005 to 2020 will grow by factor of 10. The global data volume will grow from 0.13 Zettabyte's to 40 Zettabyte's, depicting double accumulation for every two years. IBM evaluates that everyday 2.5 quintillion bytes of data will be originated. Out of which 90% of the data in the world today is created from the last two years.

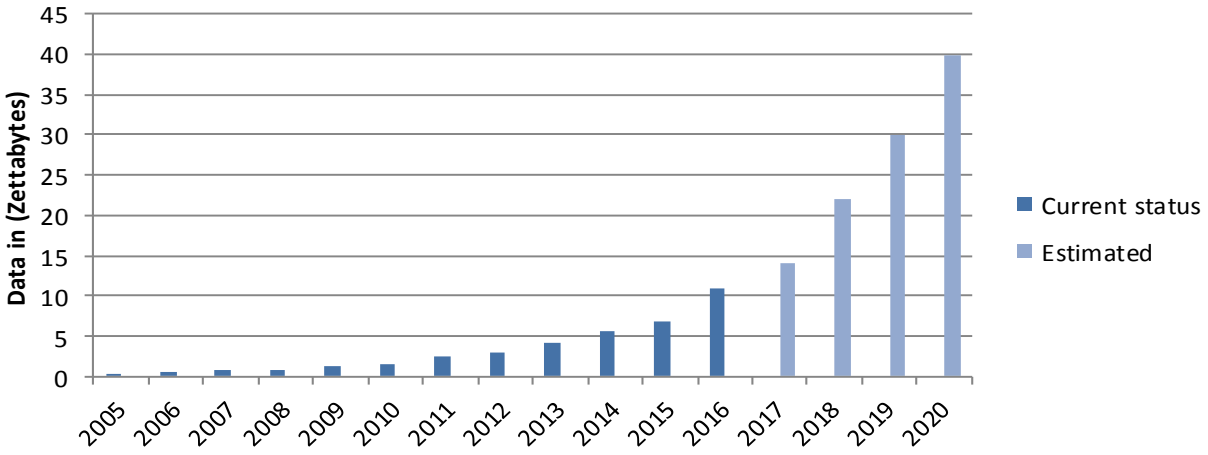


Figure 1. Data volume growth by year in zettabytes

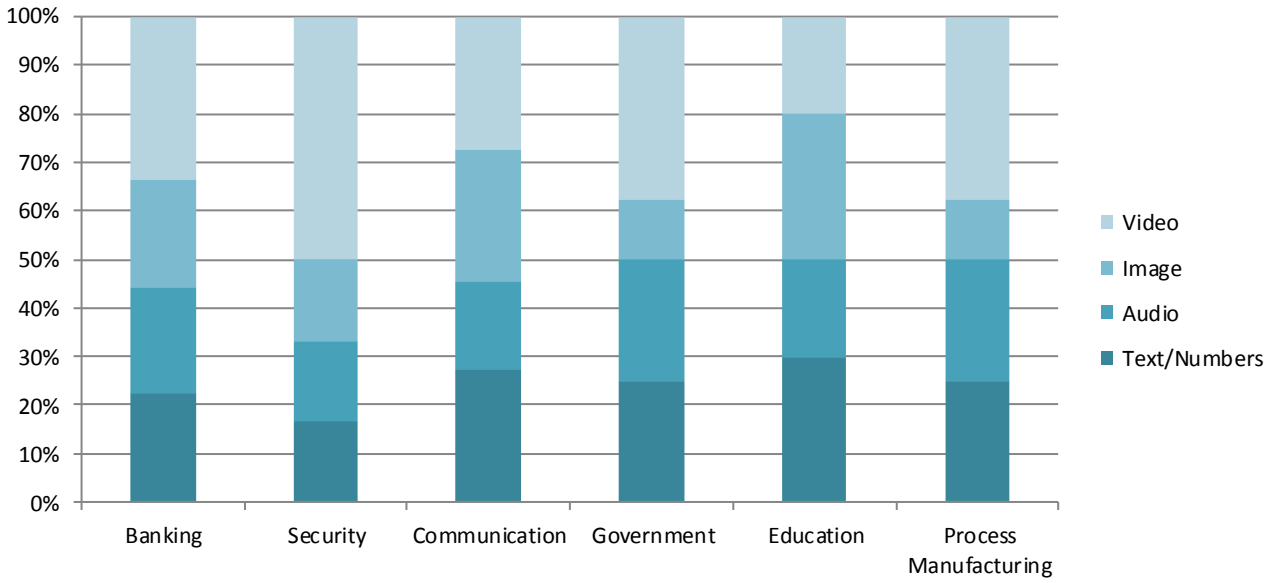


Figure 2: Variations possible in generating and growth of data

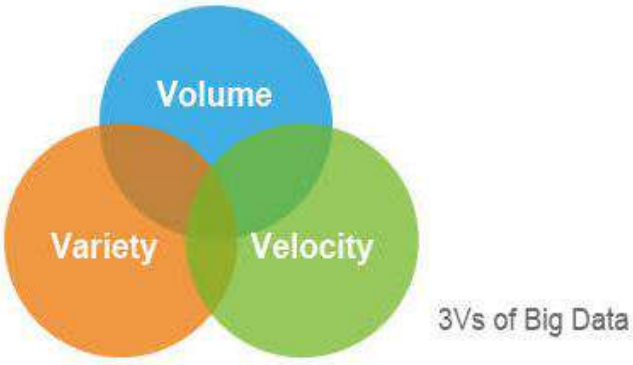
It is observed that social networking sites like Facebook have 750 million users with 350 million photo uploads per day, LinkedIn has 110 million addressee and Twitter has 320 million accounts with 500 million tweets per day. From industry, government and research community, Big Data has led to advancement in the research field that has attracted immense interest [14].

The major concern is coverage on both industrial reports and public media for example: The Economic Times, The Hindu, Times of India. Smart devices and mobile phones are the best way to get data from people in divergent aspect [17], the huge magnitude of data that mobile carrier can process to make our day to day life easier. In the Figure 1, it would present that the amount of data practically increased from the year 2005 to 2016 and the estimation, that data would increase from the 2017 to 2020. However, consider exponential growth in data from the year 2005, when enterprise system and user level data flood into data warehouse [10].

Figure 2 illustrates the diversity in data stored from different sectors. The type of data induced and stored are audio, video, digital images and text format and differ from one sector to another. Text/numeric data will be from the sectors that are directly related to research and development community, public zone like banking, government and health care [1]. Audio and video nature of data is from various fields of communication and media.

III. BIG DATA

Enormous Data is the term that contains expansive and complex datasets. It is dreary work to deal with these datasets without new innovation. The Mckinsey Global Institute (MGI) distributed a give an account of Big Data that portrays the different business openings that huge information uncover. Paulo Boldi, One of the creators says "Enormous Data does not require huge machines; it needs huge knowledge [13].



A. *V's of Big Data*

- Volume: This would allude to the information from various sources, information being in enormous limit. It can incorporate all and any sort of information, including the information that is made from all the associated gadgets, versatile information, web and every one of the information that is being come about because of this correspondence. [19]
- Velocity: Speed not just includes the speed at which the information is exchanged, however will likewise include, information streams, formation of organized records, access to information and conveyance. The issues don't just lie with the speed of approaching information additionally to stream active information for cluster preparing. [10]
- Variety: This alludes to shifted information sorts and the same can be amassed from different sources, sources being: interpersonal organizations, cell phone, sensors in the types of recordings, pictures, sound, logs and so on. This information can be profoundly organized (information gotten from the conventional database frameworks), semi-organized (nourishes like surveys, remarks) or unstructured (snaps, sounds, pictures, recordings). [8]

B. *Data Forms*

Data collected can be broadly classified into the following categories.

1) *Structured data*: This alludes to shifted information sorts and the same can be amassed from different sources, sources being: interpersonal organizations, cell phone, sensors in the types of recordings, pictures, sound, logs and so on [5]. This information can be profoundly organized (information gotten from the conventional database frameworks), semi-organized (nourishes like surveys, remarks) or unstructured (snaps, sounds, pictures, recordings). Details will be provided such that data with respect to which columns are placed where, whom are they associated with and how the columns are associated in between tables. The organization of the information can be in content or numerical, however it is regular understanding that for each individual there is a one of a kind identifier as far as Age. [7]
   The whole information is composed as far as Entities (Semantic Chunks). [5]
   - Relations or Classes (Similar elements are gathered together). [7]
   - Attributes (Same portrayals for elements existing in similar gatherings)[1]
   - Schema (All Entities in the gathering have a portrayal related with it. [2]
   - All are available and take after same request.[3]
   - All of them have same organization characterized and length characterized. [5]

2) Semi-organized information: The configuration of information don't affirm an express and correct mapping, however the labels related with the information, if discovered related with authoritative structure, at that point similar information would be less demanding to sort out

and break down. A similar idea depicted here would originate before the possibility of XML.

- Data is accessible in many configurations, in the present situation, electronically
- File Systems e.g., Web information [11]
- Data Exchange Formats, e.g., Scientific information
- Data that is not totally organized, but rather
- Similar sections will be assembled and semantically sorted out
- Entities might not have same traits in the gathering

3) *Unstructured information*: Unstructured information would be in an arrangement that can't be effectively filed. Ordering is the technique for alluding social tables with the end goal of questioning or investigation. This would incorporate the record sorts that are related with sound, video and picture documents. [10]

- Data – Any sort.
- No Format and legitimate successions.

IV. Big Data Analytics

Big Data analytics permit enterprises for better analysis of a mix of structured, semi structured and unstructured resulted due to reviews by the customers, precious business statistics. The Mckinsey Global Institute propagated a major research work in June 2011 on Big Data. Its overloading conclusion: Big Data is "a key basis of competition and growth". The expression Analytics (inclusive of Big Data form) is often used broadly to wrap up data-driven decision making. The term analytics classified into major subdivision: Corporate analytics and Academic research analytics. In Corporate Analytics, data is treated as the asset and major concentration is on increasing the revenue. In Academic Analytics, Researchers make use of data to test Hypothesis and form theories [6].

Researchers of big data analytics have found the data collected is divided into various Big Data application such as follows [17].

A. *Structured Analysis*

In structured analytics, data is generated from high degree of business organizations and scientific research fields. These data is organized and queried by RDBMS, Data warehousing, and various search algorithms. Data is grown by different research areas like Privacy, preserving, data mining, E-commerce [10].

B. *Text Analytics*

In Text analytics, text the most common way of storing the information and it includes e-mail, digital libraries, chat messages, and social media contents. Text analytics also known as Text mining, concentrate [9] on deriving correct and effective information from massive text file. Text mining system is relay on statistical pattern learning and Natural Language Processing (NLP) with importance on the letters.

C. *WebAnalytics*

The objective of Web analytics is to fetch the information from Web Pages [6]. Web Analytics also called Web mining.

D. *Multimedia Analytics*

Multimedia data includes animated sequences, graphic objects, and computer aided draft and drawing, audio-visual files. It has grown at a gigantic rate. Multimedia analytics refers to extract advantageous knowledge and semantics exist in multimedia data. Data types of multimedia data are printable characters, sound, volume, pixels.

E. *Mobile Analytics*

Mobile data traffic increased to 7.2 Zettabytes per month at the end of 2016. Vast collection of data and application leads to mobile analytics. Mobile analytics involves RFID (Radiofrequency identification), mobile phones, sensors etc.

V. Technique for Analysis of Big Data

There are several techniques that can be used to process datasets. Some techniques are machine learning, A/B testing. These techniques, analyze new combination of datasets

A. *A/B Testing*

A technique in which particular or reference group is compared with a variety test of groups to determine the best performance between variants. Reference group is a constant called as control group and the test group is the variable called as the treatment group. Changes will be implemented on the objective variable, e.g., acceptance rate of products. An example application is fraud detection with suspect as reference (constant) and forensic data collected at the crime scene as a treatment group (variable). When the variable manipulated in the treatment is more than one, technique is often called "A/B/N" testing. [15]

B. *Classification*

A method in which to recognize the classifications of new datasets and dole out into predefined classes for instance grouping of mushroom as consumable or toxic. It is utilized for information mining [14].

C. *Crowdsourcing*

A technique in which collected data submitted by large group of people or community i.e. crowd. It is usually through network media such as web [6].

D. *Data Mining*

Method in which exact pattern of data from large existing datasets are examined to generate new information by exercising certain rules. It has applications in machine learning and artificial intelligence [2].

VI. CONCLUSION

In this paper, idea of Big Data is explored. Huge information is the huge entomb related datasets and it create from different sources like web-based social networking, remarks and audits, brilliant and sensible gadgets, email connections and so on. There is many-sided quality in Big Data, for example, Velocity, Variety and Volume. These three terms are all the more trying for Big Data investigation. If writing overview indicates exponential development of information in businesses from 2005 year. There are varieties conceivable while producing and putting away information whether information is in sound, video, pictures and content. In Big Data Analytics, specialists partitioned created information into different enormous information application, for example, organized information investigation, content examination, web investigation, interactive media examination and portable examination. Many difficulties in the enormous information framework require additionally look into consideration. Investigate on common Big Data application creates benefit for business association, upgrade the adequacy of government segments among the general population.

ACKNOWLEDGMENT

I would like to thank my guide and all people who encouraged and helped me to prepare this paper. Finally, I'm indebted to all websites and journal papers which I have refer to prepare this survey paper successfully.

REFERENCES

[1] Understandable Big Data: A survey Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle LE2I UMR6306, CNRS, ENSAM, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

[2] Yuri Demchenko ―The Big Data Architecture Framework (BDAF) Outcome of the Brainstorming Session at the University of Amsterdam 17 July 2013.

[3] Vishal S Patil, Pravin D. Soni , " HADOOP SKELETON & FAULT TOLERANCE IN HADOOP CLUSTERS ", International Journal of Application or Innovation in Engineering & Management (IJAIEM)Volume 2, Issue 2, February 2013 ISSN 2319 – 4847.

[4] Sanjay Rathe, " Big Data and Hadoop with components like Flume, Pig, Hive and Jaql" International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV.

[5] Yaxiong Zhao, Jie Wu and Cong Liu, " Dache: A Data Aware Caching for Big-Data Applications Using the MapReduce Framework ",TSINGHUA SCIENCE AND TECHNOLOGY ISSNl1007-0214l 05/10l lpp39-50 Volume 19, Number 1, February 2014.

[6] Parmeshwari P. Sabnis, Chaitali A.Laulkar , "SURVEY OF MAPREDUCE OPTIMIZATION METHODS ", ISSN (Print): 2319-2526, Volume -3, Issue -1, 2014.

[7] Puneet Singh Duggal ,Sanchita Paul ," Big Data Analysis Challenges and Solutions ", International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV.

[8] Chen He,Ying Lu,David Swanson, " Matchmaking: A New MapReduce Scheduling Technique ", EECS Department, University of California, Berkeley, Tech. Rep.,April 2009.

[9] Apache HDFS. Available at http://hadoop.apache.org/hdfs [14] Apache Hive. Available at http://hive.apache.org.

[10] Apache HBase. Available at http://hbase.apache.org

[11] Apache Pig. Available at http://pig.apache.org

[12] A Review Paper on Big Data Analytics Ankita S. Tiwarkhede1, Prof. Vinit KakdeInternational Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[13] Survey Paper on Big Data. C. Lakshmi, V. V. Nagendra Kumar International Journal of Advanced Research in Computer Science and Software Engineering. Volume 6, Issue 8, August 2016.

[14] ARPN Journal of Engineering and Applied Sciences ©2006-2015 Asian Research Publishing Network (ARPN). VOL.10, NO. 8, MAY 2015 ISSN 1819-6608

[15] P. Russom, et al. Big data analytics, TDWI Best PracticesReport, Fourth Quarter.

[16] American Institute Of Physics(AIP), 2010. College Park, MD(http:// www.aip.org /fyi/2010/)

[17] http://www.oyster-ims.com/wp-content/uploads/2014/01/Global-datavolume

[18] http://www.deltapowersolutions.com/media/images/news/news-2014-big-data-3v(en)

AUTHORS' PROFILE

Dhruva M.S. completed his B.E. degree and M.Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India. Currently he is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. His areas of interest include multimedia networks, Compiler Design and Algorithms.

Shashikala M. K completed her B.E. degree and M. Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India Currently she is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. Her areas of interest include networks, algorithms.

International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3, June 2017

Implementation of Multi-node Clusters in Column Oriented Database using HDFS

P. Naresh
SICET, Hyderabad, India

G. Nataraja Shekhar
SICET, Hyderabad, India

M. Kiran Kumar
SICET, Hyderabad, India

P. Rajyalakshmi
HITS Hyderabad, India

***Abstract*—Generally HBASE is NoSQL database which runs in the Hadoop environment, so it can be called as Hadoop Database. By using Hadoop distributed file system and map reduce with the implementation of key/value store as real time data access combines the deep capabilities and efficiency of map reduce. Basically testing is done by using single node clustering which improved the performance of query when compared to SQL, even though performance is enhanced, the data retrieval becomes complicated as there is no multi node clusters and totally based on SQL queries. In this paper, we use the concepts of HBase, which is a column oriented database and it is on the top of HDFS (Hadoop distributed file system) along with multi node clustering which increases the performance. HBase is key/value store which is Consistent, Distributed, Multidimensional and Sorted map. Data storage in HBase in the form of cells, and here those cells are grouped by a row key. Hence our proposal yields better results regarding query performance and data retrieval compared to existing approaches.**

Keywords- Database; Cluster; Hadoop; HDFS; MapReduce.

I. INTRODUCTION

The term "Big Data" denotes to the unremitting huge expansion in the volume and variety along with velocity of data processing. Volume denotes to the scale of the data and processing needs. Variety denotes to the diverse layouts of information like images, booklets, videos, streams. Velocity denotes to the rapidity at which the information is being produced. Data is increasing extremely by day to day. The IDC evaluation and put the dimension of the "digital universe" at 0.18 zettabytes in 2006, and is projecting a tenfold growth by 2011 to 1.8 zettabytes.

- Structured data: It describes data that resides in fixed field files record tables, and it is standard database and data contained in relational databases and spreadsheets. The data arrangement and consistency permits it to react to basic quires to land at usable data in view of association's parameters.
- Semi-structured data: It is a database model where there is no division between the data and schema, it is not a permanent schema and it contains labels or different markers to uphold chains of importance of records and fields inside of the information.
- Unstructured data: It alludes to data that either does not have a predefined information show or is not sorted out in pre-characterized way. This sort of information can't without much of a stretch be recorded into social tables for investigation or questioning like pictures, sound, and feature documents.
- Velocity: Commonly velocity refers to the speed of the data processing. Ordinary comprehension of speed regularly is how rapidly the information arrives and put away and how rapidly, it can be retrieved. Velocity ought to additionally be connected to information in movement. The rate at which the information is streaming is found. The different data streams and the increment in sensor system arrangement have prompted a consistent stream of information at a pace that has made it unimaginable for customary frameworks to handle.

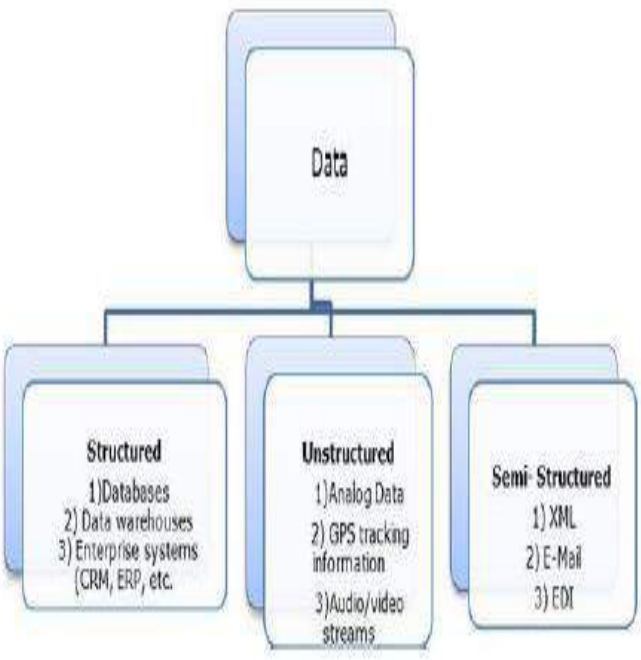


Figure 1. Types of data.

MapReduce is a programming model for expansive scale data preparing and MapReduce [5]. Programs written in different dialects like Java, Python, C++, Ruby, and so forth one essential point is the MapReduce projects are naturally parallel. MapReduce carries its own weight for substantial datasets. Here Datasets is nothing but a data can be stored as a semi-structured and record-oriented and data arrangement bolsters a rich arrangement of meteorological components, a large number of which are discretionary or with variable information lengths.

- Restricted parallel programming model meant for large clusters.
- Map Reduce divides the workload into multiple independent tasks and schedule them across cluster nodes.
- In a Map Reduce cluster [4], data is distributed to every one of the hubs of the group as it is being stacked in.

HDFS refers to "Hadoop Distributed File System" [2] HDFS is designed for the storing of large documents with gushing information access examples, running on bunches of product equipment. The key part of HDFS is its architecture itself.

A. *HDFS Architecture*

- The HDFS is designed in the form of a master and slave pattern.
- Here, the master is the NameNode and the slave is the DataNode.

B. *NameNode*

Here NameNode acts as a master node of HDFS and it organize all the data node's file system operations and it maintain replication node and maintains the file system tree which stored at local disk in the form of two files.

C. *NameSpace Image*

- The purpose of this is that it captures the snapshot of what the file system comprises of.
- The below is how a namespace image would be in the name node.
- This consists of the file_name, replication factor and block_ids.

We know that enormous data turns out consistently from the web. In previous this huge area of data is taken care of by Relational database administration frameworks (RDBMS) [4]. But know a day's predominated technology is there so at a time RDBMS can't handle by structured and semi-structured and unstructured data. The new technology NoSql is a non-relational database administration framework, not the same as customary relational database administration frameworks in some critical ways. It is intended for distributed data stores where huge size of information putting away need.

II. RELATED WORK

WUMUTI·NAHEMAN, IANXINWEI [1] they introduce the structural planning and information model of HBase database and is illustrative of NoSQL databases and did some performance test on HBase database include column family test according to their results he says query rate of HBase is moderate under single machine environment , however can be altogether enhanced in multi-machine bunch environment. Furthermore, they know the network data it is big challenge in the market it is explosive growth and NoSQL databases [2] have been widely used in some scenarios it can be combination of relational database to make up the defects of their own and above results of NoSQL database are not mature enough so his future research is integration of relational database and NoSQL database.

Jyoti Nandimath, Ankur Patil, Ekata Banerjee, Pratima Kakade, Saumitra Vaidya [7] confronted issues of effectiveness, execution and lifted base expense with the information handling in the unified environment with help of conveyed structural engineering so these expansive associations had the capacity defeat the issues of removing applicable data from a colossal information dump. This is the one of the best devices utilized as a part of the information handling issues in Apache hadoop, using the Apache Hadoop's different segments, for example, an information groups, guide decrease calculations and circulated preparing.

A. *Row-Oriented Data Stores*

- Information is put away and recovered one column at once and subsequently could read pointless information if a percentage of the information consecutively is needed.
- Easy to peruse and compose record
- Well suited for OLTP framework
- Not effective in performing operations pertinent to the whole dataset and subsequently conglomeration is an extravagant operation.
- Regular pressure instruments give less successful results than those on segment situated information store.

B. *Column-Oriented Data Stores*

- Information is put away and recovered in segments and subsequently can read just important information if some information is needed.
- Perused and Write are commonly slower operations.
- Appropriate for OLAP frameworks.
- Can effectively perform operations pertinent to the whole dataset and thus empowers total over numerous lines and sections.
- Allows high pressure rates because of couple of particular qualities in segments.

III. IMPLEMENTATION

HBase is a column oriented database that executed by Google's Big Table stockpiling building design. It can oversee organized and semi-organized information and has some implicit elements, for example, versatility, forming, pressure and waste accumulation. Since its uses compose ahead logging and disseminated setup, it can give adaptation to internal failure and snappy recuperation from individual server disappointments.

HBase based on Hadoop/HDFS [8] and the information put away in HBase can be controlled utilizing Hadoop's MapReduce[5] abilities. How about we now investigate how HBase (a segment situated database) is unique in relation to some other information structures and ideas that we are acquainted with Row-Oriented versus Segment Oriented information stores. As demonstrated as follows, consecutively arranged information store, a line is a unit of information that is perused or composed together. In a section arranged information store, the information in a segment is put away together and subsequently immediately recovered.

Row ID	Customer	Product	Amount
101	John White	Chairs	$400.00
102	Jane Brown	Lamps	$500.00
103	Bill Green	Lamps	$150.00
104	Jack Black	Desk	$700.00
105	Jane Brown	Desk	$650.00
106	Bill Green	Desk	$900.00

Figure 2. Simple HBase Table.

The Data Model in HBase is intended to suit semi-organized information that could differ in field size, information sort and segments. Moreover, the format of the information model makes it less demanding to segment the information and appropriate it over the bunch. The Data Model in HBase is made of diverse consistent parts, for example, Tables, Rows, Column Families, Columns, Cells and Versions.

A column is one example of information in a table and is distinguished by a rowkey. Rowkeys are one of a kind in a Table and are constantly regarded as a byte. All lines are dependably lexicographically by their column key. Example:

Hbase (main):001:0> scan 'table1'

ROW COLUMN+CELL

row1 column=cf-1:, timestamp=1297073325971 ..

row10 column=cf-1:, timestamp=1297073337383 ...

row11 column=cf-1:, timestamp=1297073340493 ...

row-2 column=cf1:, timestamp=1297073329851 ...

row-22 column=cf1:, timestamp=1297073344482 ...

row-3 column=cf1:, timestamp=1297073333504 ...

row-abc column=cf1:, timestamp=1297073349875 ...

7 row(s) in 0.1100 seconds

Observe very carefully the numbering is not sequence because of random sorting order. In lexicographical sorting, and every key is compared through binary level, byte by byte from left to right.

A. *Column Families*

Information consecutively is gathered together as Column Families. Every Column Family has one more Columns and these Columns in a family are put away together in a low level stockpiling record known as HFile. Segment Families frame the essential unit of physical stockpiling to which certain HBase.

The table beneath shows Customer and Sales Column Families. The Customer Column Family is made up 2 sections – Name and City, though the Sales Column Families is made up to 2 segments – Product and Am.

Row Key	Customer		Sales	
Customer Id	Name	City	Product	Amount
101	John White	Los Angeles, CA	Chairs	$400.00
102	Jane Brown	Atlanta, GA	Lamps	$200.00
103	Bill Green	Pittsburgh, PA	Desk	$500.00
104	Jack Black	St. Louis, MO	Bed	$1600.00

Column Families

Figure 3. Column Family.

IV. SYSTEM ANALYSIS

Generally cluster means group of similar things or occurring closely together. A Hadoop bunch is an exceptional sort of computational group planned particularly for putting away and examining gigantic measures of unstructured information in appropriated processing environment. Hadoop bunch is open source disseminated preparing programming keep running on thing PCs.

In Hadoop cluster environment, we have many machines are there but here we choose the following way one machine in the bunch assigned as the Name Node and another machine go about as a Job Tracker, these Name Node and Job Tracker are the masters and rest of the machine act as both Data Node and Task Tracker these two are the slaves. And here we remember that have a network connection between nodes. Hadoop clusters are used for boosting the pace of information investigation applications. Hadoop clusters are high resistance to failure due to every bit of the information is replicated onto other bunch hubs so information is not lose if hub falls flat.

Single hub bunch can be valuable to work application focus on a solitary server and Application focus treats a solitary hub group or stand-alone server single-hub bunches can profit by application focus without working in bunch environment. Before Installation of java it can require a working java 1.5 installations and it follow the commands like

$sudo apt-get update

$sudo apt-get install sun-java6-jdk

$sudo update-java-alternatives -s java-6-sun

Cluster run in only one machine and it did not take multiple machines but in this single node machine we have to create the tables and column families and this data can be stored in the HDFS database and tables are splits into regions and this regions are served by the region servers and here query performance is slow because of single cluster and see the results on single node cluster one by one. We already know the HBase [6] is a section arranged database it is open source usage of the BigTable and it is based on top of the Hadoop and HDFS and the information put away in HBase can be embarrassed utilizing Hadoop's MapReduce abilities. It is acquainted with Row-Oriented and Column-Oriented information stores and it can be intended to store Denormalized information contains wide and inadequately populated tables and bolster Automatic Partitioning. It is assembled for low Latency operations and gives the entrance to single lines from billions of records and information is gotten to through shell commands.

$start-hbase.sh

$hbase shell

Figure 4. Map-Reduce Capabilities.

Figure 5. Multi-node cluster set-up health check-up.

V. CONCLUSION

We know the world become a narrower and present generation peoples are expect the reaction of the meticulous products, actions, issues rapid on the web so we use the automatic sentiment analysis it can be recognize and foresee present and future patterns, item audits individuals verdict about social issues. Observe the present market, the incident about individuals data will be generated day-day very large-scale so this data will hoard the concepts of Google, the concepts like Hadoop and their eco-system here ecosystems are developed by some Google developers in that eco-system provide HBase so we are focussed on performance the multi-node cluster unit by using Hbase and HDFS.

REFERENCES

[1] WUMUTI. NAHEMAN, JIANXIN WEI, "Review of NoSQL Databases and performance Testing on HBase", International Conference on Mechatronic Sciences Electric Engineering and Computer (MEC), Dec 20-22, 2013, Shenyang, China.

[2] Lizhi, Shidong Huang, Leilei Chen, Yang Zheng, "Performance Analysis and Testing ofHBase Based on Its Architecture", International Conference on IEEE Machatronic Science and ISBN:978-1-4799-0174-6/13.

[3] A B M Moniruzzaman and Syed Akhter Hossain, "NoSQL Database: New Era of Databases for Big data Analytics Classification, Characteristics and Comparison", International Journal of Database Theory and Application vol. 6, No. 4. 2013.

[4] D. Carstoiu, A. Cernian, A. Olteanu, "Hadoop HBase-0.20.2 Performance Evaluation", "Politehnica" University of Bucharest. IEEE Journal Paper.

[5] I. Tomasic, A. Rashkovska and M. Depolli. "Using Hadoop MapReduce in a Multicluster Environment", MIPRO 2013, May 20-24, 2013, Opatija, Croatia.

[6] Zhou Quan, Chunming Tang, "TSHC: Trusted Scheme for Hadoop Cluster", 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies.

[7] JyothiNndimath, AnkurPatil, "BigbData Analysis Using Apache Hadoop" IEEE IRI 2013, August 14-16, 2013 IEEE, San Francisco, California, USA.

[8] Ognjen V. Joldzic, Dijana R. Vukovic, "The Impact of Cluster Characteristics on HiveQL Query Optimization", 21st Telecommunications forum TELFOR 2013.

[9] Mehmet C.Okur, Mustafa Buyukkececi, "Big Data Challenges in Information Engineering Curriculum", 2014 IEEE International Journal Paper.

[10] Santhosh Kumar Gajendran, "A Survey on NoSQL Databases ", IEEE NoSQL Database survey paper.

International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3, June 2017

A Novel Security Approach for Communication using IOT

Gowtham M
Assistant Professor
Department of Computer Science & Engineering
Rajeev Institute of Technology, Hassan
Karnataka, India

M. Ramakrishna
Professor & Head
Department of Computer Science & Engineering
Vemana Institute of Technology, Bangalore
Karnataka, India

***Abstract*—The Internet of Things (IOT) is the arrangement of physical articles or "things" introduced with equipment, programming, sensors, and framework accessibility, which enables these things to accumulate and exchange data. Here outlining security convention for the Internet of Things, and execution of this relating security convention on the inserted gadgets. This convention will cover the honesty of messages and verification of every customer by giving a productive confirmation component. By this venture the protected correspondence is executed on implanted gadgets.**

Keywords-Security; SSL; SN; IOT.

I. INTRODUCTION

The Internet of Things (IOT) is the arrangement of physical things or "things" embedded with devices, programming, sensors, and framework accessibility, which enables these articles to accumulate and exchange data. The Internet of Things empowers articles to be identified and controlled remotely across over existing framework system, making open entryways for more direct coordination between the physical world and PC based structures, and achieving improved reasonability, precision and cash related ideal position. Everything is strikingly identifiable through its embedded enlisting system however can interoperate inside the present Internet establishment. IOT is required to offer moved system of devices, structures, and organizations that goes past machine-to-machine correspondences (M2M) and spreads a collection of traditions, spaces, and applications. "Things," in the IOT sense, can suggest a wide variety of devices, for instance, heart checking embeds, biochip transponders on estate animals, electric mollusks in coastline waters, automobiles with characteristic sensors, or field operation devices that assistance firefighters in request and spare operations. These contraptions assemble important data with the help of various existing advances and after that independently stream the data between different gadgets.[5][6]

This paper addresses the security issue, by proposing a distributed security convention to fulfill this shifted condition. Secure correspondence is executed on a publicly released stage for the Internet of Things. Finally, the result exhibits that the proposed tradition is gainful to meet the specific goals and significant for the Internet of Things.

II. LITERATURE SURVEY

A. *"Examine on Security Problems and Key Technologies of The Internet of Things"*

Xu Xiaohui School of PC, Wuhan University School of money related perspectives and organization, Wuhan University Wuhan, China. 2013. The IOT is an enormous and comprehensively coursed the Internet that things interface things. It partners each one of the articles to the web through information recognizing contraptions. It is the second information wave after Computer, Internet and compact correspondence orchestrate. With the quick progression of the Internet of Things, its security issues have ended up being more idea. This paper addresses the security issues and key progressions in IOT. It clarified the fundamental thoughts and the rule of the IOT and merged the critical characteristics of the IOT and moreover the International essential research results to examination the security issues and key developments of the IOT which remembering the ultimate objective to assumes a positive part in the development and the advancement of the IOT through the exploration.

B. *"The Internet ofThings: A survey" Computer Networks, 2010"*

Atzori, Luigi; Iera, Antonio; Morabito, Giacomo, Vol.54 (15), pp.2787-2805 [Peer Reviewed Journal]. This paper addresses the Internet of Things. Fundamental empowering variable of this promising worldview is the reconciliation of a few advancements and interchanges arrangements. Distinguishing proof and following advances, wired and remote sensor and actuator systems, improved correspondence conventions (imparted to the Next Generation Internet), and appropriated insight for savvy items are quite recently the most applicable. As one can without much of a stretch envision, any genuine commitment to the progress of the Internet of Things should essentially be the consequence of synergetic exercises led in various fields of learning, for example, broadcast

communications, informatics, gadgets and sociology. In such a perplexing situation, this overview is coordinated to the individuals who need to approach this unpredictable train and add to its advancement. Distinctive dreams of this Internet of Things worldview are accounted for and empowering advancements surveyed. What develops is that still significant issues might be confronted by the examination group. The most important among them are tended to in points of interest.

C. *"Internet of Things Security Analysis" 2011*

Gan, Gang ; Lu, Zeyong ; Jiang, Jun, International Conference on Internet Technology andApplications, Aug. 2011, pp.1-4

Web of Things is an exceptional data and innovation industry, and in such a situation of the idea and the substance and the expansion of Internet of Things are not extremely unmistakable, the venture of Internet of Things which is a bit of little region with little scale and self-framework acquire satisfying accomplishment and brilliant future, it can advance the improvement of Internet of Things in some degree. In any case, there are some genuine shrouded risk and potential emergency issues. The paper concentrates on the use of Internet of Things in the country and even in the worldwide later on, breaking down the existed security dangers of the Internet of Thing 's arrange focuses, transmission, at long last we propose some suggestive arrangements because of these issues.

D. *"Securing IOT for Smart Home System"*

This paper demonstrates an approach to manage combine strong security in passing on Internet of Things (IoT) for astute home structure, together with due idea given to customer convenience in working the system. The IoT splendid home system continues running on common wifi arrange executed in view of the AllJoyn structure, utilizing a hilter kilter Elliptic Curve Cryptography to play out the verifications amid framework operation. A wifi passage is utilized as the inside hub of the framework to play out the framework introductory setup. It is then in charge of confirming the correspondence between the IoT gadgets and in addition giving an intend to the client to setup, get to and control the system through an Android based PDA running legitimate application program.

Security challenges in IOT join assurance, approval and secure end to end affiliation. Security and accommodation are the two noteworthy prerequisites for effective arrangement of IOT in the shrewd home framework in light of Wi-Fi network [6].

III. PROPOSED SYSTEM

This paper concentrates on the blueprint of a security tradition for the IOT, and the execution of this contrasting security tradition on the Sensible Things arrange. This tradition won't simply cover the respectability of messages, moreover the affirmation of each customer by giving a capable confirmation segment. It is an average stage for correspondence among sensors and actuators on an overall scale, and empowers an across the board expansion of IOT administrations. This safe correspondence gives a more proficient data transmission component contrasted and one TLS suit correspondence.

The proposed framework fills in as takes after.

1) The client registers and login through an android application.
2) Once the client enrolls the demand is sent to administrator to acknowledge or dismiss the demand. The enrollment procedure gathers information, for example, clients email ID, secret key and the IMEI number of the telephone from which the client is enlisting.
3) After the endorsement of demand from the administrator, the client can login and ask for utilizing the administrations.
4) Once the client asks for an administration the demand is sent to Authority hub, it at that point creates a declaration and sends it to both client and IOT server.
5) The authorization to utilize the administration is allowed as testament; on getting the endorsement the client can pick the vacancy to control the gadget and furthermore the rundown of administrations accessible.
6) The asked information from client is sent to IOT server alongside declaration. The IOT server assesses the authentication gotten from the client application and the expert hub in the event that it coordinates then it permits to control the gadget else the client need to rehash the whole technique.

IV. SECURITY PROTOCOL

The focal point of the P2P security system for the Internet of Things is the security tradition. This tradition is the base of all structures' correspondence and affirmation. There are two essential parts in this security tradition:

1) Registration
2) Communication

The enlistment procedure is completed between the as of late joined customer and the Specialist Node (SN). There are six sort of messages transmitted amid the primary enlistment handle.

1) Customer to SN: SSL association ask for message, which is to upgrade the security of the accompanying enrollment exchange.
2) Customer to SN: Enlistment ask for message.
3) SN to customer : Registration answer message.
4) Customer to SN: Authentication marking demand message.
5) SN to customer: Authentication marking reaction message.
6) Customer to SN : Verification tolerating message. The detail of these messages could be seen from underneath,

without the foremost SSL affiliation requests message. A mid the second procedure, correspondence process depends on the main procedure. There are at most five sorts of message, including authentication trade ask for message, declaration trade answer message, session key trade ask for message, session key trade answer message and secure message. More often than not, just secure message is utilized for data transmission.

V. MODULES

In this segment we plate about the modules exhibit in the venture and there depiction

1) Client Authentication
2) Benefit Request
3) Authentication Generation utilizing SHA
4) Authentication Distribution by IOT server
5) Session Management and Request handler
6) Installed C Programming to control equipment
7) Home PC
8) Joining

1) Client Authentication

In this module the clients downloads the android application and registers by giving the required points of interest such name email-id secret word, the application naturally gets the IMEI number of the enlisting portable number all these data will be put away in database. Once the client enlists, the demand is sent to administrator to endorse or dismiss, once the client login it sidetracks to landing page.

2) Benefit Request

After the client login he is diverted to landing page where he motivates choices to start the administration a testament from IOT server will be produced and sends a duplicate to both client application and home PC and a session of 120 seconds will be made for client, he needs to control the switches inside the session lapses else he have to login again and ask for administration and get another session.

3) Authentication Generation utilizing SHA

This model encourages the client to interface with IOT server by issuing testament as an authorization granter. Once the client enroll and ask for administration the AN of IOT server gets client points of interest, for example, IMEI number and an extraordinary ID which is naturally produced to recognize the client. At that point the An utilizations a mix of this two element and creates a remarkable ID, this procedure is finished with the assistance of Hashing calculation.

4) Authentication Distribution by IOT server

The Authority node and the IOT server integrated together to form this module. Once the user request to control the device, the request is forwarded to IOT server, the IOT server generates a certificate with unique ID to allot the session to the user and sends a copy of certificate to both user and home PC.

5) Session Management and Request handler

Once the client ask for administration a session of 120 seconds is made for the client. Each time the client controls the change he leaves controller page and each time the session of 120 seconds terminates the client leaves application, he needs to login again and ask for administration and get a session.

The ask for handler gives data about the status of session. On the off chance that the session is dealt with effectively then it advises the achievement status. In the event that the session flops then a message is shown expressing "The home PC is as of now bustling attempt later".

6) Installed C Programming to control equipment

The equipment part contains an implanted IC circuit or a raspberry Pi associated with 4 switches which empowers the client to interface numerous gadgets to control. The controlling part is coded in C programming dialect, the primary capacity of this equipment is, on accepting the flag from IOT server to control the switch, i.e. it should on or off the gadget.

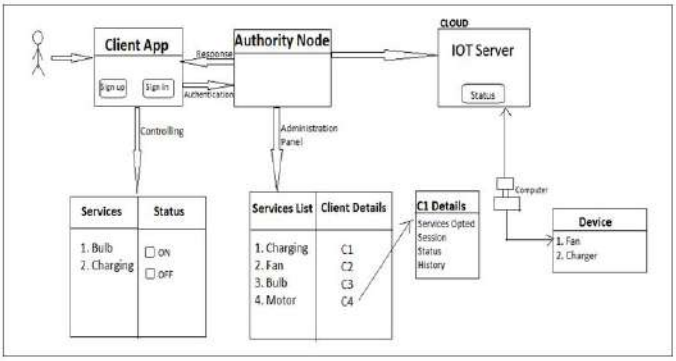
7) Home PC

The home PC is associated with the equipment. The serial to USB converter equipment requires drivers that enable the program to speak with equipment. The home PC module gets the authentication from IOT server and sits tight for client order. On getting client summon the equipment controls the switches and the status of switch is shown in the home PC screen. It additionally shows all the data in regards to association. The reset homes PC catch clears all the unexecuted demands from client, we utilize this when the application ends anomalous or when control disappointment happens. On the off chance that we don't reset the PC then the framework tries to execute the old demand whose session has been lapsed.

8) Joining

In this stage each module is consolidated together and checked if the framework is acting of course with no mistakes in any module after coordination of modules.

VI. ARCHITECTURE



A. *Customer(Client App)*

Customer/client needs to enroll to the Authority hub by an Android IOT application. Customer needs to have Android application through which he can control his gadgets remotely.

These gadgets are associated with PC which is introduced in customer/client home. PC is associated with Internet of Things server.

Customer application gives:

1) Sign up

2) Sign in

When customer is enlisted to AN, it is recognized with declaration (Say one of a kind customer id). Interesting ID alludes to a customer, that he is confirmed to control or screen the gadget.

1) User will get rundown of administrations.

2) Selection of administrations.

3) He can check his administrations status before or in the wake of observing.

B. Specialist NODE(Authority Node)

1) He keeps up all points of interest of all enlisted/validated customers. Points of interest, for example,

a. Client data (extraordinary id, administrations offered, session).
b. Currently working clients.
c. Client history (time of login, administrations picked, operations done as such for with timings).

2) AN is associated with IOT server and it keeps up
a. Clients benefit status.
b. Status of PC introduced in home.

C. Home Personal Computer

1) Personal PC with IOT programming is associated with IOT server which will continue checking the status of administrations. On the off chance that any adjustment in status PC performs as indicated by it.

2) Devices are associated with PC which is in USB to serial correspondence.

3) Personal PC is associated with equipment unit to which every one of the gadgets can be associated.

VII. Result Analysis

A few Validation checks are performed for "Verified remote exchanging utilizing IOT" application by separating it into segments. Every part has been tried and all test outcomes are working effectively.

The modules in the framework connect as for the determinations and they can be coordinated to create the coveted yield.

It is discovered that the framework fulfills every one of the necessities indicated and delivers the yield with no blunder.

VIII. Conclusion

This paper gives an effective security convention to the installed frameworks or stages separating the Internet of Things. Its effectiveness has been fundamentally made strides. As a working research facility model, comparable secure correspondence can likewise be worked for different frameworks. Outfitted with this assurance, individuals' security could be very much ensured in the Internet of Things. This in like manner advances the change of the Internet of Things.

References

[1] Xu Xiaohui School of computer, Wuhan University School of economics and management, “Study on Security Problems and Key Technologies of The Internet of Things” Wuhan University Wuhan, China. 2013.

[2] Atzori, Luigi; Iera, Antonio; Morabito, Giacomo, “The Internet ofThings: A survey” Computer Networks, 2010, Vol.54 (15), pp.2787-2805 [Peer Reviewed Journal]

[3] Gan, Gang ; Lu, Zeyong ; Jiang, Jun, “Internet of Things Security Analysis” 2011 International Conference on Internet Technology andApplications, Aug. 2011, pp.1-4

[4] Suo, Hui ; Wan, Jiafu ; Zou, Caifeng ; Liu, Jianqi, “Security in the Internet of Things: A Review” 2012 International Conference onComputer Science and Electronics Engineering, March 2012, Vol.3, pp.648-651

[5] Zou, Caifeng, Lu, Zeyong, Morabito, Giacomo, “Access control for IOT devices home automation, of computer science and electronic engineering, jan 2014

[6] Freddy K Santoso, and Nicholas C H VunSchool, “Securing IOT for Smart Home System” of Computer Engineering,Nanyang Technological University, Singapore. 2015

AUTHORS’ PROFILE

Mr. Gowtham M received his B.E. degree in Information Science and Engineering from Kalpataru Institute of Technology, Tiptur. He received M.Tech. degree in Computer Networks and Engineering from National Institute of Engineering, Mysore. He is currently working as Assistant Professor, Department Of CSE at Rajeev Institute of Technology, Hassan, Karnataka, India. His area of interest is Computer Networks, Cloud Computing.

Dr. M Ramakrishna received his B.E. Computer Science and Engineering from Periyar University, Tamilnadu. He received M.E. degree in Computer Science and Engineering from Anna University, Tamilnadu. He received Ph.D. in Computer Science and Engineering from Anna University, Tamilnadu. He is currently working as Professor & Head, Department Of CSE at Vemana Institute of Technology, Bangalore, Karnataka, India. His area of interest is Computer Networks.

International Journal of Engineering and Applied Computer Science (IJEACS)
Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3, June 2017

Secure De-Duplication in Cloud Computing Environment by Managing Ownership Dynamically

Shashikala M.K
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

Dhruva M.S
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

***Abstract*—As the arena movements to superior ability for recorded purposes, the usage of Cloud Service Providers (CSPs) are increasing more ubiquity. With the steady increment of disbursed garage adopters, records de-duplication has become a want for cloud providers. In distributed storage administrations statistics de-duplication is one of crucial structures to lessen the distance requirements of administrations by way of casting off replica duplicates of rehashing facts and setting away unmarried replica of them. But it leads to security troubles when specific clients deliver indistinguishable records to the allotted storage. As of past due, a few de-duplication plans were proposed to take care of this problem. However the general public of the plans enjoy the unwell effects of protection issues, given that they don't take into account the dynamic changes inside the obligation for statistics. In this paper, a unique server-facet de-duplication plot is proposed for combined statistics that utilizations RCE and gathering key management machine. The proposed conspire ensures that one-of-a-kind authorized access to the commonplace facts is conceivable. The security of the de-duplication schemes is furnished by means of making use of suitable encryption schemes.**

Keywords-Deduplication; Cloud Providers; Encryption; Security

I. INTRODUCTION

Cloud computing lets in get entry to unlimited virtualized sources from everywhere and at any time thru the net. Now-a-days the quick development of facts volumes positioned away within the disbursed storage has brought about an elevated interest for strategies for sparing plate space and machine facts transmission. Cloud professional co-ops, for instance, Dropbox[6], Wuala[7], Mozy[8], and Google Drive[9] always search for systems meant to restrict excess data and enlarge space reserve budget. A survey found out that best 25% of the information in facts warehouses are particular. We recognition on de-duplication, that's a specialized technique to save the digital space by way of casting off multiple copies of identical records in garage. Rather than retaining various facts duplicates with a comparable substance, de-duplication distinguishes repetition in facts and later on kills excess records by means of maintaining only a single bodily replica and alluding other extra records to that duplicate.

Despite the fact that records de-duplication brings a remarkable deal of advantages, however safety and records privacy is as yet sensitive problems. Customary encryption is incongruent with statistics de-duplication. Encryption of the indistinguishable statistics duplicates of numerous clients with numerous encryption keys will set off one of a kind parent writings, which make de-duplication unattainable. The satisfactory execution of standard encryption might define be capable of as takes after: Consider clients An and B, scrambles the file M beneath their mystery keys SA and SB and stores their referring to ciphertext CA and CB. At that factor, additionally problems emerge: (1) by using what way can the cloud server experience that the vital record M is nearly identical, and (2) irrespective of the opportunity that it can distinguish this, how would possibly it empower each clients to recoup the information, in view in their exceptional mystery keys? One critical way out is to empower on every patron to encode the record with trendy key of the disbursed garage server. By then, the server can de-duplicate the distinguished development by way of unscrambling it with its personal key combine. Still, this association allows the cloud server to get the outsourced simple statistics, which may separate the security of the records if the cloud server cannot be definitely depended on. Hash table calculation defeats the drawback which examined previously. It encodes/decodes facts replica with a joined key, that's gotten through registering the cryptographic hash estimation of the substance of the information duplicate. At that factor, customers maintain the keys and give the ciphertext to the cloud server. Since the joined encryption is deterministic, indistinguishable files will create a comparable parent content.

Suppliers of cloud-primarily based storage, for instance, Google power would save be able to on restriction costs via de-duplication: should clients exchange a similar file, the management perceives this and stores only unmarried replica. Concurrent encryption has been proposed to execute information safety whilst making de-duplication manageable. It encodes/unscrambles facts replica with a united key, that's gotten via registering the cryptographic hash estimation of the substance of the facts replica. After key generation and records encryption, customers maintain the keys and ship the determine content to the cloud.

II. LITERATURE SURVEY

Literature review is the process of presenting the summary of the conference papers and journal articles study resources. The information de-duplication conspires over encoded information has been produced and enhanced further into Convergent Encryption (CE), Leakage-Resilient (LR) De-duplication plot, Randomized Convergent Encryption (RCE) and Dynamic Ownership Management Scheme.

A. Merged Encryption (CE)

LI[1] With a particular true objective to keep information security against inside cloud server and furthermore outside challengers, clients may require their data encoded. Be that as it may, customary encryption under various clients' keys makes cross-customer de-duplication incomprehensible, since the cloud server would constantly watch unmistakable ciphertexts, regardless of the possibility that the information are the same, paying little respect to whether the encryption calculation is deterministic. Douceur [2] presents Convergent Encryption, which is the promising answer for this issue. In CE, an information proprietor gathers an encryption key over information by using cryptographic hash work. At that point registers the ciphertext utilizing square figure over information alongside their encryption key. CE deletes information and keeps just encryption key subsequent to transferring ciphertext to the distributed storage. Since encryption is deterministic, on receipt of same record CE delivers same ciphertext for it and the server does not store the record yet rather refreshes meta-information to show it has an extra proprietor.

Benefits: Provides promising arrangement over ordinary encryption and jelly information security.

Restrictions: Convergent Encryption experiences some security issues i.e. label consistency issue. It implies that honesty and security of information has been traded off because of the absence of PoW process and dynamic possession administration.

B. Ramp Secret Sharing Scheme (RSSS)

Li [3] formalizes a joined key administration plot i.e. Dekey which is proficient and solid for secure de-duplication. Dekey set de-duplication between united keys and circulates those keys over various key servers while saving the semantic security of joined keys and protection of outsourced data. Dekey is actualized utilizing the Ramp mystery sharing plan. Dekey utilizes RSSS to gather concurrent keys. It's thought is to allow de-duplication in joined keys and circulate the merged keys over different KM-CSPs. Rather than scrambling the joined keys on a for every client premise, Dekey manufactures mystery shares on the first merged keys (that are in plain) and appoints the offers over different KM-CSPs.

Benefits: Provides dependable, proficient and adaptation to internal failure united key component for secure de-duplication.

Constraint: This plan does not bolster dynamic proprietorship administration issue in secure de-duplication.

C. Spillage Resilient (LR) De-duplication Scheme

Xu[4] proposes a spillage versatile de-duplication plan to unravel the information respectability issue. It tended to a key security worry in cross-client customer side de-duplication of encoded documents in the distributed storage: protection of clients' delicate records against both outside challengers and the honestbut-inquisitive distributed storage server in the limited spillage display

Benefits: Resolves information trustworthiness issue i.e. anticipates label consistency assault.

Restrictions: Secure verification of possession (PoW) plot in the standard model remains an open issue. Another downside is the absence of dynamic possession administration among the information holders.

D. Approved De-duplication Hybrid Cloud

LI [5] proposes an approved de-duplication conspire where differential benefits of clients, and in addition the information, are considered in the de-duplication strategy in a cross breed cloud condition. He exhibited a few new de-duplication developments supporting approved copy check in cross breed cloud design, in which the copy check tokens of documents are created by the private cloud server with private keys. The figure demonstrates the engineering of approved de-duplication.

Benefits: This plan gives approved de-duplication over half and half cloud for clients who have distinctive benefits.

Restrictions: Data spillage.

Far site – a de-duplication framework that spotlights less on approval of clients. Information on the cloud is put away in typical frame. De-duplication is performed just on content and pictures and it does not bolster all document formats [10]. In 2013, Neal Leavitte[4] examined the diverse issues landing in the de-duplication in multi-inhabitant condition. Distinctive creators proposed the utilization of the single key encryption.

III. EXISTING SYSTEM

In existing framework, Cryptographic methods were connected to get to control for remote stockpiling frameworks. The information proprietors scramble records by utilizing the symmetric encryption approach with content keys and after that utilization each client's open key to encode the substance keys. It requires every information proprietor to be online constantly. A few strategies convey the key administration and circulation from the information proprietors to the remote server under the suspicion that the server is trusted or semi-trusted.

Disadvantages

- The key management is very complicated when there are a large number of data owners and users in the system.
- The key distribution is not convenient in the situation of user uses the system dynamically.

- The user needs to know private key.
- Less protect security.

These de-duplication systems cannot support differential authorization duplication check.

IV. PROPOSED SYSTEM

We propose a de-duplication plot over scrambled information. The proposed conspire depends on Elliptic Curve Cryptography and declines the key length while giving securities at an indistinguishable level from that of different cryptosystems gives. The proposed plan ensures that solitary affirmed access to the regular data is possible. We give the unusual state security and avoid the replication of archive in the cloud expert association. To secure the assurance of tricky data while supporting de-duplication, the centered encryption methodology has been proposed to encode the data before outsourcing .In proposed framework, we are utilizing hash capacity to produce key for the record. By utilizing hash capacity to keep away from the duplication in cloud. After that we are applying cryptographic system for security reason. We are utilizing ECC calculation for encryption and decoding process. The proposed scheme has the going with purposes of enthusiasm in regards to security and profitability:

To begin with, dynamic proprietorship organization guarantees the retrogressive and forward secret of de-duplicated data upon any ownership change. Second, the proposed plot ensures security in the setting of PoW by showing a re-encryption framework that uses an additional social event key for dynamic ownership gathering.

A. *Secure Data De-duplication Architecture*

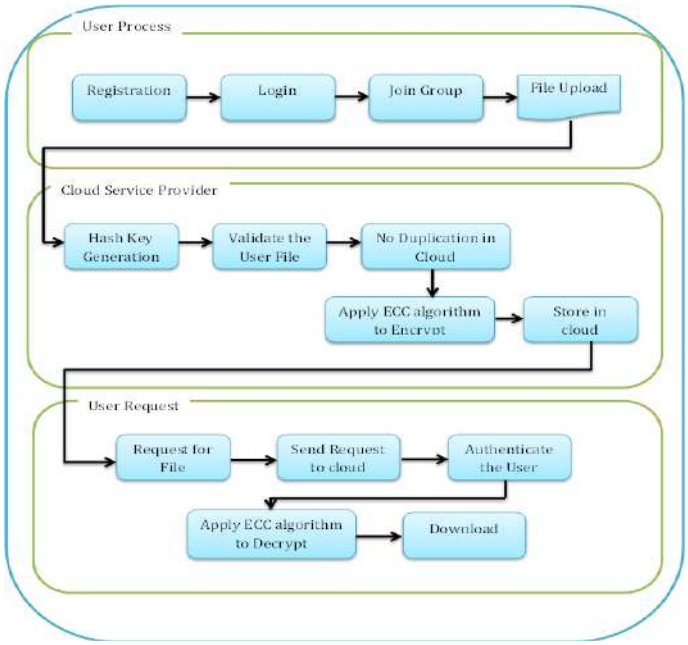


Figure 1

B. *Sequence Diagram*

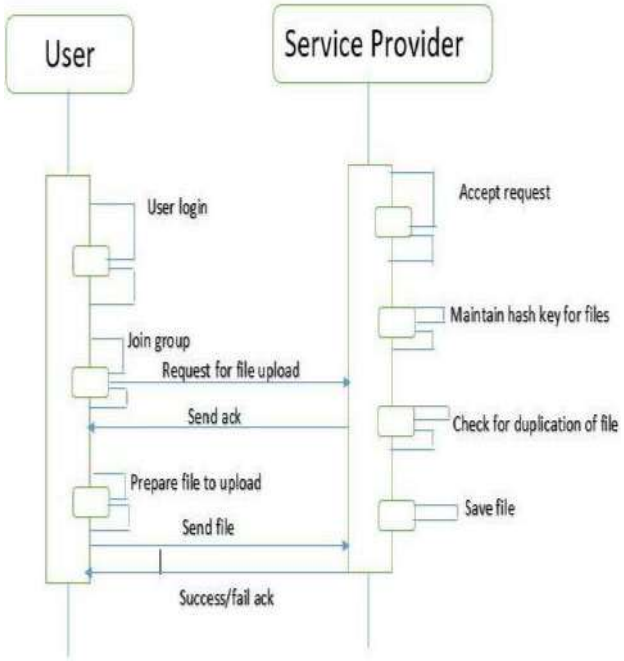


Figure 2.

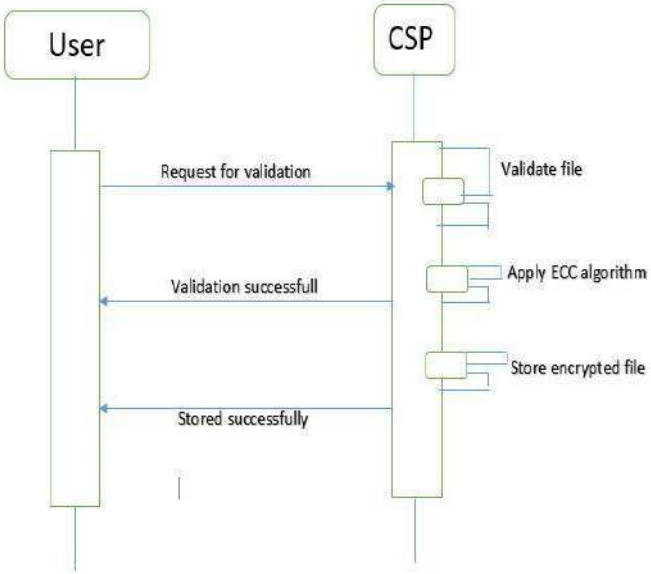


Figure 3. User interaction with Service Provider

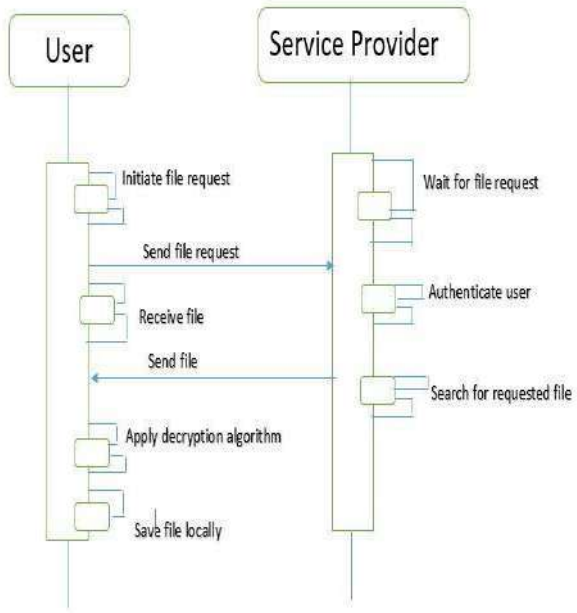


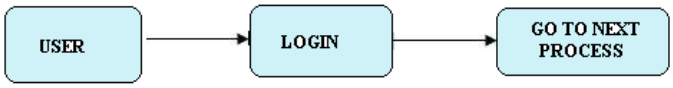
Figure 4. User interaction with Service Provider

V. METHODOLOGY

MODULES:

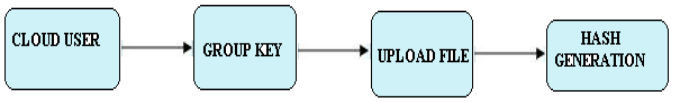
1) User Registration and Login

In this Module If he is a new user he needs to enter the required data to register the form by providing the user details like name; dob etc. and the data will be stored in server for future authentication purpose. After registration user will get the username and password for further process. Using Username and Password, user login into Group. Group generate key for the valid user and process inside the group under the valid key.



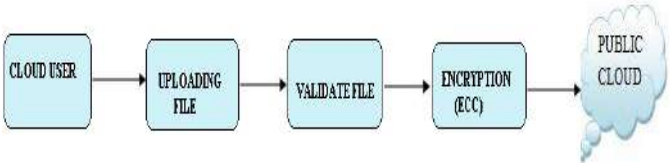
2) User Joining The Group and File Upload

For every user a key would be generated using which the user gets the authentication to join the group with a key. In file upload process, user choose the file from the system and generate hash key for each file. Hash key generation is provided to avoid duplication of the file to the cloud. If the file is already in the cloud the user cannot upload the file.



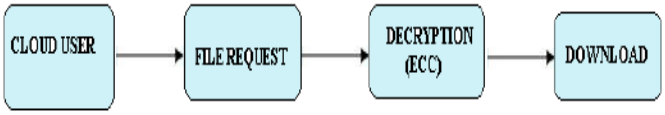
3) File Encryption and Storage in Cloud

If data duplication check is negative, the data user encrypts its data using ECC algorithm in order to ensure the security and privacy of data, and stores the encrypted data at CSP. We implement ECC algorithm which converts a file in to a binary format and it gets encrypted and is stored on to the cloud. The data that is stored on to the cloud will be in encrypted format.



4) User File Request and Download

Any user who has registered earlier and joined the group with a valid key can request the file to the cloud. The cloud service provider after authenticating the user can receive the file request, decrypt the file using ECC algorithm and send the requested file to the user. Then the file will be downloaded in the user's location.



VI. RESULTS

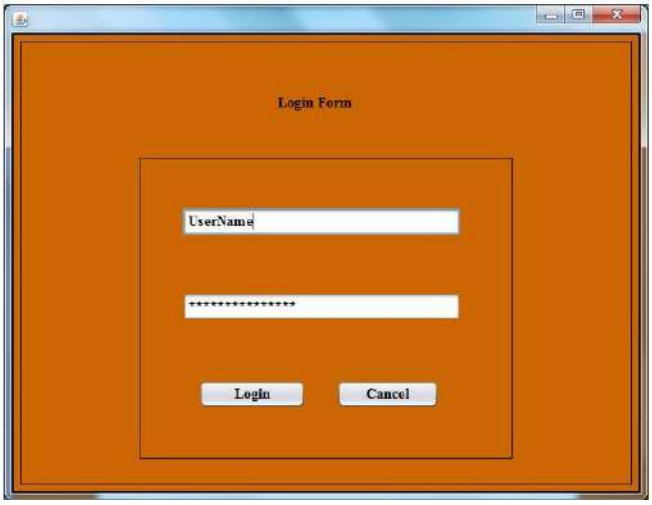


Figure 5. Login Page

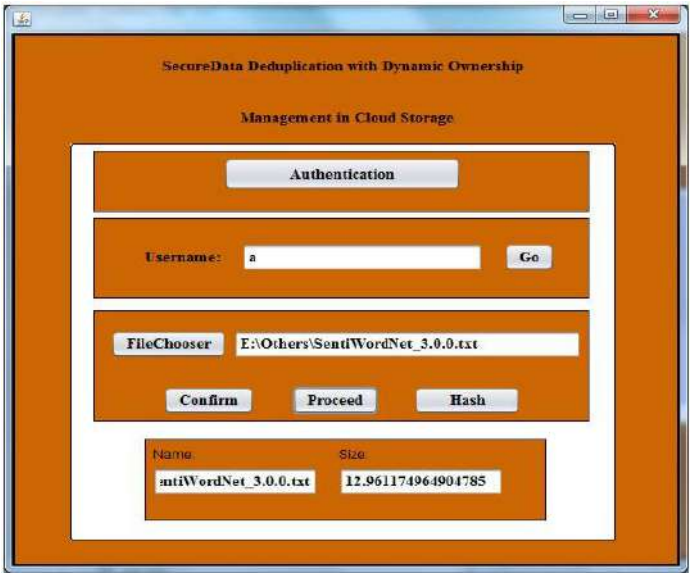


Figure 6. Uploading the file

Figure 7. Generating hash key and verifying file

Figure 8. Checking Duplication of File

VII. CONCLUSIONS

Managing encrypted facts with de-duplication may be very critical and important in examine for achieving a prospering cloud storage provider, mainly for massive data storage. . In this paper, we proposed a realistic scheme to control the encrypted big facts in cloud with de-duplication based on possession venture. This plan enhancements facts safety and privacy in disbursed garage towards any clients who don't have good sized responsibility for data, and also in opposition to an actual however inquisitive cloud server. By utilizing hash key successfully can accomplish de-duplication in dispensed storage. Elliptic curve cryptography is applied for encryption and decryption method to decrease probability of attacking the document.

REFERENCES

[1] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. Hassan, and A. Alelaiwi, "Secure Distributed Deduplication Systems withImproved Reliability," IEEE Transactions on Computer, Vol. 64, No. 2, pp. 3569–3579, 2015

[2] R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, "Reclaiming space from duplicate files in a server less distributed file system," Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.

[3] Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 6, 2014.

[4] Xu, E. Chang, and J. Zhou, "Leakage-resilient client-side deduplication of encrypted data in cloud storage."

[5] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.

[6] Dropbox, http://www.dropbox.com/.

[7] Wuala, http://www.wuala.com/.

[8] Mozy, http://www.mozy.com/.

[9] Google Drive, http://drive.google.com

[10] Rev De dup: A reverse deduplication storage system optimized for reads to latest backups 2013.

AUTHORS' PROFILE

Shashikala M. K completed her B.E. degree and M. Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India Currently she is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. Her areas of interest include networks, algorithms.

Dhruva M.S. completed his B.E. degree and M.Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India. Currently he is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. His areas of interest include multimedia networks, Compiler Design and Algorithms.

Call for Papers

International Journal of Engineering and Applied Computer Science (IJEACS) invites authors to submit their manuscripts for categories like research papers, review articles, survey papers, technical reports , report of unsuccessful research projects , case studies , tutorials , book reviews, short communications and cross talk, extended version of conference papers for publication in our monthly issue.

IJEACS reviews manuscripts through double blind peer review process. Authors can submit their original, unpublished manuscript which is not under consideration for publication in any journal or conference through email. IJEACS publish papers in major streams of Engineering and Computer Science include following but not limited to.

Computer Science

- Software Design and Modeling
- Service Oriented Architecture
- Open Source Software
- Software Testing and Maintenance
- Software Measurement and Reliability
- Knowledge based Systems
- Image Processing and Computer Graphics
- Extreme Programming and Formal Methods
- Artificial Intelligence, Image Recognition and Bio metrics.
- Machine Learning and Computer Vision
- Algorithm Analysis and Design
- Computational Mathematics
- Data Structures and Graph Theory
- Video Coding and Data Compression
- Database Systems and Big Data Analytics
- Internet of Things, Architecture and Computing
- Parallel and Distributed Computing
- Cloud Computing, Agent Oriented System

- Communication Network and Systems
- Embedded Systems and Applications
- Cryptography and Information Assurance
- Computational Biology and Bioinformatics
- Human Computer Interaction
- Natural Language Processing

Engineering

- Micro Processor Architecture and Design
- VLSI/IC Microelectronics and Computer Design
- Parallel Processing and Digital Systems
- Wireless Transmission and Mobile Communication
- Antenna and Wave Propagation
- Semiconductors, Circuits and Signal System
- Material Science and Metallurgy
- Machine and System Design
- Robotics and Automated Control
- Manufacturing Processes and CAD/CAM
- Quality Control and Assurance
- Digital Signal Processing
- Photonics, Fiber Optics and Optical Communication
- Biosensors, Electrical-based Interactions
- Nano electronic Devices and Silicon Micro/Nano Electronics Applications
- Distributed monitoring systems and Smart Systems
- Fluid Dynamics and Implementations
- Mechanics and Vibration
- Heat Transfer
- Combustion Engines and Automobiles
- Health Instrumentations and Technologies
- Thermal Engineering
- Solar Power Systems
- Ergonomics

Submit your manuscript to: **submit@ijeacs.com** **or** **ed.manager@ijeacs.com**

Innovations continue to serve the humanity

International Journal of Engineering and Applied Computer Science

Volume: 02, Issue: 06, June 2017

ISBN: 978-0-9957075-7-3