

ACADEMIA

Accelerating the world's research.

IC-CSOD -2018 Conference Proceedings

M E PURUSHOTHTHAMAN NAIDU

Related papers

[Download a PDF Pack](#) of the best related papers ↗



[005 MOBILEINNOVATIVECOMPUTING](#)

Rohini Kadlag

[Journal of Computer Science IJCSIS February 2016](#)

[Journal of Computer Science IJCSIS](#)

[Journal of Computer Science and Information Security March 2012](#)

[Journal of Computer Science IJCSIS, Jagruti Save, Kadar Shereef](#)



**ISO 9001:2015
Certified**

**NBA
Accredited**

**NAAC
Accredited**

www.facebook.com/mcw.tcet

www.twitter.com/multiconw



Proceedings of
**International Conference on
Communication, Security and Optimization
of Decision Support Systems**

Organized by,
Department of Information Technology
February 23rd & 24th 2018, Friday & Saturday

Compiled & Edited By:

Dr. Rajesh Bansode

Mr. Vikas Kaul

Dr. Radhika Kotecha

Ms. Vandana Munde

Ms. Pranjali Kasture

Ms. Purvi Sankhe

Ms. Sangeeta Vhatkar

Ms. Nishta Mathur



Zagdu Singh Charitable Trust's (Regd.)

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai*)

Institute Accredited by National Assessment and Accreditation Council (NAAC), Bangalore#

• ISO 9001 : 2015 Certified • Accredited Programmes by National Board of Accreditation, New Delhi**

*Permanent Affiliated UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (w.e.f. A.Y. 2015-16)
• Electronics Engineering (w.e.f. A.Y. 2017-18)

**1st time Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (3 years w.e.f. 16-09-2011)

**2nd time Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology • Electronics Engineering (3 years w.e.f. 01-07-2016)

1st cycle of Accreditation : • "A" Grade for 5 years (w.e.f. 30-10-2017)



ISO 9001:2015
Certified

NBA
Accredited

NAAC
Accredited

IC-CSOD - 2018 Conference Proceedings



MULTICON-W 2018

In Association with





In Association with



"International Conference on Communication, Security and Optimization of Decision Support Systems(IC-CSOD 2018)"

Chief Patron

Mr. V. K. Singh, Chairman, TEG

Patron

Mr. Jitendra R. Singh, Trustee, TEG

Mr. Karan V. Singh, CEO, TCET, TIMSR, TIMSCDR

Program Chair

Dr. B. K. Mishra, Principal, TCET

Advisory Committee

Dr. S. N. Merchant, IIT, Bombay.

Dr. Dhiren Patel, Director, VJTI Mumbai,

Dr. Weifa Liang, Emeritus Professor, Australia

Dr. Prakash D. Vyavahare, Professor, S. G. S. I.T.S, INDORE,

Mr. Saurabh Srivastava, Assistant Vice President, Reliance Jio Info comm Ltd

Dr. Manoj Kumar Kowar, Ex-Director, B.I.T, Chhattisgarh, India

Technical Program Committee

Witold Pedrycz, Research Chair, Canada

Dharma P. Agrawal, University of Cincinnati, Cincinnati, USA

B. P. Patil, Army Institute of Technology, Pune, India

G. Sinha, Professor, CMR Technical Campus, Hyderabad, India

K. Rajesh, Head, High-Performance Computing, BARC, India

Program Co- Chair (IC-CSOD-2018)

Dr. Kamal Shah, Professor, IT Dept.

Dr. Deven Shah, Professor, IT Dept.

Convenor

Dr. Rajesh S. Bansode, Professor & HOD- I.T. Dept

Co-Convenor

Mr. Vikas Kaul, Assistant Professor, IT Dept.

Ms. Sangeeta Vhatkar, Assistant Professor, IT Dept.

Dr. Radika Kotecha Assistant Professor, IT Dept

Treasurer

Ms. Mary Margarate, Assistant Professor, IT Dept.

Publication Committee

Ms. Vandana Munde, Assistant Professor, IT Dept.

Ms. Pranjali Kasture, Assistant Professor, IT Dept.

Ms. Purvi Sankhe, Assistant Professor, IT Dept.

Thakur College of Engineering & Technology, Mumbai, India

Copyright©2018 by the International TCET. Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page, Copyrights for components of this work owned by others than above mentioned bodies must be honored, Abstracting with credit is permitted.

To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and or a fee. Request permission to republish from : Publications Depat. TCET. For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center.

Notice to Past Authors of Multicon-W Published Articles

TCET intends to create a complete electronic archive of all articles and or other material previously published by TCET, if you have written a work that was previously published by TCET in any journal or conference proceedings prior to 2015, and you do NOT want this work to appear in the TCET digital Library, please inform multiconw@thakureducation.org, stating the Title of the work, the author(S), and where and when published.

ISBN Numbers

ISBN : 978-0-9994483-1-1

Printed in India

Thakur College of Engineering & Technology, Mumbai India

Preface

The department was established in the year 2001 with the objective to support quality technical education to engineers in field of IT. Till current year 13 batches have passed out with excellent placement and academic record. To continue the journey of excellence in research the department has been organizing International Conference since 2013 with the objective of providing a common platform to researchers from various domains, departments to share their contributions under the umbrella of Multicon-W. The conferences held in the past have developed the experienced faculties, researchers and industry experts to learn and share the mechanism to improve on critical thinking, problem solving and innovative skill to face the future challenges. The domains at the department has also gained enough technical strength due to the conduct of conference each year. The primary objectives are knowledge sharing, promoting research & upcoming technologies, understanding future trends and challenges. This endeavor provides direction to technical experts to re-orient future technical educational programmes to meet the global demands. It also supports experts to share and understand the life-long learning with professional and social values.

This International Conference Proceedings volume contains the versions of contributions presented during the 6th International Conferences on Communication, Security and Optimization Decision Support Systems. In the previous years the conference provided a setting for discussing recent developments in a wide variety of topics including Communication, Security, Cloud Computing, Data Compression. The Conference has been a good opportunity for authors contributing from Chile, Tunisia, China, Iran & USA to share & present their topics in their respective research areas. We would like to thank all participants for their contributions to the Conference program and for their contributions to these Proceedings. The technical papers received from participants were from throughout India and 11 International papers from abroad with 03 industry invited papers to our conference, made it truly international in scope. The conference has received around 135 papers that are to be presented during the conference and provided ample opportunity for deliberations & discussion. The knowledge shared through conference provides a ISBN no. for the conference proceedings paper making them easily discoverable, readable and citable. Most regional conferences go unpublished and have limited durability beyond the closing ceremony and are routinely overlooked by international databases rendering them invisible to the world. The conference is a platform that enables authors to benefit from a range of services saving time and raising the impact and durability of your conference.

We are looking forward to the 10th International Conference on Communication, Security and Optimization of Decision Support Systems that will be held during last week of Feb. 22nd -23rd, 2019 at the same location. We hope that it will be an interesting and knowledge gathering platform to all authors.

Last but not the least I wish to thank all who directly or indirectly involved in the compilation of conference proceedings.

Dr. Rajesh S. Bansode

Convener IC-CSOD, 2018

CONTENTS

EMERGING TRENDS IN COMMUNICATION AND NETWORKING

1.	Document Summarization As A Service (Dsaas): A Web Application For Summarizing Technical Documents <i>Zainab Mirza, Abdul Samea, Khan Shaheer, Abuzer Malvi</i>	1-4
2.	Work-Stake Hybrid: A Trustless Consensus Mechanism for Blockchain <i>Yash Chetnani, Nupur Giri</i>	5-8
3.	Smart Surveillance using Internet of Things <i>Bhushan Nemade, Deekshi Chavan</i>	9-12
4.	Event Management System <i>Rishabh Dubey, Pranay Rao, Ankit Gupta, Amit Singh</i>	13-16
5.	3-Tier Website as a SaaS Model <i>Ruth Baptista, Deven Shah</i>	17-22
6.	Comparative Study of Project Management Software <i>Vandana Munde, Vijaya Yele</i>	23-26
7.	Android Based Voting Application <i>Mary Margarat, Aanchal Singh, Rupesh Vishwakarma, Abhishek Singh</i>	27-30
8.	E-Commerce Android Application for Solid Waste Management <i>Alireza Chaviwala, Huney Pandita, Vequar Ansari, Nazneen Pendhari</i>	31-34
9.	BlockChain Technology: Future Aspects <i>Hetal Amrutia, Anil Vasoya, Kamal Shah</i>	35-38
10.	IoT & Cloud based vehicle toll collection system <i>Neha Kapadia</i>	39-42
11.	A proposal for monitoring system for secure health using IoT <i>Vinesh Bagul, Rajesh Bansode</i>	43-48
12.	Wireless sensor network for environmental monitoring and Control <i>Pankaj Malviya, Kautilya Joshi, Deep Modi, Zahir Aalam</i>	49-50
13.	Design of Guest Tagging System and Surveillance using RFID <i>Satyendra Yadav, Akash Singh, Lokesh Ucheniya , Deven Shah</i>	53-58
14.	To Analyze Student's Learning Experience in social media <i>Swathi Yadav, Kamal Shah</i>	59-62

15.	To study and implement IEEE 802.11g security aspects using index policy method <i>Yamisha Savalia, Ashwini Nakil, Jasmina Ambani, Rajesh Bansode</i>	63-67
16.	Livestock Management using IOT <i>Veena Kamat, Nemish Gandhi, Pooja Mistry, Namdeo Badhe</i>	68-70
17.	Digitalizing health care system using aadhar card <i>Seema Jamal, Sangeeta Vhatkar</i>	71-74
18.	Timing Side Channel Attack Implementation and countermeasure for an AES-ECC hybrid system <i>Nishtha Mathur, Purvi Sankhe, Shruti Mathur, Neha Patwari, Ratna Nayak</i>	75-79
19.	Accessing Operating System Using Finger Gesture <i>Parth Parikh, Bijith Marakarkandy</i>	80-82
20.	Health Monitoring System <i>Mary Margarat Valentine Neela</i>	83-85
21.	A proposal for smart health care centre using IoT for Indian scenario <i>Rina Yadav, Rajesh Bansode</i>	86-90
22.	Raspberry Pi based Cost effective interactive device for Blind, Visually Impaired, Deaf and Dumb People <i>Shubham Maheshwari, Vijay Karotra, Tanishq Mehra, Shridhar Kamble</i>	91-95
23.	A Comparative Study of Load Balancing Using SDN and Traditional System <i>Vikas Kukreja, Dinesh Panchi, Jatin Sajnani, Hitesh Seedani, Nupur Giri</i>	96-98
24.	Web Application on Result Analysis <i>Kirtan Patel, Maulik Shah, Rishab Tiwari, Ashish Soni</i>	99-103
25.	Digital India and its Impact <i>Sandeep Yadav, Sangeeta Vhatkar</i>	104-108
26.	Efficient and Secure Data Dynamic Operations in Cloud Computing <i>Shital More, Swati Abhang, Sudhir Dekhane, Shridhar Kamble</i>	109-116
27.	Android App for Renting Daily Use Items <i>Harsh Parekh, Hetal Amrutia, Kartik Iyer, Harsh Doshi</i>	117-121
28.	Smart Mirror using IoT <i>Charvika Banjan, Saurabh Khandelwal, Vamil Gandhi, Anil Vasoya</i>	122-126

RESEARCH CHALLENGES IN SECURITY

29	Video Processing in Visual system for color detection for people with Tritanomaly <i>Nazneen Pendhari,' Raghavendra Sedamkar</i>	127-130
30	NOSQL DB with Python PaaS <i>Mirza Zainab, V.S. Dhaka</i>	131-134
31	Real-Time Location Detection using Face Recognition through CCTV cameras/Mobile camera <i>Huzaifa Saboowala, Mustafa Poonawala, Mufaddal Kapadia, Farhana Siddiqui</i>	135-139
32	<i>Analysis of Real Time Surveillance System on Hadoop Image Processing Interface</i> <i>Mohammed Ahmed Abdul Mannan, Shamuwel Ahmed Ansari, Shehbaz Ansari, Sayed Adnan Husain</i>	140-143
33	Agricultural Updates via SMS using Cloud Computing Approach <i>Rakesh Shirsath, Mandar Shinde, Mayank Mehta, Yogesh Kolhe</i>	144-147
34	A Survey on Visual Concept-Detection <i>Prajakta Bhosale, S.D. Sawarkar, Nita Patil</i>	148-151
35	Reliable Human Identification using Iris as a Biometric <i>Jayshree Sali, Akram Pathan, Rakesh Kushwah</i>	152-154
36	A New Fighter Alarm Based on Interaction <i>Hamidreza Hajiaghai, Mehdi Gheisari</i>	155-156
37	SQLIProtectParser For Preventing SQL Injection Attack <i>Neha Patwari, Ratna Nayak, Nishtha Mathur, Shruti Mathur, Purvi Sankhe</i>	157-160
38	Biometric Identification And Real Time Control and Security System in Smart Car <i>Riya Parikh, Pooja Kammala, Manasi Kadam, Neha Patwari</i>	161-165
39	Sign Language Recognition Using Computer Vision <i>Radhika Gokani, Bhavesh Mishra, Dinesh Gowda, Sudhir Dhekane</i>	166-168
40	Design of Guest Tagging System for Surveillance using IoT Analytics <i>Prachi Jain, Lokesh Ucheniya, Satyendra Yadav, Akash Singh, Deven Shah</i>	169-174
41	Homomorphic Encryption for Cloud Security <i>Prathmesh Padwal, Misal Dave, Jaykumar Yadav, Vijaykumar Yele</i>	175-179
42	Authentication Access Control <i>Akanksha Jain, Shivani Rege, Pooja Jain, Sudhir Dhekane</i>	180-182

43	Web Server Log Analysis <i>Sangeeta Vhatkar, Vishal Patel, Shubham More, Harsh Raval</i>	183-187
44	Cart Free Shooping <i>Nishant Gandhi, Ajinka Padave, Kajal Gogri, Rahul Neve</i>	188-191
45	Dynamic Digital parking <i>Siddesh Rao, Nikita Sharani, Yash Shah, Neha Kapadia</i>	192-196
46	Vehicle Tracking/Parking System <i>Vandana Munde, Shreyas Sanil, Yash Mujumdar, Shweta Panchal</i>	197-200
47	Finger Print Recognition for Examination System <i>Ayush Dutta, Harshal Kale, Shubham Tare, Vijaykumar Yele</i>	201-203
48	Website and Application for Result Analysis <i>Arvind Gupta, Akhil Baliga, Sandip Bankar</i>	204-207
49	<i>Log Stacking and Threat Hunting</i> <i>Pranjal Aswani, Aman Cheval and Abhishek Naik, Hetal Amrutia</i>	208-211
50	Look based Media player <i>Vandana Munde, Atulkumar Mishra, Abhijit Chaudhary, Rahul Vishwakarma</i>	212-215
51	RFID based Electronic Payment System <i>Shreya Mistry, Nishita Desai, Tanay Sanghavi</i>	216-219
52	Result Analysis and Prediction <i>Hetal Amrutia, Siddhant Acharya, Ujjwal Sharma, Sudhesh Singhania</i>	220-224
53	An Approach to Maintain Attendance Using FTP Tools <i>Saurav Jangir, Mahendrasingh Parmar, Aakash Tripathi, Purvi Sankhe</i>	225-229
54	Mathematical Morphology Based Algorithm for Image Noise Reduction <i>Divya Sharma, Arpit Vakharia, Juhi Vadia, Bijith Marakarkandy</i>	230-233
55	Tracking System For Mentally Challenged People <i>Chirag Rai, Ayush Kedia, Aditya Rao Nuty, Namdeo Badhe</i>	234-237
56	Anti-Theft Detection and Tracking System for Car Using IoT <i>Harish Baheti, Akshata Desai, Abhishek Pathak, Zaheer Alam</i>	238-241
57	Secure Cloud storage of Biometric Templates with Secure Network Coding <i>Jewel Correia, Vikas Kaul, Deven Shah, S. K. Narayankhedkar</i>	242-246
58	Chat-Bot for Banking Industry <i>Meghna Singh, Abhileen Pandey, Raj Singh, Pranjali Kasture</i>	247-249
59	RFID Based Attendance Management System <i>Shashank Shukla</i>	250-255

ADVANCES IN OPTIMIZATION OF DECISION SUPPORT SYSTEMS

- 60 Soft Computing Based Empirical Analysis for Detection of Occlusion in Face Images 257-261
Rohit Raja, G. R. Sinha, Rajkumar Patra, Su Su Hlaing
- 61 A Comprehensive Review and Efficient Implementation of Privacy-Preserving Associative Classification 262-268
Darshana Patel, Avani Vasant
- 62 A hybrid (IWO-TLBO) meta-heuristic optimization method for data clustering application 269-276
Seyed Mohamad Bagher Davoodi, Mehdi Gheisari, Mehrnoosh Bazrafkan
- 63 Diagnosis of Heart Disease Using Neural Networks 277-282
Chaitanya Roygaga, Suraj Punjabi, Swapnil Sampat, Tanuja Sarode
- 64 Machine learning approach to predict farmer's loan/credit repayability using weather prediction and credit history 283-286
Waheeda Dhokley, Umair Shaikh, Samad Ansari, Nehal Ansari
- 65 Android Based Car Tracking System 287-290
Karan Pawar, Aditya Gupta, Radhika Kotecha
- 66 Supply Chain Management and Data Mining Techniques: A Review 291-294
Pranjali Kasture
- 67 Survey on Improving Web Search By Analyzing Various Facets 295-297
Gunjan Deshmukh, Govinda Borse
- 68 A Literature Review on Drowsiness Detection System 298-304
Raees Ahmad Noor Mohammad, Rahatullah Khan
- 69 Development of Pathology Informatics tool using disease symptoms network: Identification of Ailment and Remedies 305-308
Lutuf Islam, Ataul Islam, Aiman Khan, Veena Misar, Himanshree Nagvekar, Mrinmayee Rane
- 70 Machine Learning Methods for Online News Popularity Prediction 309-312
Ravi Sitapara, Pradeep Kotian, Prashant Chaudhary, Shridhar Kamble
- 71 Splunk for Big Data Analytics 313-315
Ratna Nayak, Purvi Sankhe, Shruti Mathur, Nishtha Mathur, Neha Patwari
- 72 To Understand Drug usage by Mining Social Media Data 316-319
Shwetha Yadav, Kamal Shah
- 73 Project Title Allocation & Monitoring System 320-322
Prince Jain, Pravin Jodha, Jagdish Choudhary, Neha Kapadia

74	Santander Customer Satisfaction <i>Rajat Nareshkumar Khemka</i>	323-326
75	Health Expert <i>Vinayak Sharma, Gayatri Mulchandani, Shivani Subnani, Gaurav Gianani, Anjali Yeole</i>	327-329
76	Social Network Analysis using Data Mining Techniques <i>Sneha Yadav, Vikas Kaul</i>	330-333
77	Opinion Mining of Users Feedback on the Web <i>Sanjana Puthran, Shweta Mishra, Isheta Tilara, Purvi Sankhe</i>	334-337
78	Price trend prediction of stock market using data mining algorithm <i>Vinesh Yadav, Pramod Chauhan, Sunil Chaurasia, Pranjali Kasture</i>	338-340
79	A Survey on Sentiment Analysis Algorithms for Opinion Mining <i>Dhiraj Patil, Zaheer Aalam</i>	341-345
80	An Empirical Study of the Impact of Knowledge Management on E-Supply Chain Management in Indian Public Sector Oil Companies <i>Aaditya Desai and Sunil Rai</i>	346-352
81	Adaptive Automated Attendance Analysis for Education System <i>Kameshwar Gupta, Pratik Mishra, Akanksha Das, Bhushan Nemade</i>	353-356
82	College Management System and Data Analysis with Machine Learning <i>Mary Margarat, Raj Asha, Saurabh Maydeo, Meet Gadoya</i>	357-360
83	English to Hindi Machine Translation <i>Ankita Gaud, Chanda Jha, Akash Naik, Shruti Mathur</i>	361-363
84	Smart Warehouse – Reduction in Post-Harvest Losses <i>Aishwarya Nair, Anand Ranka, Vikram Deshmukh, Kamal Shah</i>	364-366
85	An Intelligent College Enquiry Voice Chatbot <i>Vedprakash Jha, Bharat Paliwal, Vivek Kumar, Neha Patwari</i>	367-370
86	Automatic Human Detection System <i>Bhushan Nemade, Shreysijay Singh, Pooja Yadav, Kirti Vishwakarma</i>	371-374
87	Webpage Ranking Search Engine with Search Engine Optimization Suggested Project <i>Harsh Bharucha, Tejas Sutone, Vikas Kaul</i>	375-377
88	Recommendation System Leveraging Heterogeneity of Trust in Social Networks <i>Vijay Varma, Rohit Yadav, Deepak Yadav, Aaditya Desai</i>	378-380
89	Automation Using Raspberry Pi 3 Smart Mirror <i>Suresh Rajpurohit, Harsh Singh, Yash Kosambia, Sandip Bankar</i>	381-383

90	Business Exchange <i>Rajat Nareshkumar Khemka</i>	384-388
91	Patient Data Analysis using Structured Data Marts <i>Akanksha Chand, Jyoti Vikram, Padmaja Shukla</i>	389-392
92	Intelligent Tourist System <i>Rohan Shetty, Priyanka Verma, Dharti Singh, Radhika Kotecha</i>	393-397
	<i>Appendix</i>	399-414
	<i>Index</i>	415-420

International Conference on
Emerging Trends in
Communication and Networking

Document Summarization as A Service (Dsaas)

A Web Application for Summarizing Technical Documents

Zainab Mirza
Department of I.T,
MHSSCE, Mumbai

Abdul Samea
Department of I.T,
MHSSCE, Mumbai

Khan Shaheer
Department of I.T,
MHSSCE, Mumbai

AbuZerMalvi
Department of I.T,
MHSSCE, Mumbai

Abstract— In this paper, a document summarization service is implemented (DSaaS). This paper discusses the summarization method of a technical, word-only document that is taken from a user. The implementation uses sentence scoring, extraction and compression of sentences from a document to produce a summary. DSaaS uses Natural Language Processing APIs and Machine Learning techniques for extracting/abstracting informative sentences, Amazon Cloud Services for deployment of Web Application and management of Database services.

Keywords-document, summaries, extraction, sentences, cloud, natural language processing, machine learning.

I. INTRODUCTION

Now a day, many people rely on wide variety of sources to stay informed from news stories to social media posts to search results. Technical documents also provide a way of conveying information to interested users. Being able to develop NLP models that can automatically deliver summaries of longer text can be useful for digesting such large amounts of information in a compressed form, and is a long-term goal. To summarize, models need to be able to comprehend documents and refine the important information, tasks which are highly challenging for computers, especially as the length of a document increases. This summarization system is based on the work of Luhn et.al [3] which focuses on deriving informative sentences from a given paragraph based on the removal of stop words and identifying essential words or more importantly, paragraph related keywords for summarization. This procedure provides an algorithm for extracting relevant sentences which in turn, are formatted or possibly trimmed down to form abstract sentences using parse tree analysis and reduction of words from input sentences, which is based on work done by YuyaUnno&Jun'ichiTsujii [12].

II. LITERATURE REVIEW

The most cited paper on summarization is that of (Luhn, 1958), that describes research done at IBM in the 1950s. All sentences were ranked in order of their significance factor, and the top-ranking sentences are finally selected to form the auto-abstract. Related work (Baxendale, 1958), also done at IBM and published in this same journal, provides some useful information on a particular feature helpful in finding salient parts of documents: the sentence position.

In his method of extraction, the author examined 200 paragraphs to find that in 85% of the paragraphs the relevant topic sentence came as the first one and in 7% of the time, it

was the last sentence. Thus, this method was simplistic but lacked informative approach.

A Trainable Document Summarizer [Julian Kupiec, Jan Pedersen and Francine Chen, Xerox Palo Alto Research Center]- The scoring criteria employed include participation in predefined semantic roles, rhetorical relations, inclusion of phrasal index terms, document-specific keyword frequencies, and the assessment of sentence similarity structure. Methods either assume the document exists in isolation that is, unrelated to other documents, or in the context of a larger collection. According to this paper, summaries that are 25% of the size of the average test document, it selects 84% of the sentences chosen by professionals. For smaller summary sizes, an improvement of 74% was observed over simply presenting the beginning of a document.

Information Extraction architecture is based on Abstraction Schemes. An abstraction scheme consists of Information Extraction rules, content selection and one or more generation patterns, all created by hand. Each abstraction scheme is designed to address a theme or subcategory. Thus, rules that extract information for the same aspect within the same scheme will share a similar meaning. An abstraction scheme aims to answer one or more aspects of its category, and more than one scheme can be linked to the same aspect. This system, can generate useful summaries for a given category. The key elements of information are present, stated in a way that is easy to understand. The summary generated is compared against other methods, which contain up to an average of 100 words, but through this method, the summary generated is around 25 words long.

Text Summarization via Hidden Markov Models and Pivoted QR Matrix Decomposition [John Conroy Center for Computing Sciences Institute for Defense Analyses]- The authors of this paper proposed two models and are asfollows:

The first method summarizes a document by seeking the main ideas. Co-location is used to disambiguate the meaning of words that rely on context. This method uses the SRA's NameTagTM (NameTag is an automated text indexing system that recognizes and classifies names and other key phrases such as time and numeric expressions.) To recognize named entities and WordNet to associate synonyms. Once the terms are defined, a term-sentence matrix is formed. The job of the automatic summarization system is to choose a small subset of these vectors to cover the main ideas (terms) in the document. The method proposed is taken from numerical linear algebra: the QR decomposition with partial pivoting.

The second method proposed for text summarization is a Hidden Markov Model (HMM). Jing and McKeown [JK99] previously proposed a HMM for decomposing a human summary, i.e., mapping the component parts of a summary generated by a human back into the document. The HMM has two kinds of states, corresponding to summary sentences and non-summary sentences. The beginning of the document is handled by special states so that early document structure can be easily captured.

The results were compared to human generated summaries. To test how well human summaries could be predicted, 4 people generated extract summaries for around 40 articles from the latwp-test data set. The QR method and the HMM with two features with the QR method of extraction of sentences, trained using the latwp-dev data and summaries generated by M(a single person), were compared against each of the four human summaries. In addition, the four human's summaries were compared against each other. The summaries generated by M included the title 98% of the time, while the other summarizers had been instructed to omit it. Comparing the HMM to the four human summarizers the scores range from 51 to 58; comparing the QR with the four human summaries gave scores between 50 and 59(on a scale of 0-100).

III. EXISTING SYSTEM

There are many Automatically Summarizers available online but mostly based on Extractive Summary method.

- Splitbrain.org
- Summarizing.biz
- Autosummarizer.com
- Smmry.com
- Online Summarize Tool

From our prospectus, the issues with these systems is that it is extractive method based and just cut down the sentences from the document uploaded. For e.g. If you upload a document and select the summarization ratio to 40%, it'll remove 40% of the sentences and give the remaining text as output (Splitbrain.org). Wherein some systems summarize on basis of sentences the user desire to have in his summary (Smmry.com). The Summarizing.biz summarizes the text on basis of the words the user needs. For e.g. If the user inputs a text and selects 100 words, the summary of 100 words will be returned for the user.

Hence, most of the existing systems work on the threshold specified or selected by the user while uploading a text for summary.

IV. OBJECTIVE

Summarization is a problem of Natural Language Processing because, to do it properly, one has to really understand the point of a text. An approximation called extraction is more feasible today. To create an extract, our system needs simply to identify the most important/topical/central topic(s) of the text, and return them to the reader. Although the summary is not necessarily rational, the reader can form an opinion of the content of the original. The objective of this project is to develop a satisfactory abstractive web application for the readers to give them a complete idea of the lengthy documents/stories/articles etc. they upload. The summary produced helps the reader to reduce time

instead of reading the whole document and it provide swift Information from the large document.

V. WORKING OF THIS SYSTEM

Our proposed system is a Web Application (DSaaS) which will be deployed on Amazon Cloud. This system will be deployed using AWS Elastic Beanstalk, which will be used for deploying and managing the web application on cloud platform. our application also uses Amazon RDS service for using MySQL, as it is required for the systems database requirements. The application will have a login/register page for the users to create and manage their accounts. After logging in, the users will be able to access their previous summarized documents. When a user logs in, he will be able to upload documents. After uploading a document, user can view the generated summary for the document.

DSaaS will provide a user to upload a doc/pdf file for summarization or the user can use Text Box to paste some long text. Each user can save his/her document, along with the related summary that is generated, and can view it later. DSaaS will use the Amazon cloud based RDS for storing documents, user's credentials, sentence stores, summaries for future summarization on similar topics, etc. Feedback feature enables a user to provide feedback for the resulting summary. Upon receiving feedback, the application re-evaluates the document for generating a more precise summary. The proposed system achieves the output through the following steps as depicted in figure 1:

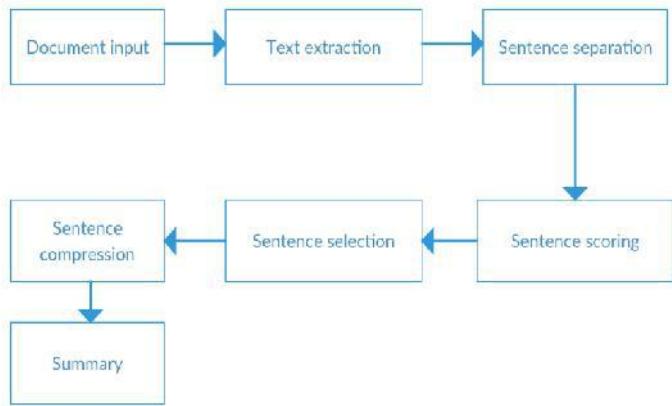


Fig 1: Summary Generation Steps

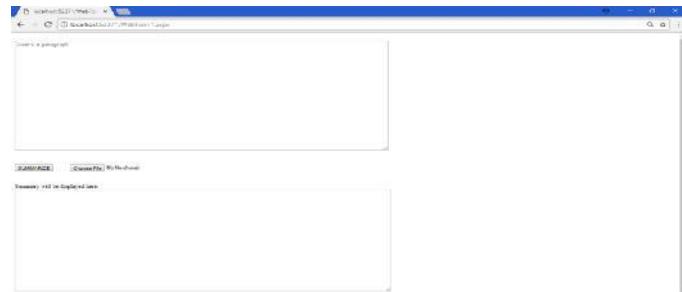


Fig 2: Sample UI

VI. ADVANTAGES OF PROPOSED SYSTEM

In the recent years several software were developed, for the purpose of creating summaries of texts and documents automatically, without human intervention. In this article we will cover the advantages of such programs over traditional approach to summarization.

- A. Reading the entire article, dissecting it and separating the important ideas from the raw text takes time and effort. For example, reading an article of 500 words can take at least 15 minutes. Automatic summary software summarizes texts or documents in a split second. This allows the user to read less data but still receive the most important information and make solid conclusions.
 - B. The critical words are usually words with tactical importance, such as ‘bomb’, ‘explode’, etc. While humans can oversee an important sentence, computers will not miss it so important ideas will always be mentioned.
 - C. DSaaS store the data on cloud database which help developer easy access and security.
 - D. Each user can save his/her document, along with the related summary that is generated, and can view it later.
- v.The DSaaS application is being developed in C#. C# is a server-side scripting language and shall be easy to maintain.

vi.The system shall be available 24 hours a day. User should be able to login anytime and access the system.

VII. SUMMARY GENERATION PROCEDURE

The proposed system requires text only document as an input. The steps for converting an initial document into a summarized document include the following:-

A. Sentence Separation-Each of the sentences are separated and stored in a database table, according to their occurrence in the document. For this method ,we use a MySQL service provided by Amazon RDS(Amazon’s cloud based Relational database management service).

B. Sentence Scoring-the straightforward approach for selecting a sentence is based on combination of parameters such as sentence length,number of stop words removed, and keyword identification in a sentence, that is identifying the best words to be included in a summary. We use a method based on Inverse document frequency formula which highlights the importance of a word ,or more precisely, identifies how rarely a word occurs in given number of documents.IDF for a term “t” is defined as:

$$Idf_t = \log_{10}(N/df_t)^{[8][1]}$$

where N defines the number of documents, df defines the no. of documents that contains a particular word. The value of IDF for a particular word hence is always changing, and so is the importance of a word in every document. The score is normalized into a range of 1-10 for simplicity in calculating average IDF of a sentence.

C. SENTENCE EXTRACTION-Once the sentences have been assigned their IDF,sentences are extracted based on specific threshold defined by the user. This threshold may lead to decrease or increase in size of a summary. Also, if summary is found irrelevant, the user may provide feedback for inclusion of specific sentences in the summary by providing specific keywords to be searched in the sentences.

D. SENTENCE COMPRESSION-This technique represents abstraction of sentences that are finally selected for the summary. For this purpose, we have selected the Google Natural Language Processing API for generating a parse tree of a given sentence. A parse tree signifies the connectivity or dependency among related words in a sentence and also specifies the parts of speech for every given word. A simple example of a parse tree generated by this API for a sentence “I have no doubt in him” is shown below:

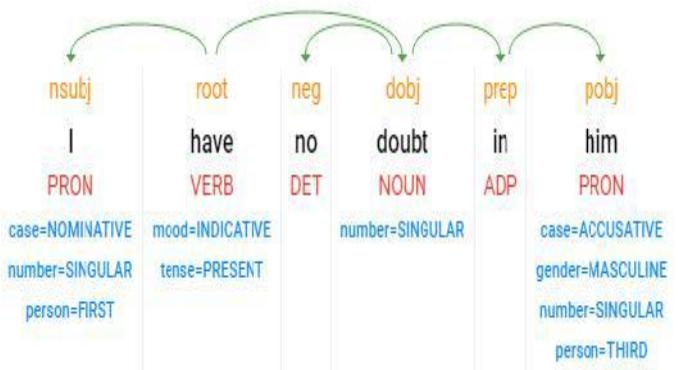


Fig 3: Sample Parse tree for a sentence generated through Google NLP API

A parse tree generated as above will help the software to remove some of the less relevant words in the sentence based on their parse label. Also, compression may require to change the tense of a sentence. For this purpose, we have selected the “tensify” API available for Visual Studio.

An example of sentence compression that can be expected from our system is as follows: “The compression algorithm is good as it deletes and reduces many irrelevant words” can be trimmed down to “The algorithm is good as it deletes irrelevant words”. For our system, we assume that sentences require compression only when their length exceed a minimum (around 10-15 words), and not for relatively smaller sentences. We focus on maintaining maximum number of nouns given in the input sentence, while maintaining compression by reducing number of verbs related to a given noun.

VIII. USER OF DSAAS

User of the DSaaS will be mostly students. This software will aid them to understand a lengthy document through its summary and hopefully, help them to get an idea about initial document.

IX. SCOPE

Currently our focus for creating the software is mostly on the extractive summarization with a limited ability of abstractive summarization. The system will only accept a document of about 2-3 pages or an input paragraph and not very lengthy documents

such as a book. Also, currently this system is designed to extract paragraphs from a text-only document and does not take documents with figures or diagrams along with its textual content.

X. EXPECTED RESULT

For a document summarization, we expect a representative summary or abstract of the entire document, by finding the most informative sentences. We aim to provide a summary of the long document to the user in a doc or pdf format. This summary may or may not be equally relevant to every reader, as it can derive only a portion of entire document. The user can upload a long text document in the provided area or can upload a document/article from the internet in pdf format through a simple GUI.

XI. CONCLUSION

In this paper, implementation of Document Summarization as a Service is discussed. Further, the short historical background related to the research in the area has been elaborated, also attention has been paid to the general approaches in summarization currently used and proposed. The working of DSaaS was further discussed with implementation modules and services. Improvements are expected in future to make the system more stable, effective and accurate as per the user.

XII. ABBREVIATIONS

DSaaS-Document Summarization as a Service.

NLP-Natural Language Processing.

VS- Visual Studio.

AWS- Amazon Web Service.

DB- Database.

RDS-Relational Database Service.

API-Application Program Interface.

REFERENCES

- [1] Manjula.K.S, Sarvar Begum, D. Venkata Swetha Ramana- "Extracting Summary from Documents Using K-Mean Clustering Algorithm," International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 8, August 2013.
- [2] Tadashi Nomoto, Yuji Matsumoto- "A New Approach to Unsupervised Text Summarization," 24th annual international ACM SIGIR conference on Research and development in information retrieval.
- [3] Hans Peter Luhn. 1958. The automatic creation of literature abstracts. IBM Journal of research and development 2, 2 (1958), 159–165.
- [4] Anjali R. Deshpande, Lobo L.M.R.J. -"Text Summarization using Clustering Technique". International Journal of Engineering Trends and Technology (IJETT), Volume -4, Issue-8, July 2013.
- [5] Ayush Agarwal, Utsav Gupta- "Extraction based approach of summarization using k-means clustering," IJSRP, Volume 4, Issue 11, November 2014 Edition.
- [6] Roshan Chettri, Udit Chakraborty- "Automatic Text Summarization," International Journal of Computer Applications (0975 – 8887) Volume 161 – No 1, March 2017.
- [7] Giri Virat, Apte S.S, BewoorMrunal- "K-means driven document summarization," International Journal of Computer Applications and Business Intelligence, April-June 2010, Volume:01, Number:02.
- [8] Eduard Hovy and Chin-Yew Lin- "Automated Text Summarization in SUMMARIST," In Advances in Automatic Text Summarization, I. Mani and M. Maybury (editors), 1999.
- [9] Jing and McKeown- "Text summarization using Hidden Markov Model," 24th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 406-407.
- [10] Pierre-Etienne Genest, Guy Lapalme- "Fully Abstractive Approach to Guided Summarization," ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2 pp. 354-358.
- [11] Julian Kupiec, Jan Pedersen and Francine Chen- "A Trainable Document Summarizer," SIGIR '95 Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval pp. 68-73.
- [12] YuyaUnno, Jun'ichiTsujii- "Trimming CFG Parse trees for sentence Compression using Machine Learning approaches," ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Sydney, Australia, 17-21 July 2006.

Work-Stake Hybrid: A Trustless Consensus Mechanism for Blockchain

Yash P. Chetnani
Computer Engineering
V.E.S.I.T., Mumbai University
Mumbai, India
yash.chetnani@ves.ac.in

Nupur Giri
Computer Engineering
V.E.S.I.T., Mumbai University
Mumbai, India
nupur.giri@ves.co.in

Abstract-The authors aim to propose a new mechanism for trustless consensus in peer-to-peer networks using Blockchain. The previously used techniques of Proof of Work and Proof of Stake have been combined into a new parameter called the ‘Cost factor’ for a more efficient system. The Cost factor will now be the criterion the miners will have to fulfil in order to be eligible to mine new blocks. This new parameter takes into account both, a miner past performance, and its current capacity to determine its right to mine a block. The block size and buffer period have been thus introduced to achieve the objective of this paper. The proposed system leads to faster confirmations, reduced operating costs and immunity against the Sybil attack.

Keywords- Block chain, trustless consensus, security

I. INTRODUCTION

Bitcoin [1] in 2009 started the trend of blockchain based peer to peer crypto-currency. Since then the applications of blockchain have been expanded into traditional banking, healthcare, etc. The building block of this system was the Proof of Work (PoW) [2] protocol. The purpose of this protocol was to establish a consensus among members involved in the system while maintaining complete anonymity. This is called as Trustless consensus. Later developments led to the Proof of Stake (PoS) [3][4] system. Both these protocols have their advantages and disadvantages.

Building up on these systems this paper provides a hybrid between PoW and PoS. The proposed protocol takes into account combines the advantages of PoW and PoS, improving both, their security and efficiency, while at the same time eliminating the obstacles faced by them individually such as the ‘nothing at stake’ problem. The ability of this system to reevaluate itself after every block is mined helps us counter dynamic attacks.

II. REVIEW

As the essence of Blockchain lies in the fact that all its participants are anonymous to each other, building trust people to come to a single decision without any direct contact with each other. The two main techniques used by blockchain to solve Byzantine General’s problem and thus achieve trustless

consensus are, Proof of Work and Proof of Stake. Both these techniques follow the same principle, that is, they make cheating the system an expensive operation, making it non-profitable. The PoW technique does so, by making ‘mining’ a computationally heavy task. Here any attacker would require more computing power than the rest of the network to succeed, which is not easy but still possible. This method has led to a rise in the operating costs. PoW can be considered to be a Markov process [7] as the past events do not have any consequence on the future. On the other hand, PoS has the opposite approach. PoStakes into account the involvement of the miner in the system in terms of its stake. This stake can be calculated in various ways, for example in Peercoin[4] it is the number of coins held. A user with a large number of coins can continuously mine blocks without actively invested in the system, that is, even with no present investment in the system the user can still influence it. As a consequence, both these techniques have deficiencies which can be exploited by the attacker. In a PoW based system, an attacker can forge new identities to fool the system while in PoS based systems, an attacker can distribute its resources (stake) among its own multiple virtual identities. Even though these multiple virtual identities do not physically exist, they extend the attackers influence over the system. Another Attack, such as the Sybil attack, uses these multiple virtual identities to obtain a majority in the system, thus giving the attacker full control.

III. SYSTEM

A. Work-Stake Hybrid

To overcome the aforementioned drawbacks of existing techniques, this paper proposes a new system, called the Work-Stake Hybrid. Both, the computational power required in PoW and stake required in PoS are what we call the cost of the system. The cost helps us, one, determine how invested a user is in the system, and two, dis-courage malpractices. The cost in the proposed system is called the ‘Cost factor’. A node’s Cost factor is determined by its past performance. Every time a node among them becomes a problem [6]. The objective of Byzantine Generals problem is to find a way for a group of nodes to agree on a decision and thus achieve consensus.

its record is stored. Depending upon how close to the PoW target a particular node has been able to reach, it'll be assigned some Cost factor. The difficulty in a PoW based system is defined by a random hash. The task for the miner is to find a hash smaller than the set difficulty. A hash in the system is represented as a hexadecimal number. The difficulty can also be defined in the terms of the number of leading zero's in the hash. The Cost factor earned is determined by the number of leading zero's in miner's hash closest to the target. A difference of 'n' zero's will give the miner a Cost factor of

$$10^{-n} \quad (1)$$

till it reaches a difference of 10 zeros. This Cost factor accumulates over time till the node mines a block. The Cost factor decides both, whether or not a node can mine a block and its priority over other miners. A minimum Cost factor of 12 is required to be eligible to mine a block. As the maximum achievable 'Cost factor' in a single round of mining is 10, it'll require a miner to continuously mine using the same identity. Also, since the Cost factor is unique for every miner, it cannot be transferred or shared. Every time a node mines a block, its Cost factor becomes zero forcing it to start over again. This creates a dependency in the system for every miner to possess both, a past stake and continuous presence in the system.

B. Block Size and Buffer Period

For the miner to prove its 'Cost factor' for the current 'n'th block it requires to attach all the 'n-8' unsuccessfully mined blocks along with the block it's attempting to mine. These 'n-8' attached blocks along with the current 'n' block establishes the miner's right to mine the current block. The set difficulty will allow the miner to achieve a minimum 'Cost factor' of 2 within the set time.

The two challenges faced now are, one, there might be a point where the entire network will be with a Cost factor of less than 2, due to shorten set time, making it impossible for anyone to mine a block. Second, 'n' and 'n+1' block simultaneously propagating through the network.

The paper proposes the introduction of buffer period to handle the above challenges. The buffer period is the time between receiving a block to be mined and sending a block to the network to be mined by others.

In order to achieve a minimum 'Cost factor' of 2, the buffer period should lie in the range of 80-100 seconds. To understand this one requires to establish the relationship between buffer period (time) and block size.

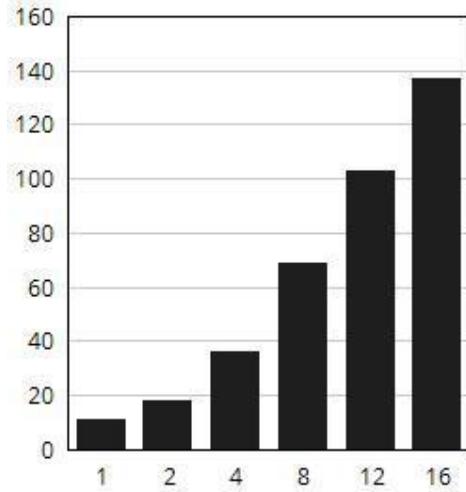


Fig 1: Establishes the direct relationship between the size of the block and the time taken to propagate data through the network of 3000 nodes. [8]

One can see in the Fig 1. That within the proposed 80-100 seconds that the size of total data to be transmitted is limited to approximately 5Mb. For the proposed system with 9 successful blocks, each block is of approximately 512kb.

The miner in this buffer period is supposed to perform two tasks, one, and check all the blocks attached with the block to be mined for the authenticity of the claimed Cost factor and second, find a nonce for the remainder of the period for the next block. Thus, the buffer period ensures that every miner gains a minimum 'Cost factor' of 2.

IV. SIMULATION

A. Set-up

Two chains, the honest chain and the attacker's chain, compete against each other with a fixed share of resources. The goal of both these chains is to reach a predefined number (1000) upon which no further alterations can be made to the blockchain. This share of resources is decided by a fixed number between 0 and 1 for each scenario. The winner is decided by generating a random number between 0 and 1. If the number generated is higher than the attacker's share, then the honest is incremented, else the attacker's chain increments. Since the attacker will have to divide its resources among at least two nodes to mine consecutive blocks (due to the Cost factor), the share of resources that the attacker can use at any instance will be half of what it owns. This also reduces the attackers mining probability by half.

B. Analysis

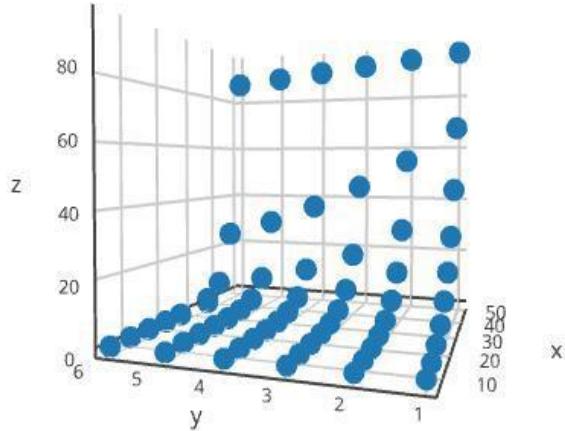


Fig 2: The original pure PoW based system. The X-axis denotes the share of resources controlled by the attacker and the Y-axis represents the initial lead the honest chain has in terms of blocks. The Z-axis gives the probability of the attacker chain overtaking the honest chain at any given time. Each case was simulated $10^6 \times 100000$ times with the final result being the average.

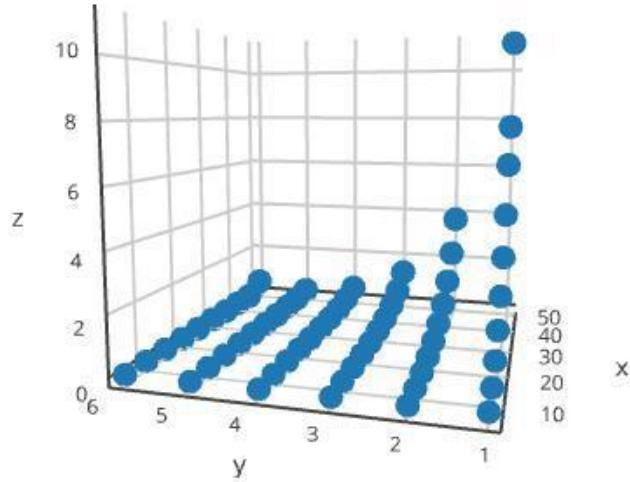


Fig 3: The proposed system employing the Cost factor

The resources allocated to the attacker which is represented on the X-axis is the total percentage and not the amount available at any and every instance. The considered course of action on the attacker's part is to distribute these resources evenly among two nodes to be able to consecutively mine blocks.

As can be seen, even with a single block lead and 50% resources controlled by the attacker the probability of overtaking the honest chain is below 12% in the proposed system

V. RESULTS

A. Faster Confirmations

As proved earlier, a single miner can only dedicate his entire hashing power to a single node. This node then is by virtue of its Cost factor, ineligible to mine the next block. In order to mine continuous blocks, a miner will have to divide its resources among the many identities used to mine the blocks which effectively reduces its hashing power by the number of nodes it is dedicated to. Considering a miner divides its resources among two nodes, every node now has half the original hashing power and hence, half the probability of mining a block. Earlier, control more than 50% of the resources virtually guaranteed a win for the attacker, but now even with 70% of the resources the probability of overtaking the honest chain is well below 20%. With this reduced capability of an individual to mine blocks, the chain can now be secured with fewer confirmations as compared to the earlier magic number of 6 blocks [1].

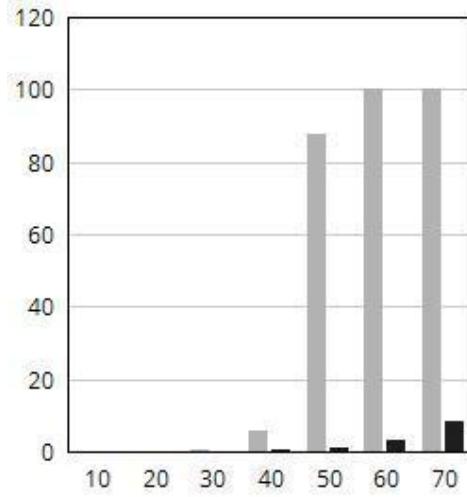


Fig4: Comparison of Probability of Attacker chain overtaking Honest chain in Old System vs New System. Grey represents the old system with 6 block lead, and black represents the proposed system with a 3 block lead. X-axis denotes the percentage of resources controlled by the attacker and the Y-axis gives the probability of success for an attacker

B. Reduced Operating Costs

With the recent advances in mining hardware, the cost of operating a mining rig has increased too. A mining ASIC can consume as much as 1200W of power. This over the period of 24 hours can incur costs greater than that earned by mining. One way to increase the efficiency of the system is by increasing its productivity. If more blocks are mined in the same amount of time than previously possible, it would make the operation more profitable. Since the proposed system mines at a rate '5' times greater than before, it helps us overcome another obstacle previously faced by the industry. With the move away from a pure PoW based system, a reduced energy consumption can be seen.

C. Nothing at Stake Problem
As the system progresses from a purely PoW based system to a hybrid between PoW and PoS, the author also intends to

tackle the problems faced by a PoS based system. The major issue with any PoS based system is called the ‘Nothing at Stake’ problem.

A miner can simultaneously bet his wealth on all the current chains. Irrespective of which chain wins the race, the miner will gain his share of profit. With nothing to lose, the miner is free to act opportunistically and disobey the rules laid down for the system. Since, unlike wealth, the Cost factor from one chain cannot be used in another, the miner is forced to work on only a single chain at a time. This makes working on the honest chain the only profitable option for the miner.

D. Sybil Attack

As the system progresses from a purely PoW based system to a hybrid between PoW and PoS, the author also intends to tackle the problems faced by a PoS based system. The major issue with any PoS based system is called the ‘Nothing at Stake’ problem.

A miner can simultaneously bet his wealth on all the current chains. Irrespective of which chain wins the race, the miner will gain his share of profit. With nothing to lose, the miner is free to act opportunistically and disobey the rules laid down for the system. Since, unlike wealth, the Cost factor from one chain cannot be used in another, the miner is forced to work on only a single chain at a time. This makes working on the honest chain the only profitable option for the miner.

VI. CONCLUSION

A new system for establishing a trustless consensus in a P2P environment using Blockchain has been successfully developed. The introduction of Cost factor and its ability to combine the merits of both PoW and PoS systems has resulted in a more secure and efficient blockchain mechanism. Various alterations to be made to existing system have been explained. The effectiveness of the proposed system has been measured using different parameters such as number of confirmations required, time and types of attacks possible. These have been confirmed by simulating different real life scenarios.

REFERENCES

- [1] Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008
- [2] akobsson, Markus; Juels, Ari (1999) Proofs of Work and Bread Pudding Protocols Communications and Multimedia Security. Kluwer Academic Publishers: 258–272.
- [3] Popov, Serguei (2016) A Probabilistic Analysis of the Nxt Forging
- [4] King, Sunny PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake Retrieved 2014-11-17.
- [4] Douceur, John R. (2002) The Sybil Attack International workshop on Peer-To-Peer Systems. Retrieved 23 April 2016.
- [5] Lamport, L.; Shostak, R.; Pease, M. (1982). "The Byzantine Generals Problem". ACM Transactions on Programming Languages and Systems. 4 (3): 382–401. doi:10.1145/357172.357176 .
- [6] R. Serfozo, Basics of Applied Stochastic Processes, Probability and its Applications, Springer-Verlag Berlin Heidelberg 2009
- [7] TradeBlock (2015, June 23) Bitcoin Network Capacity Analysis – Part 6: Data Propagation Retrieved from <https://tradeblock.com/blog/bitcoin-network-capacity-analysis-part-6-data-propagation>

Smart Surveillance Using Internet of Things

Bhushan Nemade

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Dipti Chavan

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Abstract- The proposed system uses Internet of Things (IoT) for developing a home security surveillance system. Internet of Things has lately been an emerging concept that attempts to connect all potential objects to interact with each other on the internet to provide a secure, comfortable life for humans. The cameras is used for face detection to identify the owner and permit entry without having to go through the procedure. This way a user can be able to know if there is any wrong activity going on in the house using just his/her phone. The idea of implementing a Smart Home Surveillance System is to enable the user to remotely control the activities of his house, more specifically security. The prominent advantage of this system is that it reduces the need for much human interference in the monitoring of a home security, which is quite necessary in these busy times. Proposed system is intended to design and implementation of a home surveillance system with the help of Raspberry Pi and also Android application to alert the user of a possible intrusion of their homes by sending a picture of the person detected over the Internet, thus facilitating remote surveillance. The solution to this setback is to use a system that allows remote monitoring or surveillance of the house, without demanding the owner's presence to monitor the surveillance. In emergency conditions, required measures

Keywords— IOT, face detection, Raspberry Pi

I. BACKGROUND

Proposed system deals with providing the security for home automation system. Home security Systems are of prime importance for family because these system helps in proving safe environment for homes and provides secure, smart and efficient living.



Fig 1: Different Sensors

There exist various systems for surveillance such as camera, CCTV etc. In Proposed systems, the person who is present in a particular region can observe about the activity happening in the place. When person is migrating from one region to another region, Proposed system can also perform the observation about the suspicious activity happening in that place. Main advantage is that proposed system provides privacy on both sides since it is observed by only one person. Another advantage is that the proposed system can be developed with simplified circuit. Raspbian operating system is used by proposed system. Image can be easily transferred to the smart phone by installing Raspbian operating system. The use of Cloud Computing enables storage of images in large quantities on the cloud; so on board memory is not used and can be saved. It is necessary and important for every home owner to consider adding a home security system, as burglaries, thefts and murders have become routine in big cities. This paper focuses on implementation of Smart surveillance system using PIR sensor and Raspbian operating system for smart phones. It helps to use smart phone technology to facilitate required security to the homes and other sensitive areas using IoT. The proposed surveillance system receives information and sends it a cell phone using WI FI dongle with the help of mobile

application. Using Infrared sensor, human detection can be achieved. When motion is detected, the cameras start recording the human presence and the RaspberryPi system informs the owner about suspicious activity over a smart phone.

The major contribution of proposed system is to provide a factor of convenience to the users regarding the security and safety of their houses. Instead of the conventional doorbell system, the two strong concepts of Internet of Things and Cloud Computing coming together to produce a home security surveillance system is part of an idea called Home Automation. Of late, crimes related to theft and home-intrusion is taking a toll on this city and this called for a better system for security, although the intrusion detection systems existing today made a huge difference, it is not affordable by most families.

So a home security surveillance system using Raspberry Pi and PIR sensors which is a low relatively low budget project is an effective way to battle the security issues at various sensitive area.

II. RELATED THEORY

P Reddy proposed the IOT based surveillance system to collect a wise remote control and seeing with embedded web server. They develop the system with cheaper cost, easy to use and with adaptable response for web enabled estimation and control system by using the key features.

S. Jadhav proposed reconnaissance framework which can be seen by proprietor remotely through android application. The framework related with IOT sends push notification to mobile device when interference is perceived inside the room. Simply endorsed customers can access to their system remotely by web interface using mobile phones. This system contains hardwired perception framework which controls security features of surveillance system. This system can be used in various circumstances and situation like bank lockers, homes, manufacturing unit monitoring etc.

Lu Lai has proposed framework to identify the human identification and to bypass the false programmed entryway framework activity for enhancing security features. This framework focuses on face localization and discovery.

Raspberry Pi 2 model B is used as the base system for operation of the home security system. Each of the other components is interfaced/connected to the Raspberry Pi. The USB web camera is activated only when the PIR sensor detects motion, and then the camera captures the image, later to be sent to the mobile application. PIR (Passive Infrared) sensor is a type of infrared sensor that will detect the presence of any living being by measuring the light radiating from objects in its field of view.

The timer module will act as a relay that measures a certain amount of time after which the camera is activated so as to make sure that the person trying to enter the house is not necessarily an intruder and maybe a known person.

This system makes use of cloud storage, and the platform used for cloud services is Microsoft Azure. This is done to save the onboard space in the Pi, and use Internet as a medium for interaction of devices.

The WIFI module is used to use internet connection to connect with the server/cloud, and the GSM module/ dongle is used to Connect with the user's mobile phone

III. PROPOSED METHODOLOGY

The working of the proposed system is simple. The idea is to allow the user to remotely keep track of the activities of his/her home using a simple mobile application.

1. Consider a visitor ringing the doorbell to the house.
2. This triggers the motion (PIR) sensors in the system.
3. Whenever a PIR sensor detects motion, it provides binary '1' as output to the system, after which the web camera interfaced with the Raspberry Pi is activated.
4. The camera then captures the image(s) of the person(s) detected by the sensors.
5. The image(s) undergoes face recognition algorithm.

In the image processing, For Face detection and Recognition, there is a built in function which is a part of the Image class called detect Faces, this function takes the image as input, and searches for faces within it; when faces are found, bounding rectangles are returned to a list of Python tuples. These tuples contain the x and y position of the match, and the width and height of its bounding rectangle.

- a. If the image(s) matches with any of those stored in the SD card, the process is aborted.
- b. Otherwise the image(s) is first compressed and uploaded to the cloud/server and eventually transmitted to the user's registered mobile via an android application.
6. The image sent to the cloud/server is then retrieved on the client side(user's mobile) using JSON and the Android page uses Async Task to allow the application to fetch data in the background.

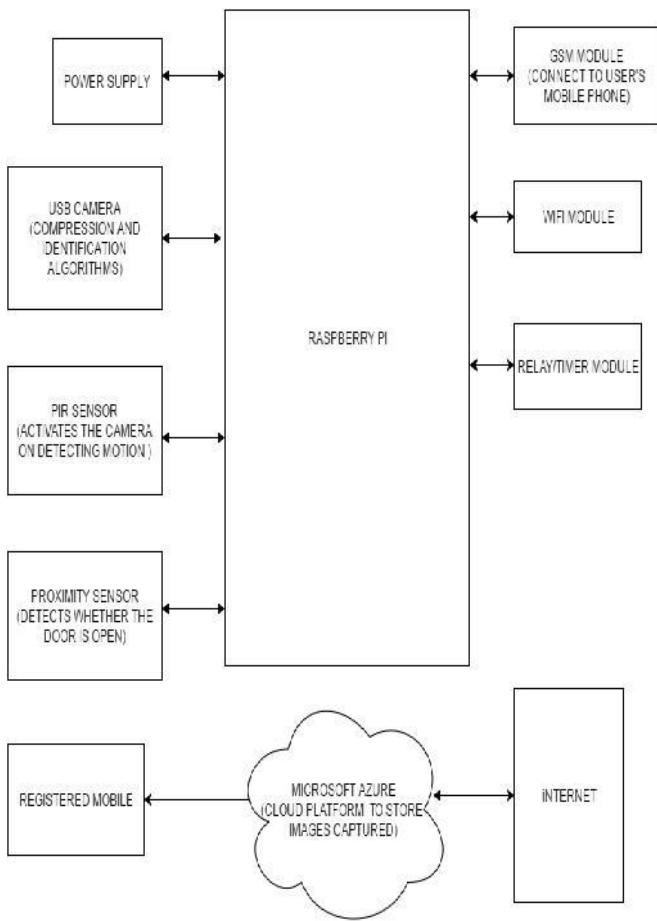


Fig 2 Working of the system

An exception to the normal working of this system would be if the door is open. In this case the proximity sensor which is a set of magnets attached to the door will sense if the door is open or closed by identifying a magnetic field in between them. When the door is open, even though the PIR sensor picks up motion in the hallway/premise, the camera will not be activated.

Here's what we will need:

- Laptop
- Raspberry Pi
- Card reader
- Micro-SD card
- Micro-USB cable
- Ethernet Cable
- Wireless Internet Connection

Steps to setup the Raspberry Pi:

- Download the latest Raspbian image from the Raspberry Pi website.
- unzip the downloaded Zip file and extracted image file (.img) is used for writing to the memory card.
- Insert the SD card into your card reader and plug it into your laptop.
- Win32DiskImager utility should be downloaded zip format. Then the executable is extracted and finally run the utility.
- Do the selection extracted image file

- Do the selection of drive letter assigned of SD card in the device box. (correct drive should be selected carefully) If SD card slot need be checked in the computer and if drive is not detected then external SD adapter need to be used.
- Click Write and wait for it to complete.
- Exit the imager and eject the SD card.

IV. RESULT AND DISCUSSION

The expected time of transmission for the images captured by the webcam without using compression algorithms is 1 minutes and 39 seconds if the image resolution is 512x512 which has an uncompressed size of 786 KB. The CPU processing time varies for each compression algorithm, shown in the following table:

Table 1: CPU processing time

Algorithm	PSNR values OF image (in dB)	CPU time	
		Encoding	Decoding
Wavelet	34.66	0.35 sec	0.27 sec
JPEG	31.73	0.12 sec	0.12 sec
VQ	29.28	2.45 sec	0.18 sec
Fractal	29.04	5.65 hrs	1.35 sec

The following table shows the size, transmission bandwidth and transmission time for all kinds of multimedia data such as image, audio and video data:

The face recognition algorithms when used to identify if the person in the picture is a known entity, is expected to detect as many faces as is present in front of the camera. This ensures complete security, and all the faces must be first verified with those existing in the database and then further processed for the routine procedure. The time required to first verify with the database and then transmit the image to the application should be as less as possible to enhance the performance of the system, because this system demands low latency in all the processes

Table 2: Experimental Results

Methods	Experiment 1		Experiment 2	
	Intel (seconds)	Arm (seconds)	Intel (seconds)	Arm (seconds)
Eigenfaces	1.91	30.65	432.08	3600
Fisherfaces	5.96	8.3	549.84	3360
LBPH	55.02	136.52	143.61	907.02

This table shows the timing performance of the given Face Recognition techniques. As seen in the table, in experiment 1, Eigenfaces scored the best of all three. But Fisherfaces also resulted good with 5.96 seconds. But overall, the LBPH method is

robust to change in the hardware architecture, especially in Arm platform LBPH is made the best results.

In the smart home security system, the sensors connected to the Raspberry Pi device help to detect the presence of an outsider, and the web camera attached to one of the USB ports gets activated and further captures the picture of the person(s) detected. The images are stored on the server or cloud which is then sent to the user who can monitor the system from anywhere using just a simple GUI on an Android application. So far with this system, the advantages are remote surveillance and reduction in ambiguity. This system can be improvised by addition of various different features to produce a powerful security system that maybe later used in high performance applications.

V. CONCLUSIONS

The system in discussion makes use of PIR and Hall Effect Proximity sensors to ensure high security which is the basic purpose of implementation of the system. This work includes the use of a smart GUI Android application to communicate with the standalone system from anywhere. This will also help in versatility and reduce the operation cost. The captured picture of the guest or intruder after face detection is sent to the user. The user can further forward the same photograph to the police station if he wishes.

Also, the proposed system can be improvised with the inclusion of Voice Pattern Analysis using advanced DSP Processors.

REFERENCES

- [1]. Aamir NZ, Mohamed S, "An Internet of Things approaches for motion detection using Raspberry-Pi",
- [2]. International Conference on intelligent computing and Internet of Things (ICIT) 2015.
- [3]. Sunil K, Rahul R, "An Internet of Things approach for motion detection using Raspberry", International Journal of Advanced Technology and Innovative Research 2016; 8:3622-3627.
- [4]. Priya BP, Viraj MC, "Smart Motion Detection System using Raspberry-PI",J. International Journal of Applied Information System (IJAIS) 2016; 10:37-40.
- [5]. Dhiraj Sunehra, Ayesha Bano "An Intelligent Surveillance with Cloud Storage for Home Security", 978-1-4799-5364-6/14 ©2014
- [6]. Baoan Lia, Jianjun Yub,a*, "Research and application on the smart home based on component technologies and Internet of Things", 1877-7058 © 2011 Published by Elsevier Ltd.
- [7]. Rudolf Volner, PhD., MUDr. Lubomir PouSek, "Intelligence Security Home Network" , 0-7803-7882-2/03©2003 IEEE
- [8]. Yi-Tsung CHIEN, Yea-Shuan HUANQ Sheng-Wen JENG, Yao-Hong TASI, Hung-Xin ZHAO, "A Real-time Security Surveillance System for Personal Authentication", 0-7803-7882-2/03©2003 IEEE
- [9]. Pavithra R, Usha R, George CC,"Scale Invariant Feature Transform based Face Recognition from a single sample per person",International Journal of Computational Engineering Research (IJCER) 2014; 4:41-47.
- [10]. Deepak M Shaila, "Face detection using genetic based SIFT algorithm", Green Computing, Communication and Conservation of Energy (ICGCE) 2013.
- [11]. Anil J, Padma S, "Literature Survey on Face and Face Expression Recognition", IEEE conference 2016.
- [12]. Huansheng N, Hong L, "Cyber-Physical-Social Based Security Architecture for Future Internet of Things",J. Scientific research 2012; 2:1-17.
- [13]. Helen D, "INFSO 0.4 Networked Enterprise & Rfid, INFSO g. 2 Micro & Nano Systems, working group Rfid of the ETPEPOSS", Internet of Things in 2020 Roadmap for the future 2008.
- [14]. Li YG, Jiang MF, "The Reinforcement of Communication Security of the Internet of Things in the Field of Intelligent Home through the Use of Middleware",Fourth International Symposium on Knowledge Acquisition and Modelling 2012.
- [15]. Ravi Kishore Kodali, Vishal Jain, Suvadeep Bose ,Lakshmi Boppana, "IoT based smart security and home automation system", 2016 International Conference on Computing, Communication and Automation (ICCCA)
- [16]. An Iot Based Integrated Home Security, Safety And Automation System,]<Http://Kecece.In/Project/Occaeat-Cupiditate/>

Event Management System

Rishabh Dubey

*Department of Information Technology
Thakur College of Engineering of Technology
Mumbai, Maharashtra, India
rishabhdubey020@gmail.com*

Ankit Gupta

*Department of Information Technology
Thakur College of Engineering of Technology
Mumbai, Maharashtra, India
ankitgupta9551@gmail.com*

Abstract—Development of the Android application is going in an exponential way. The basic of application is available to all net users. Android platform is now a major platform for developers and application(app) is major method to communicate and get information. The event management system app will be android based due its simplicity and popularity. As this application would be developed considering all the surveys and the problems related to it related earlier and fulfilling the necessary needs of the Admin/User. As NoSQL would be holding the data of the app which will be updated by the admin of the application. The android app will be with admin and user's (participant in event) which contain the basic GUI giving the information of all the events and enquiry. NoSQL is a relational database management system contained in a C programming library.

Keywords-

Android App, NoSQL, Event Management

I. INTRODUCTION

This android app is primarily aimed for custom events to meet the demand of sponsors and participants of event by organizer in given time. The present event system like C event event management, Event briteetc as per reference [6][7].

These apps are now popular for their performance and giving a satisfactory result to sponsors and participant peoples. The services they give are included in our system too but what we are giving is a custom event management and best performance. The system side feature is that it can generate performance and evaluation (individual of staff team and group of service providers) reports such as progress reports, merit lists, etc. This events management systems work in a pre-plan manner which is less time to consume and efficient which is very beneficial for nowadays fast life. It is an area that is growing rapidly in these modern era which is emerging at varied levels successfully. It is one of the most effective marketing communication tools that can create awareness of its organization, and also about its product. Events play

Pranay Rao

*Department of Information Technology
Thakur College of Engineering of Technology
Mumbai, Maharashtra, India
pranayrao96@gmail.com*

Amit Singh

*Department of Information Technology
Thakur College of Engineering of Technology
Mumbai, Maharashtra, India
fireonthunder@gmail.com*

an important role in our society. Any happening or an activity can be referred as an event. Individuals often find they lack the expertise and time to plan events themselves. Independent planners are needed to step in and give these special events the attention they deserve. In the current scenario, planning an event requires a lot of patience and hustle bustle right from deciding the theme to deciding venue and events. Lots of factors need to be considered while making each decision. Also once the party is planned lot of on the day issues such as maintaining low noise levels after a particular time, or neighbors complaining about the noise levels etc takes the fun out of the party/event.

II. EXISTING SYSTEM

Majority of events have many service providers to work simultaneously and it is very hard to manage these providers to schedule. It is also important for the event organizer to have contact information of service providing members of events for checking the progress of work or maybe a problem or new variable is to appear. To manage all these activities, we are developing this software.

To get affluence in the event management business, a user should have the strong connection to the service provider. To make a successful event, a manager needs different service provider like Sound systems services,

Lighting providers, Canteen services, stage construction, etc. In present event management system event organizer have to do all management work manually. Event organizers keep all details of events as payment information on papers. There is no analysis system software to check the past and present expenses on any events.

To do the analysis organizer need to check payment file logs or register and this task is actually wastage of time and tiresome which have become a burden for present event organizers. Keeping these all problem in check we are developed this event management system. This system helps the event manager to manage their paperwork on the app and they can also retrieve a report of past events organizer/user have done.

Application nowadays only reminds through SMS and people nowadays do not check message inbox. In some of the applications even unauthorized user can disseminate information and do not provide proper hierarchy of authentication and also no instruction or manual for layman user as per reference [1][2][3][4].

PROPOSED SOLUTION

This paper 'Event Management System for College' is Android app and is alternative to existing manual process which needs more time and effort. This system is automated and will provide features of better management, monitoring, and maintenance of the resolving process. It will ease the entire manual process and lower the effort and work needed to make a great event and give satisfaction to all participants of events. Our system is concerned with college, which will take event information that to be organize and will be managed by an administrator or organizer person.

This system is automated maintaining detail about users (admin/participant) and feedback can be retrieved easily from user and will prove to be an important aspect to administrator. As our paper is based on an app, the user can check it out through their mobile and just have to login to the app and search for the relative application. This isn't time consuming. Normally and legally organizer have to go to the department of admin have some forms filled up probably there are some transaction too which is the only manual work to be done while using this app. By using this app, event organizer can use their Smartphone to get the updates and contact if some problem arises and resolved faster.

Problems regarding conference events are present in any college is major part for development of institute. Nowadays organizers face problem relating to time. In the sense, they have to go to the admin office to lodge any problem they need to for better growth of institute and wait for their response. This tends to waste of time. So taking this issue in consideration we thought of developing an application which will be useful to the organizers.

The main purpose of our paper is to save the time, work in coherent manner, efficiently using resources. Through the analysis we found that there is need to develop such an application will be used to the organizer and this motivated us to develop such an application which will fulfill their required needs.

The main objective is to reach to the participants and help them to get their conferences. The second main objective is to let organizer know the problem of any room or facility that is lagging and get it solved.

Planning and Analysis, provides basic process and analytic frameworks to support Corporate/Business Unit managers in both their day-to-day management tasks and their quarterly/annual financial management responsibilities. Decision is made of how best way to manage the process to provide the optimal solution.

IV. METHODOLOGY

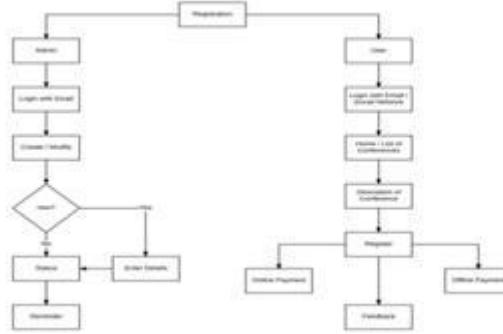


Fig. 1: Methodology

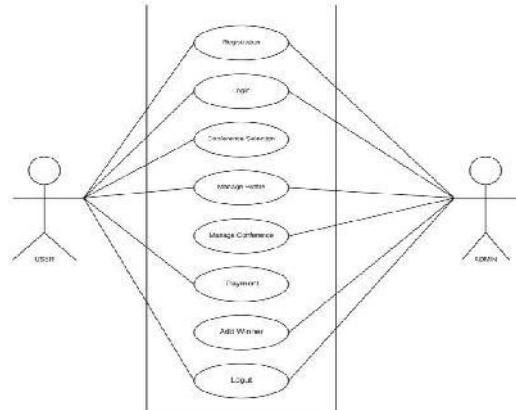


Fig. 2 : Use Case

V. RESULT

The Application is built on an android platform and implemented accordingly to android architecture. The system will work on all android devices with android version 5.1 and higher. System will have an interface where participant can login, if want can sign up in system as well.

Then event and its facilities will be managed by the administrator he can view, respond or delete and is completed or resolved. This system will replace existing manual process and will be completed automated having better management, and monitoring of complaints. A database will help to view registration and maintain record of the participants. It will ease the process of event and reduce the time and effort, increasing the reliability and efficiency of organization. The planning of using resource in an efficient way and without wastage of resource and time. Gathering the necessary information and analyzing the information for finding the research gaps. The integration of database to frontend app is perfectly usable and designing of app frame is focused in its great look. The software code will be developed in such a way that it will be able to bridge all the gaps in the previous proposed solutions. The integrated database and app will be functioning as desired by the client. The proposed system will

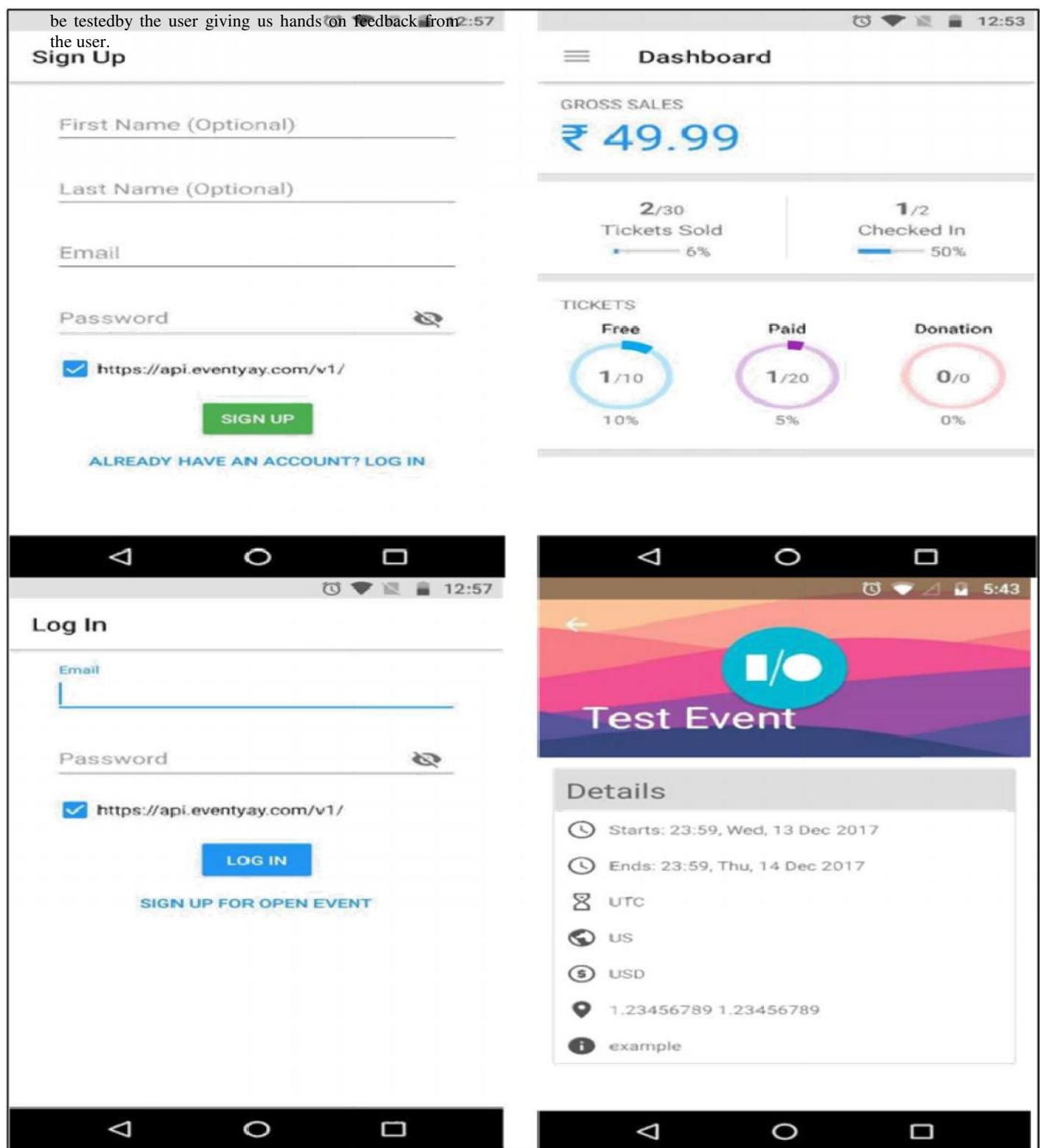


Fig 3: User Interfaces

VI. SCOPE

The event management app is made to use online database as data center for collection and using it by NoSQL to make data useful for event management and records. The application is using master-slave method as master will be one an only administrator or core team and slave will be user's app. The system being design is economically with respect to the students and other users point of view.

To further develop and enhanced user information management system that can help solve drawbacks of existing system. It helps us in current all works relative to Event Management System. It will also reduce the cost of collecting the management and collection procedure will go on smoothly.

This Event management system app can be used by normal android mobile user. It provides secure Administrative panel. This application can be used for any event. The participants have to register for these events before the deadline. This application best feature is to having strong network with service providers.

We have left all the options open so that if there is any other future requirement in the system by the admin for the enhancement of the system then it is possible to implement them as per reference [10]. Such as the following points mentioned are the possible facilities can be added in application in future.

We can give more advance software for event management system including more facilities. We will host the platform on online servers to make it accessible worldwide. Integrate multiple load balancers to distribute the loads of the system.

VII. CONCLUSION

The system offers reliability, time savings and easy control. Also users can view details, notifications anywhere and anytime. The application will greatly simplify and speed up the result preparation and management process. It provides high security and a system that reduces the work and resources required in traditional process.

The proposed system provides the new way of computing and displaying an operations with responsive and attractive user-

Application development is defined as a very useful and efficient development that fulfills the required needs. Our team is working on developing an application in android because Android as an OS, currently developed by Google, based on Linux kernel and is designed primarily for Smart phones and Tablets.

Firebase, a NoSQL database is used to store and retrieve data. Firebase DB uses JSON indexing format. It helps us develop high-quality apps quickly.

Our project is only a humble venture to satisfy the needs to manage event. Several user friendly coding have also adopted this package shall prove to be powerful in satisfying all the requirements of the college. The objective of event management is to provide a framework that enables the

manager to make reasonable estimates and work progress within a limited frame of time.

Thus by this we conclude that the proposed app (event management system) custom organizing can be used effectively by organization. Also not only organizer but a normal man can also use as it is simple to understand. This app making lead us to understanding of a good understanding of management and its resourceful way of using.

VIII. ACKNOWLEDGMENT

We sincerely thank to our guide Ms. SangeetaVhatkar for her guidance and support for carrying out the need project work.

We also thank our project coordinators Mr. Aaditya Desai and Ms. Neha Kapadia for arranging all the necessary facilities to carry out the project work.

We also thank our HOD Dr. Rajesh S. Bansode and Principal of our college Dr. B.K Mishra and the college management for their handful of support in our cause.

REFERENCES

- [1]. MallapaGurav, PrashantBiradar "Andi e-resources application for students based on android" IRJET vol. 3 issue 7 July 2016.
- [2]. SaipriyaSairam, Shruti Suresh "Event Management an Android Application", IJIRSET vol. 5 issue 3 March 2016.
- [3]. KinjalModi, Unnati Chauhan "Greeting Reminder Application Based Android" ijrcce vol. 1 issue 2 April 2013.
- [4]. A.J. Kadam, Aradhana Singh "Mobile web based android application for college management system" ijebs vol. 6 issue 2 Feb. 2017.
- [5]. <http://developer.android.com/index.html>
- [6]. <https://www.eventbrite.com>
- [7]. <https://www.cvent.com>
- [8]. <https://theeventscalendar.com/support/forums/forum/event-tickets/>
- [9]. <https://www.simplifiedcoding.net/android-application-development-tutorial-from-scratch>
- [10]. https://www.scribd.com/document/351140927/Event-Management-System#fullscreen&from_embed
- [11]. Android SDK|Android Developers <http://developer.android.com/sdk/index.html>
- [12]. <https://www.tutorialspoint.com/java>
- [13]. <https://www.tutorialspoint.com/nosql>
- [14]. <http://www.jdbc-tutorial.com/>

3-Tier Website As a SAAS Model

Ruth Baptista

*Department of Information Technology
Thakur College of Engineering and Technology*

Deven Shah

*Department of Information Technology
Thakur College of Engineering and Technology*

Abstract— Abstract- Software- as- a -Service (SaaS) is an application delivery model that enables users to utilize a software solution over the Internet, thereby eliminating the need to download and install the application on the user's own computer thus the user's burden of maintenance, ongoing operation, and support is eliminated too. Large number of SaaS solutions are based on multi-tenant architecture. In this model, a single version of the application, with a unique single configuration is used for every user ("tenants"). The project idea lies for proposing a model without the user intervention. Any user will be able to access and modify the existing web application for their own use with a unique URL specific to the user. The proposed website will be offering a service to various dentist who wish to use the website on a pay per use basis.

Keywords— Design pattern, Software-as-a-service, Tenants, User intervention .

I. INTRODUCTION

3 Tier Web Architecture is a unique system of developing web database application that works around all the 3 tier model. This comprehensive 3 tier architecture module is the framework for most Web Applications on the Internet.[1] This architecture of the system helps to separate the Business Logic from the Application, Data Storage and database.

While migrating the non-SaaS applications into SaaS application, there will be certain issues that needs focus, such as database architecture, data partitioning, UI customization, scalability issues, and workflow management.[1] The idea for migrating a three tier application into a saas application is surfaced from design pattern. In a software design pattern it is a general, reusable solution to a commonly occurring problem within a given context in software design. It is not a finished design that can be transformed directly into source. It is a description or template for how to solve a problem that can be used in many different situate

The concentration of the project lies in the process where an existing three tier web application for dentist is to be transformed or converted to a multi-tenant SaaS application. This transformation process and changing of the home page according to the user login is the real crux of the project and implementation of the proposed framework to achieve. The variables involved in the project are essentially centered

towards the requirements of the dentist and his patients. Data Sensitivity of the variables involved will shape the approach adopted towards the Database module and the level of customization in terms of the UI and Configuration module. The website will be offering a service to various dentist who wish to use the website on a pay per use basis. The main goal is to provide a look and feel to the dentist as their own website with a unique url and a logo while managing the patients and their records.

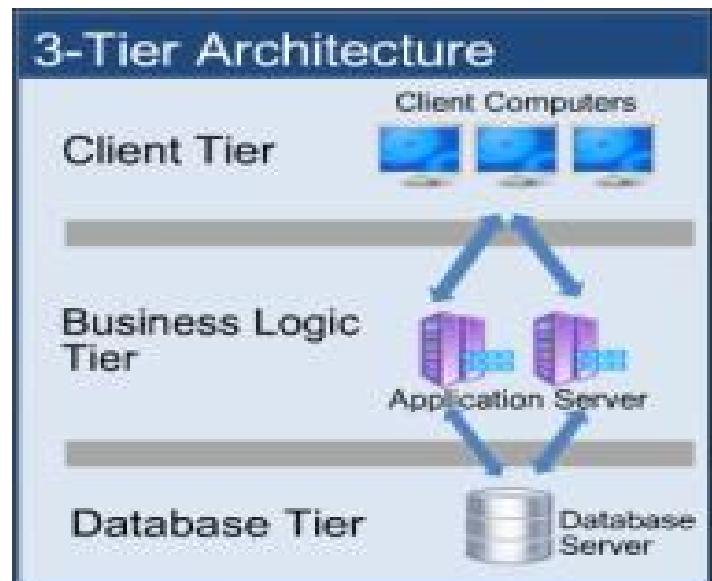


Fig 1: Traditional three tier architecture

Thus, providing a framework to migrate a traditional three tier web applications into multi-tenant SaaS.[9] The framework provides a detailed overview of the proposed system architecture that will helps software architects and developers to migrate their applications into a saas model.

The main focus of the paper is the idea for converting a three tier web application to a saas based application, mainly, the frontend customization, as well as the business logic configuration and database configuration for the framework. This paper is organized as the following: Section II discusses the related work, Section IV discusses the customization of front end, Section V discusses business logic

layer, Section VI discusses database layer for the application and finally Section VII discusses the concludes the paper and outlines the future work.

II . RELATED WORK

In past few years there is large amount of research that has been carried for SaaS and multi-tenancy. But there are only a few who proposed a complete framework for migrating traditional web applications into SaaS application.

The ASP concept had goals totally the same to the goals of the SaaS model and motivated multi-tenancy. However, these approaches focused upon giving out an instance per tenant, and they slip asleep the level 1 of maturity model proposed by Chong et al. Data multi-tenancy is the most explored right of entry below multi-tenancy, and is often implemented upon peak of a database. Both Jacobs et al and Chong have outlined three main approaches for data outlook in a multitenant deployment- (a) surgically remove databases: gives each tenant its own database, (b) shared database when surgically remove schemas: gives each tenant its own tables and (c) shared database along in the middle of shared schemas: shares the same table in the middle of many tenants and enforces security at the adjacent lump in the architecture. Much of the existing research upon multitenant SaaS have focused upon shared data architecture and security running and middleware extensions to habitat the expertly-founded concerns due to data/security.

III. FRAMEWORK INTRODUCTION

3-tier application architecture is characterized by the functional decomposition of applications, service components, and their distributed deployment, providing improved scalability, availability, manageability, and resource utilization. During an application's life cycle, the three-tier approach provides benefits such as reusability, flexibility, manageability, maintainability, and scalability.[14] Each tier is completely independent from all other tiers, except for those immediately above and below it. You can share and reuse the components and services you create, and you can distribute them across a network of computers as needed. You can divide large and complex projects into simpler projects and assign them to different programmers or programming teams. You can also deploy components and services on a server to help keep up with changes, and you can redeploy them as growth of the application's user base, data, and transaction volume increases.

Logic layer is moved outside the presentation layer and into the business layer as it enhances reuse. As applications grow, applications often grow into other realms. Applications may start out as a web application, but some of the functionality may later be moved to a smart client application. Portions of an application may be split between a web site and a web or windows service that runs on a server.[14] In addition, keeping logic helps aid in developing a good design (sometimes code can get sloppier in the UI).

This framework is proposed to convert any traditional web architecture to SaaS model. This architecture is proposed in "A Framework to Convert Traditional Application to SaaS Model" by Eyad Saleh, Nuhad Shaabani, and Christoph Meinelasso-Plattner-Institute University of Potsdam, Germany.

We are going to follow the proposed architecture to convert our 3-tier traditional website to cloud based website and offer it as a service to our tenants or clients. Multi-tenancy is different [1] from multi-instance architecture (Figure 2) where separated instances of the same software are hosted on different servers to serve different tenants.



Fig 2: Multi-instance vs. Multi-Tenant

Multi-tenancy is the principle of running a single instance of the software on a server to serve multiple companies (tenants). Re-engineering traditional web applications from scratch into multi-tenancy requires tremendous efforts in terms of cost, manpower, and time. Thus, we are using a framework to migrate traditional web applications into multi-tenant SaaS. The framework provides a detailed overview of the proposed multi-tenant architecture that helps software architects and developers like us to migrate their applications into multi-tenancy.

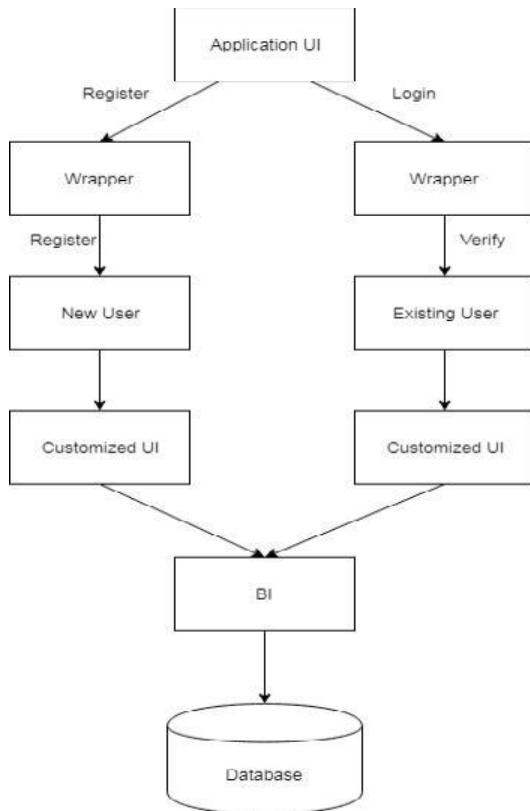


Fig 3: The process flow

Tenant administration must be role based. In a multitenant environment, the tenant admin is able to distribute the management and access capabilities to appropriate administrators within the tenant. Multiple web tenant enables a multiple instances of your application to serve thousands of your vendors. Single web tenant to serve thousands of your customers.

Multi-tenant data model designs added which become smart configurations with SaaS Tenant.[13]

Provide option form a single coupled database, specific database schema per web-tenant or an isolated database per web-tenant.

The process flow of the migrated application will be as follows[1]:

- a) A user belonging to a certain tenant logs into the system by entering his username and password.
- b) A dedicated authentication module is used to map this user to the tenant he belongs to, to create a token for the tenant including the Tenant-ID as well as other relevant information (such as locale settings), and finally to pass it to the customization layer.[12]
- c) In the customization and configuration layer, the UI components, such as logos and colors, the business logic, and workflow configuration data for this tenant are restored, and passed to the application server.
- d) The DB configuration data will be passed to the DBserv-

- er for query transformation.
- e) The application server receives the above specified data from the upper layer and passes it to the run-time customization engine, which integrates all components and lunches the application instance.
- f) A log service is used to record the application actions and store them in text files.
- g) A dedicated monitoring service is used to monitor the performance and status of the application, and detects any faults or bad resource usage.

IV. PRESENTATION LAYER (CUSTOMIZATION OF FRONTEND)

The 3-tier Web Architecture is designed to provide a higher degree of flexibility and increased security that can be designed for each service at each level. This unique system of framework for web application development with the 3tier web architecture also ensures that there is increased performance as the task is shared between servers. Web 3-Tier architecture is a connection and composition of the three links that facilitates the smooth functioning of the website.

The presentation layer implements the "look and feel" of an application. It is responsible for the presentation of data to the users, receiving user events and controlling the user interface. Most ecommerce applications are web-based[11]. Web-based applications now often contain most of the data manipulation features that traditional applications use. This is accomplished through use of Dynamic HTML and client-side data sources. The programming languages used are the combination of HTML, CSS and Javascript. JSP or ASP are used for dynamic content.

- HTML is a Web authoring markup language for defining content structures and rendering a web page.
- Javascript is commonly used for client-side validation. Javascript does have some control over the look-and-feel of a page in dynamic HTML.

The proposed system's presentation layer will be same as the 3-tier architecture with a little modification in the presentation layer including the idea from design pattern. This layer is presented same as it is with a wrapper on the top which will be asking for the user details. After accepting this user details the UI for application will change to the name and logo of that particular user, this gives the look and feel of using their own website. The user (dentist) when registered will be provided with a unique url where the user can update his data and add patients data. This unique url can then be provided to other users (patients) to register or book a appointment.

V. BUSINESS LOGIC LAYER

The logic tier is pulled out from the presentation tier and, as its own layer; it controls an application's functionality by performing

detailed processing. Logic tier is where mission-critical business problems are solved. The components that make up this layer can exist on a server machine, to assist in resource sharing.[12] These components can be used to enforce business rules, such as business algorithms and legal or governmental regulations, and data rules, which are designed to keep the data structures consistent within either specific or multiple databases. Because these middle-tier components are not tied to a specific client, they can be used by all applications and can be moved to different locations, as response time and other rules require. For example, simple edits can be placed on the client side to minimize network round-trips, or data rules can be placed in stored procedures.

In the logic layer the logic for data retrieval and authentication will be applied according to which the data will be displayed, hence giving the access to add data about the patients.

Authentication Module:

- The authentication module has two parts which includes tenant register process and user login process. The tenant register process is as flow:
- Tenant registers in SaaS application,
- SaaS administrator checks the tenant information. If the tenant information is passed, the SaaS application will generate an account for the tenant;
- The tenant tries to use SaaS application with the generated account and he can select the application functions according to his need.
- The tenant can create user accounts and assign permissions for his enterprise's users, who can use the generated account to access the application;
- If the tenant is satisfied with the SaaS application, he can pay for the application functions needed;[15]
- As soon as the SaaS application receives the rent, the tenant will change to official users. Certainly, apart from differences from register, the multi-tenant application is much different from typical web application in login process. The multi-tenant application login process is shown as flow:
- User inputs his username and password.
- SaaS application verifies the user's input.
- If the user passed the verification, the multi-tenant application can get the user's tenant information through the userID and store the userID and tenantID into session variable.
- Users access the SaaS application.
- Then, the user's role information is got.
- The user's permission is got through his role.
- User gets his using environment and uses the application.

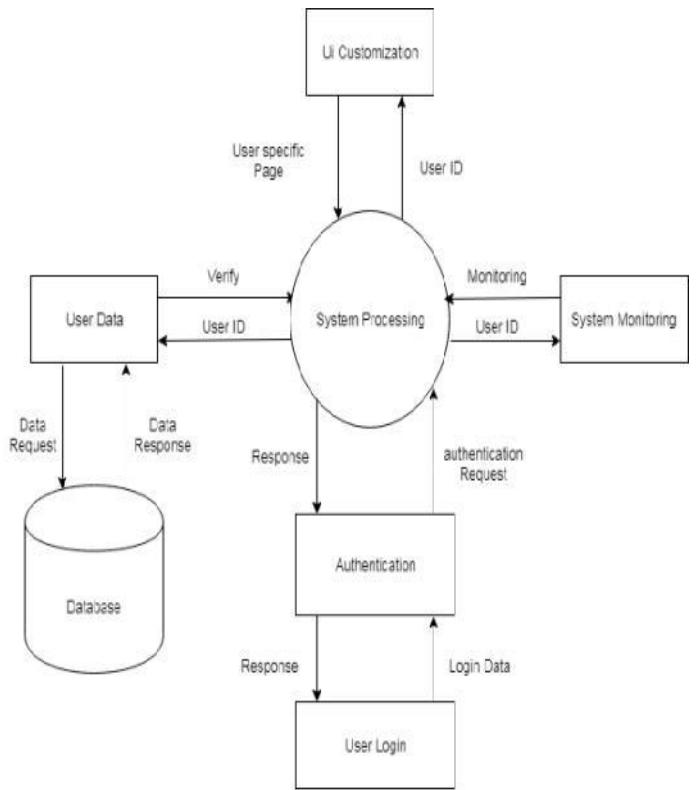


Fig 4: Authentication data flow

From the above process we can see that authentication module is responsible for the tenant management and rights management for all tenants. According to user's TenantID, authentication module can navigate user to the appropriate tenant environment. In this module, tenant information management functions need to be added to migration platform and the original web login module need to do a small amount of code modification.

VII . DATABASE LAYER (CUSTOMIZATION OF BACKEND)

This tier consists of database servers, is the actual DBMS access layer. It can be accessed through the business services layer and on occasion by the user services layer. Here information is stored and retrieved. This tier keeps data neutral and independent from application servers or business logic. Giving data its own tier also improves scalability and performance.[14]

This layer consists of data access components (rather than raw DBMS connections) to aid in resource sharing and to allow clients to be configured without installing the DBMS libraries and ODBC drivers on each client. An example would be a computer hosting a database management system (DBMS), such as a Microsoft SQL Server database.

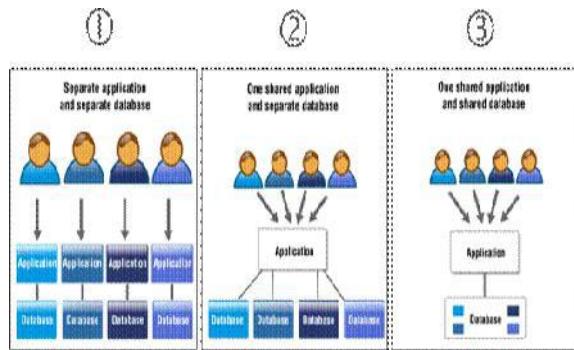


Fig 5: Database re-configuration

In this layer the data will be partitioned according to the user, each time a new user registers a new user will be created and their data will be stored in that users. Each user will have only rights to access its own tables where its patient data will be saved.

Database transformation module includes database transformation function and tenant filter function, the database transformation function is charged for transforming original system database into multi-tenant model and the tenant filter function is to filter user operation and data access.

In multi-tenant SaaS system development, database is commonly designed for three models: separate database, shared database independent schema and shared database shared schema which have different methods to meet the SaaS application requirement.[12]

Each tenant's data stores in an independent database and each tenant shares the same application. In this model, It is a simplest storage way and is convenient for user to backup and restore their own data and has a good data isolation, but the shortcoming is occupying too many resources. This model adapt to users who need high security but not focus on cost, such as bank, hospital etc.

All tenants store their data in a shared database, but each tenant has dependent schema, that means every tenant have a dependent set of table. Despite the advances in virtualization and cloud provisioning, there are inherent problems in the ability to scale up this model as more customers come on board.

In this database model, all tenants store data in a same database schema. A column tenantID was added to each relation table to distinguish tenant to others. This model has the highest data sharing and lowest data isolation and has lowest hardware and maintenance cost. In our transformation platform, we use the third database model. The database transformation module is inserted between presentation layer and database layer. Firstly, let us look up the Database transformation function.

Extraction and Generation Meta database schema. database transformation module reads the database schema from original system and creates Meta database schema.

Generation multi-tenant database. According to different user requirements, database transformation module can generate multi-tenant database with Meta database schema by adding corresponding management tables and inserting “TenantID” column to each table to distinguish various tenants' data. Because of only adding “TenantID” column to every table in multi-tenant database, the old web system can work correctly and needn't any changes in business process. Owing to the database transformation module using Meta database schema, user can select to generate various kinds of multi-tenant database such as Oracle, DB2, My Sql and SQL Server database.

Modification database connection configuration. In order to make the original system connecting to new database correctly, we must modify the database connection configuration file.

Operation filters. In SaaS software model, users can only access data belonging to their own tenant. In order to achieve this goal, “TenantID” column is used to distinguish each tenant's data. DTM can shield other tenant' data by data filter function when users access the application by operation such as inserting, updating, deleting and browsing.

Database transformation module obtains the SQL sentence submitting to the new multi-tenant database from the application. It parses the SQL sentence and adds filter condition such as “TenantID=LoginTenantID” which will guarantee the data security in shared multi-tenant database and ensure the user can only access his own tenant's information.

VIII . CONCLUSION

The key aspect of using software as a service is multi-tenancy. Migrating traditional web applications into multi-tenant SaaS poses several research challenges in terms of software customization, database architecture, and isolation between tenants. That is using a single resource by many users or sharing the resource. Developing a new web application from scratch is more easy than re-engineering the already existing web application, doing such change requires lot of time, money and man-power. It is difficult to re configure the entire present system.

Our Proposed system thus provides a alternative of wrapper for the web application which will not make much changes to the entire system. We propose a model to simplify and speed up the application . The application can be used by many dentist who practice dental study and cannot afford their own website.

IX. FUTURE WORK

The work of proposed system would like to be extended with the idea of developing SaaS applications to provide better service to clients. There can be various modifications that can be made to the SaaS project to make it highly scalable, increase performance and depending on the sensitivity of the data which will be stored various database approaches could be adopted. Endeavouring to help others come up with their own SaaS business where existing traditional applications can be reengineered and given a face lift to generate monetary advances.

The SaaS model can be applied in a wide-range of domains and there is endless prospect in terms of SaaS businesses. As SaaS gives companies an alternative, they can plug in and subscribe to services built on shared infrastructure via the Internet. The SaaS models are yet to flourished in few years because of the many benefits it offers to businesses of all sizes and types.

- [12] Craig D. Weissman and Steve Bobrowski, "The design of the force.com multitenant internet application development platform", In Proc. of the SIGMOD, 2009,pp. 889-896.
- [13] Stephan Ried, "Forrester's SaaS Maturity Model: Transforming Vendor Strategy while Managing Customer Expectations," <http://www.forrester.com/Research/Document/Excerpt/0,7211,46817,00.html>, 2008
- [14] Concepts-of-three-tier-architecture <https://alitarhini.wordpress.com/2011/01/22/concepts-of-three-tier-architecture/>
- [15] D. Jacobs and S. Aulbach, "Ruminations on multitenant databases," BTW Proceedings, 2007.
- [16] F. Chong, G. Carraro, and R. Wolter, "Multi-Tenant Data Architecture," MSDN Library, Microsoft Corporation, 2006.

REFERENCES

- [1] Eyad Saleh, Nuhad Shaabani, and Christoph Meinel "A Framework To Convert Traditional Application To SaaS Model", Hasso-Plattner-Institute University of Potsdam, Germany, INFOCOMP 2012.
- [2] Jianbo Zheng and Weichang Du "Cloud Services for Deploying Client-Server Applications to SaaS", Faculty of Computer Science, University of New Brunswick, Fredericton, Canada
- [3] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia "Above the clouds: A Berkeley view of cloud computing", Technical report, University of California, Berkeley, USA, 2009.
- [4] C. Guo et al. "A Framework for Native Multi-Tenancy Application Development and Management", 9th IEEE Intl. Conf. on E-Commerce Technology and 4th IEEE Intl. Conference on Enterprise Computing, Ecommerce and E-Services (CEC-EEE), 2007.
- [5] Manoj A. Thomas, Richard T. Redmond, H. Roland Weistroffer "Moving To The Cloud: Transitioning From Client-Server To Service Architecture", Virginia Commonwealth University, USA
- [6] Nadir K.Salih, Tianyi Zang "Modeling and Self-Configuring SaaS Application", School of Computer Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, China
- [7] Ekasari Nugraheni "Migration of Web Application SIMA into Multi-tenant SaaS" Research Center for Informatics, Indonesian Institute of Sciences Jl. Sangkuriang 21/154D, Bandung 40135, Indonesia
- [8] Yangpeng Zhu "A Platform for Changing legacy Application to Multi-tenant Mode", International Journal of Multimedia and Ubiquitous Engineering Vol.9. No.8(2014). pp.407-418
- [9] Software as a Service, http://en.wikipedia.org/wiki/Software_as_a_service
- [10] Cor-Paul Bezemer, Andy Zaidman "Challenges of Reengineering into Multi-Tenant SaaS Applications" Delft University of Technology Software Engineering Research Group Technical Report Series
- [11] Ganesh Olekar,Vikram Sreekumar "Cloud Computing: Migration from Traditional Systems to the Cloud" International Journal of Advanced Research in Computer Engineering & Technology (IJAR-CET) Volume 2, Issue 3, March 2013

Comparative Study of Project Management Software

Vandana Munde

*Thakur College of Engineering & Technology
Kandivali East, Mumbai*

Vijaykumar yele

*Thakur College of Engineering & Technology
Kandivali East, Mumbai*

Abstract—Projects role is important in the financial development of a country. Companies have been investing large amount of money in projects related to various areas to gain more profit so these projects are designed with the aim of efficient planning and implementation, for earning adequate return on investment with their own resources. Managing project development involves techniques and skills that are proprietary to the IT industry. This study discuss the importance of project and Project management and also describes the importance of project management software for the project success. The study focus on popular project management tools in the market and compared with various aspects of project management software useful for the industry.

Keywords-: *project, project management, project management software, criteria, RationalPlan ,Zoho,Basecamp,*

I. INTRODUCTION

Project Management is a process which includes planning and implementation various Tasks. There are several project management tools that are used in the process of managing a project. Since there are many techniques and tools some of them are selected and analyzed. This is to help managers in making a decisions to choose a tool to use depending on their needs. The purpose of this study is to research on the most used project management software . The softwares that are reviewed are the currently being used. General information on each of the selected software will be discussed with emphasis on their key points, advantages and disadvantages.

A. Project

Project is a powerful application that helps you plan and manage a wide range of projects. From meeting crucial deadlines and budgets to selecting the right resources, you can be more productive and realize better results using the set of features Project offers. A project is a well defined task, which a collection of several operations is done in order to achieve a goal every project may have a unique and distinct goal. Project is not a routine activity or day to day operations. Project comes with a start time and end

time. Projects can also be viewed in terms of their attributes as time frame, purpose, ownership, resources, roles, risks and assumptions, interdependent tasks, organizational change, and operating in an environment larger than the project itself. Project ends when its goal is achieved and hence it is a temporary phase in the lifetime of an organization. Project needs adequate resources in terms of time, manpower, finance, and material and knowledge bank[1].

B. Project Management

The idea of project management has been around for a long, long time. Today, project management has emerged as its own field. Project Management is the process of defining, the application of knowledge, skills, tools and techniques Planning ,organizing, leading and controlling the development of an Information System Project[2].The goal of Project Management is to deliver an Information System that is acceptable to users and is developed on time and within budget The effectiveness of project management is critical in assuring the success of any substantial activity. Project

Management methodology provides a structure for understanding how projects are initiated, conceptualized, planned, carried out, terminated, and evaluated. Now adays Traditional approach of project management is supported by a body of knowledge and research across many discipline. Project Management methodology provides a structure for understanding how projects are initiated, conceptualized, planned, carried out, terminated, and evaluated

C. Project Management Software

Project management software is designed to make the job of a project manager easier and more efficient, providing applications to aid in planning, to manage project costs, and to track activities and monitor schedules. As more and more public works departments face the realities of increasing workloads and shrinking resources, finding technology applications that allow productivity gains

becomes ever more important. The use of project management software as a tool for managing and organizing work has grown and continues to grow at a rapid pace in all industries. There are many project management software used by industry and project management team should be aware of about latest technology tools used for managing project effectively. The project management theories, tools and techniques were widely used by different industries and organizations long ago .

The project management software need to have strong planning and monitoring capabilities if there is lack of input at a planning phase and also if the project is working on a tight budget. Projects are looked differently by different people at various positions. Project management software has the capacity to help plan, organize, and manage resource tools.

D. Project Management Software

Project management software is designed to make the job of a project manager easier and more efficient, providing applications to aid in planning, to manage project costs, and to track activities and monitor schedules. As more and more public works departments face the realities of increasing workloads and shrinking resources, finding technology applications that allow productivity gains becomes ever more important. The use of project management software as a tool for managing and organizing work has grown and continues to grow at a rapid pace in all industries. There are many project management software used by industry and project management team should be aware of about latest technology tools used for managing project effectively. The project management theories, tools and techniques were widely used by different industries and organizations long ago .

The project management software need to have strong planning and monitoring capabilities if there is lack of input at a planning phase and also if the project is working on a tight budget. Projects are looked differently by different people at various positions. Project management software has the capacity to help plan, organize, and manage resource tools and management and documentation or administration system. administration systems.

II. NEED OF PROJECT MANAGEMENT SOFTWARE

Technology is improving every day, there are always some new inventions, some new upgrades to the older one. And with these changes, there is a change in the way people work in their daily life, they have new demands.

Companies have to do anything they can, to satisfy their customer, because there are many other competitor companies in the market and they would not want to lose their existing customers. The trend has been to try to hold onto the existing customers rather than to look for new customers. The competition has not been to get new customers but not to lose old customers to new companies.

Customers' demands are ever changing, so it is not an easy task for companies to cope with the changing demands, it requires planning; planning for the future [3]. This is when companies started to realize how project and project management are important in winning the race against time, while it was seen as an option before, it is now considered to be a must.

The traditional tools can still be used to run the project, but for the success of the project, the new more improved approaches should be taken and the most successful companies are taking the most of the ever-growing technology and expertise to manage their project; the technology that helps to plan, organize and manage resources, and develop resources estimates. Scheduling, cost control and budget management, resource allocation, collaboration software, communication, decision-making, quality management and documentation, all are included in the project management software depending upon the software.

A. RationalPlan

RationalPlan is a project management platform that puts the emphasis on improving productivity, simplified use and fast learning. It was developed to address various types of users, ranging from novice project managers, users that require an environment to manage multiple projects and up to enterprises that are in need of a distributed solution with concurrent access. With RationalPlan, project managers and team members will find it easy to build project plans that are consistent with their requirements and end goals, effectively manage resources and evaluate workload, monitor project progress, estimate expenditures, and track and manage budget. Benefits of RationalPlan presents include effortless management and monitoring of projects, budget, and resources and undo or redo steps on multiple levels. Team members can build and implement project plans immediately and plan schedules and quickly identify issues & potential problems that might hamper a project success[5].

B. Zoho

Zoho Projects is the project management software which provides features like Gantt Charts, project coordination to collaborate, Time management, Report tools and Bug

tracker.Zoho provides various free services along with a fairly simple UI. It also provides integration with Google, giving access to the Google apps. With Zoho Projects, you can get work done on time, and use its milestones, tasks, and tasklists to plan your work in advance. What the system does is in fact very simple: it divides your large and complex projects into manageable units, and schedules recurring tasks, dependencies, and subtasks according to your deadlines. It is also packed with a variety of seamless collaboration tools thanks to which your employees and consultants can stay in line with your progress. Zoho Projects gives access to in-depth insights displayed in understandable charts and graphs, when projects are accomplished, so that team could report on your activities, track opportunities, and capture eventual drawbacks. The pricing scheme is flexible, and there is even a free plan for freelancers and small teams[6].

C. Basecamp

Basecamp is one of the most popular online project management suites today and branded as a reliable service that offers simple setup and short learning time. setting up an account and jumping right into project management is quite easy with this software, enabling users to organize people, delegate tasks, and monitor progress right from the get go. Now, serving more than millions of users and still being a popular pick for over 10 years, Basecamp clearly is a top choice when it comes to online project management. Most beneficial feature of Basecamp is flexible pricing scheme, where no company will be charged per user, and all clients will have access to the same features, depending on the purpose for which they're using this system. Enterprise-grade users are encouraged to contact the company and ask for an individual quote, and cover usage per year instead of paying monthly installments.

III. COMPARATIVE STUDY

Table 1.below provides the summary of project management softwares discussed in this article. The various project softwares like Basecamp,Rationalplan, Zoho are compared on the basis of different parameters like Scheduling ,budget management the programming language used, issue tracking, report generation.

Table1. Comparison of Project Management software

Parameters	Software		
	Basecap	Rational Plan	Zoho
Scheduling	No	Yes	Yes
Resource Management	Yes	Yes	Yes
Document Management	Yes	Yes	No
Issue	No	Yes	Yes

Tracking System			
Collaborative Software	Yes	Yes	Yes
Workflow system	No	Yes	Yes
Reporting & Analyses	No	Yes	Yes
Budget Management	No	Yes	Yes
Time Tracking	No	Yes	No
Invoicing	No	Yes	No
Integration	Dropbox, Github , Iphone App	Microsoft Project , Google Drive, Dropbox ,SkyDrive	MS outlook, Office, Google apps, Google Docs

IV. CONCLUSION

Project management software are useful for the success of the project. Project managers should select software that they are familiar with and is most suitable for the project. This paper has only looked at three software with much focus on the features part. Features like reporting & analysis, time tracking, invoicing implementation of meeting strategy, risk and cultural profiles are the features offered by limited number of the existing tools. some tools are strong in PM activities whereas some tools are strong at collaboration of teams. So for project management activities some tools covers scheduling resource management, some covers budget management, some covers issue tracking and visualization. There is not a single tool offering all project Management activities collectively. Project management software can be selected depending on the requirement of the project .The study of selected software is done which is the ground work for the further research which include number of software . Further technologies have rapidly changed and developed, which has resulted in shifting of requirements, and definitions of a good project management software.

REFERENCES

- [1]. Mrs.Kalaivani S., Mrs. Kavitha S., Comparative Study On Software Project Management Models ,Volume: 02 Issue: 04 | July-2015. p-ISSN: 2395-0072. www.irjet.net
- [2]. Schwalbe, Kathy. 2013. Information Technology Project Management.
- [3]. https://www.theseus.fi/bitstream/handle/.../Hoang_Nhat_Minh_Shrestha_Swostik.pdf?by NM Hoang - 2014

[4]. Wysocki Robert K., McGary Rudd. 2003.
Effective Project Management. 3rd. Wiley
Publishing.

[5]. <https://reviews.financesonline.com/>

[6]. <https://www.capterra.com/project-management-software>

Android Based Voting Application

Mary Margret

*Department of Information Technology
Thakur College of Engineering Mumbai*

Rupesh Vishwakarma

*Department of Information Technology
Thakur College of Engineering Mumbai*

Aanchal Singh

*Department of Information Technology
Thakur College of Engineering Mumbai*

Abhishek Singh

*Department of Information Technology
Thakur College of Engineering Mumbai*

Abstract- This project is about Online Voting System that is automation of present voting system which is done manually. According to this system, people who are citizens of India and age greater than 18 years can directly vote online instead of going to polling booths. For the process of automation, a database is used to store the details of voters regarding their names, ages and addresses and is maintained by election commission of India.

The main objective of the proposed system is to increase the voting percentage and even to improve the security of voting system with valid votes. We can conclude that through the online voting system the voting percentage increases and even the cost and maintenance are controlled and reduced. But at the same time voters need to be educated so that the present system can be utilized properly else complexity arises. And security should be implemented properly. Application should be adaptable and responsive during crash of system.

As of now the target audience for this application are people who are digitally literate, have the access to unhindered data connectivity and also the mobile phones which have a fingerprint sensor, basically people living in urban areas as all of the above things are easily available with current advancement in technology and devices enabled with fingerprint sensor are affordable for many

I. INTRODUCTION

Election allows the public to choose their representatives and express their selection for how they will be governed. Information and communication technology is linked for the proper execution of democratic rights. The principle of election process is fundamental to purity of democracy itself. The election system must be sufficiently sensible to withstand variety duplicitous behavior and must be appropriate so that voters and candidate can accept result of an election. A paper ballot is used to cast the vote in election and may be a piece of paper or small ball is used in voting. Actually it was a small ball used to record decisions made by voters. Each voter use only one ballot and ballot cannot be shared. At polling station voter cast their vote on ballot and put that ballot in box. Another voting system is EVM. The full form of EVM is "electronic voting machine". Votes are stored by using Control unit and it is used by poll workers. Voters used Ballot unit which is placed at election booth. This system uses a six volt alkaline battery so that it can be easily used in rural area. "Online voting" or "E-Voting" is also called as Internet voting. In this system by using internet connection voter can cast his/her votes. Gujarat was the first Indian state which implements this system in April 2011. Many people find the EVM system time consuming. This system is developed with the help of TCS. As in rural area, internet facility is not up to the mark as compare to the urban area because of the limited number of towers. So people cannot use internet for voting purpose. The current voting system is time consuming as voters have to wait in a long queue to cast their votes.

II. BACKGROUND

Voting was, is and always will be the most crucial aspect of any and every democracy. There exist many ways to cast vote for elections. These involve traditional means of voting which includes manual voting process that is using ballot papers. In ballot paper system user used to cast his vote by imprinting the stamp of his desired candidate on a paper and put it in a secret box. The other voting method is using electronic voting machine (EVM). EVM are majorly used when the population is large and it is practically impossible to count so many papers. So an EVM is like remote where there are buttons or every candidate on a single machine itself, user just has to cast his vote by pressing the button in front of candidate. Advantage of using EVM over ballot paper system is votes need not be counted afterwards instead they are counted while vote casting is done within machine itself. Security is an important part of voting. But, the issues of cheating in election and other malpractice leads in many difficulties and challenges.

III. IMPORTANCE OF PROJECT

E-voting is overcomes every major hurdle involved in the traditional system. The primary focus being the convenience of voting. Adapting to e-voting also increases the response of people for elections due to increased convenience. It is quicker and private than the traditional method. E-voting is also cost effective because it replaces the EVM machine and ballot papers. Also, the task force involved for conducting elections gets replaced except the higher authorities. Replacing the paper ballot system also makes it eco-friendly. Authentication and cross verification becomes easy when the data is recorded digitally. The E-voting application also gives the real time results. Thus, by reducing the errors involved in counting.

IV. SCOPE OF THE PROJECT

Presently the voting consumes a lot of money which would be reduced in considerable amounts. Furthermore the flexibility for voter increases as they have to vote from any place, any time of the day when the election is scheduled. The project has various modules like authentication, registration and voting and displaying results. For voting and authentication fingerprint scanner is used to maintain the secrecy. The application will have its own database and maintain information about voters, candidate and all necessary information about upcoming elections.

V. PROBLEM DEFINITION

The proposed system encompasses the voting feature for an institution over the application. It has processes like registering, authentication, voting and displaying the results. Registration should include signing up of new user and should not allow existing users to create a new account. Registration should be done using email id phone number and fingerprint. In the proposed system authentication of the user should be done by OTP which will be sent over SMS and email. Also the fingerprint should be recorded using the cell phones fingerprint sensor to maintain the uniqueness. Phone number and email-id of each and every user will be unique and also as the user-id should be generated and provided by the application. Thus, user id also remains unique. While voting, authentication should be done by users fingerprint to make sure the user is unique. As

the voting process continues, simultaneously the votes will be recorded and counting will be done. After the completion of voting process, the results should be declared on the application.

VI. METHODOLOGY

Initially user will be registering to our application using his/ her contact no. and email then application would be checking whether the email and mobile no already exists. If yes alert message is given to the user. If no, he can proceed to further authentication. Eventually user ID and password would be generated for the user. Now for voting process user has to log in to the application where he can read about the candidates information. While voting will be done using fingerprint and if the user has already voted he would be alerted about the same. After entire voting process is completed the votes will be counted and results will be displayed

- [1] Android application is to be created with the help of android studio which will contain different pages like login page, registration page, voting page, home page, Notification page, and Profile page of user as well of candidate.
- [2] Google firebase is used to store, collect, analyze, process, modify, retrieve, and many other operation.
- [3] With the help of Google firebase we will maintain the data of user as well as candidate, votes of the user will be stored on firebase, it also helps in counting the votes and provide the result instantly.

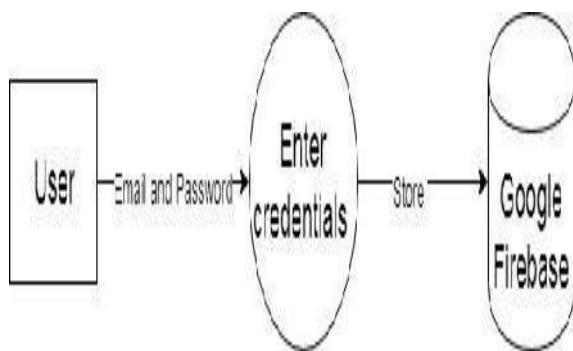


Fig.1 DFD Level 0

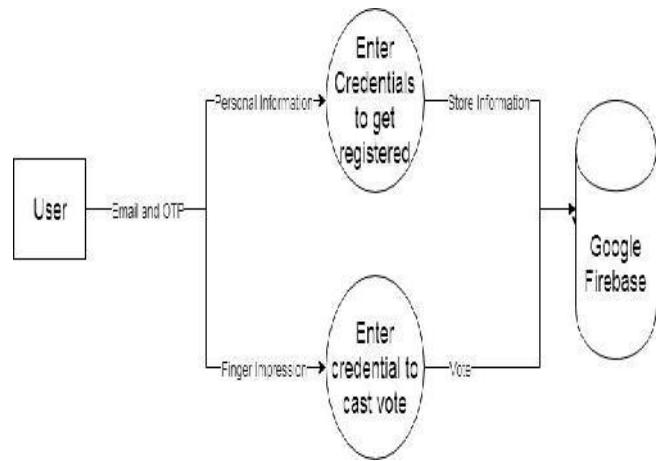


Fig.2 DFD Level 1

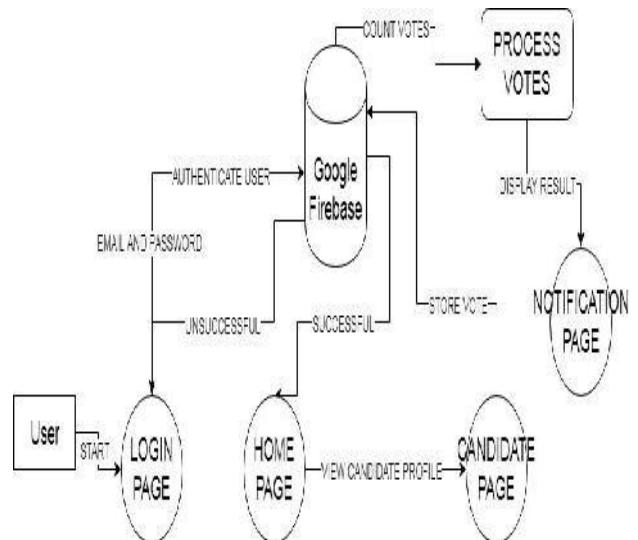


Fig.3 DFD Level

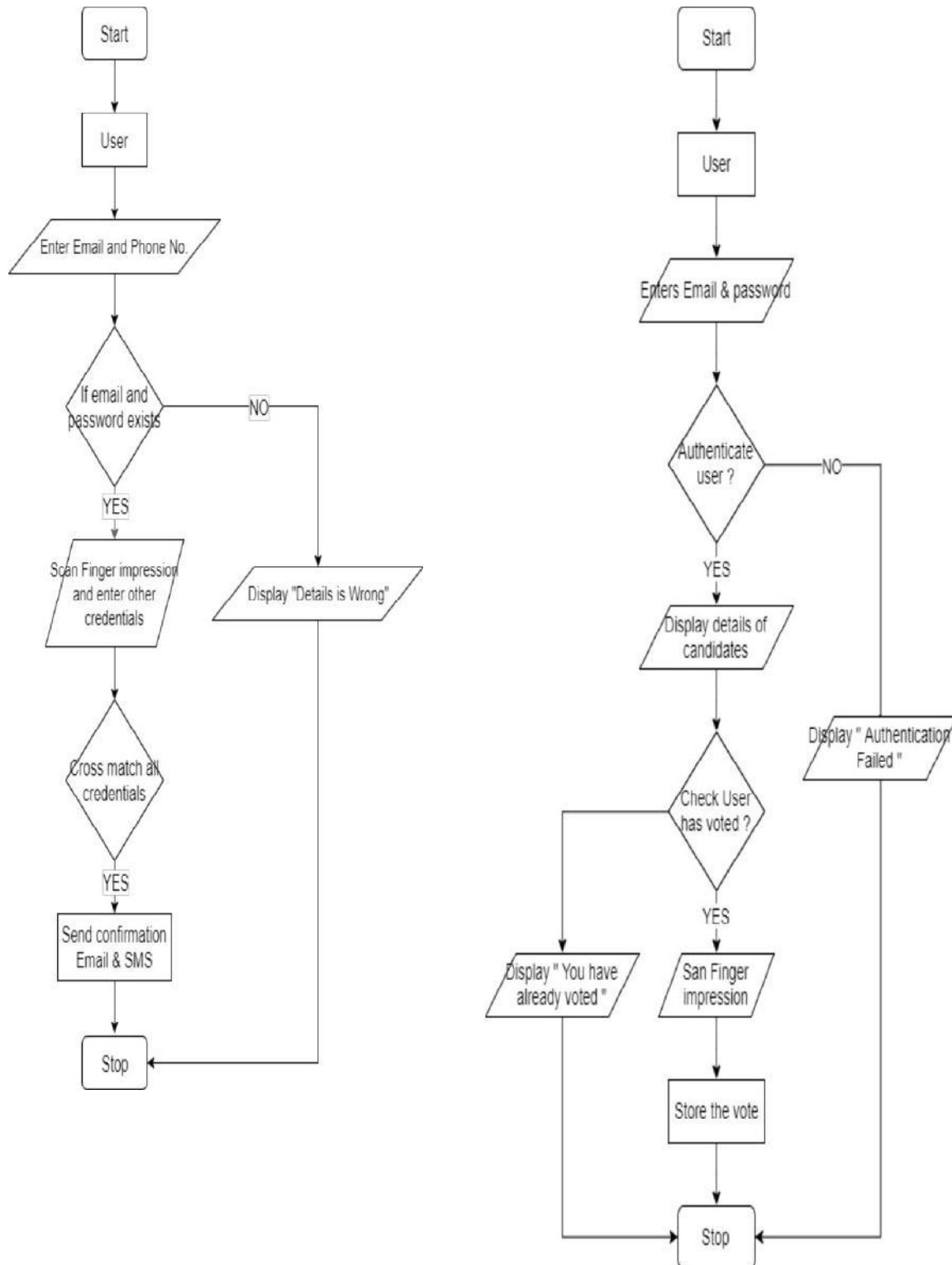


Fig.4 (a), 4(b): Flow of Voting Process

VII. RESULT AND DISCUSSION

A. Expected Result

Phase I: Planning and Analysis

Discussion with the team members and project guide about the project definition, finalized the project problem definition. Finalize roles of each member in development of project. All the necessary information for the development of the project is supposed to be collected and one is expected to analyse the feasibility for the same.

Phase II: Designing and Implementation

Design the GUI of the application; prepare the activity for all the pages like Login page, Signup page, Main Menu, About, Help, etc. Connectivity between both of them such that to achieve the basic functionality.

Phase III: Coding and Testing

Google firebase and its inbuilt packages and classes is used to create and maintain the data. Using Java, firebase and applications connectivity would be done. Application is expected to work properly with all the functionality and features. After the completion of designing and implementation part of each and every module, every module will be tested and if in case there is any error present in module it is supposed to be corrected accordingly.

B. Future Scope

With the help of this project one will be analyzing the traditional voting system and suggest a new and flexible way for voting. Thus, by this project there will be an alternate fast way to cast the vote via this application. With the help of the application it is expected that the cost and time required for conducting the election will get reduced. It is expected from the above project that it will help the institute and the organization to conduct the election as per there requirements in a very convenient way.

Currently the application is suitable to use only in the urban areas as many of people in our country lack basic digital services like data connection and slow data rates. Moreover, people also don't have mobile phone which has a fingerprint sensor in it. Many people living in rural areas also digitally illiterate. Same is not the case in urban areas as many have such phones embedded with fingerprint sensor and continuous data connectivity. So at initial stages it can be applied in urban areas.

With current advancement in technologies and reducing cost of devices it is expected that within few years everyone will be able to afford and buy a mobile phone which has an embedded fingerprint sensor. So the process of voting can be completely digitized across entire country.

VIII. CONCLUSION

This project focused on the analysis and development of e-voting application on an android platform which can tackle all earlier issues that were encountered in a conventional voting system. It will be definitely helpful for the users who wish to vote and the voting process will be made very easy by using this application. E-voting application systems have many advantages over the traditional voting system. Some of these advantages are less cost, faster generation results, easy accessibility, accuracy, and low risk of human and mechanical errors.

It is very difficult to develop e-voting system which can allow security and privacy on the high level. Due to option of viewing candidate profile the voters get easily acquainted with all the candidates easily. Using android voting system percentage of voting increases and cost and time of voting process is decreases. It is easy to use and it is less time consuming.

The system is first analyzed for both SMS and internet based voting. It's then developed by implementing both techniques using android platform. Future development focuses to design a system which can be easy to use and will provide security and privacy of voters and votes on acceptable level by proper authentication and

processing section. Fingerprint recognition is used as a form of biometric to recognize identities of human beings. Thumb impression of every individual is unique. Instead of providing authentication based off a single thumb print using both the thumbs guarantees greater amount of security and makes the process of authentication more efficient and evolutionary. Hence, it helps in maximizing the security.

Our system maintains integrity of the voting process from the minute the voter casts his/her vote until the cast vote is registered. While observing fully fledged voting transparency, the system is capable of denying access to any unauthorized voter and also prevents multiple votes from same voter. With the use of the e-voting system, many of the issues that have challenged traditional voting systems in the past are thus resolved.

IX. ACKNOWLEDGMENT

Before going on with the contents of our project we would like to extend our gratitude to the people without whom this project would not have been possible.

We sincerely thank to our guide Mrs. Mary Margarat for his guidance and support for carrying out our project work.

We also thank the project coordinators Mr. Aditya Desai and Mrs. Neha Kapadia for arranging the necessary facilities to carry out the project work.

We also thank our respective Dr Rajesh S. Bansode (HOD – IT), Dr R.R. Sedamkar (Dean Academic), Dr B.K. Mishra (Principal) and the college management for their support.

REFERENCES

Journal References:

- [1] AashishPedbulkar, PranavRajhans, Shweta Joshi. "Empirical study for e voting system" at IJARCEVolume 4, Issue 3, March 2015
- [2] Dr.AreeAliMohammed RamyarAbdolrahmanTimour "Efficient E-voting Android Based System" at IJARCE Volume 3, Issue 11, November 2013.
- [3] KirtiAutade, PallaviGhadge, Sarika Kale, Co-authors- Prof. N. J. Kulkarni, Prof. S. S. Mujgond "E-voting on Android System" ISSN 2250-2459, Volume 2, Issue 2 in February 2012.
- [4] Mary Maragarat, HitendraBhandari, ParthModha ", E-Voting System ", ISBN 978-0-9972393-4-8,"Multicon-w 2017" ICCCV in February 2017.

WEB REFERENCES:

- [1]<https://pdfs.semanticscholar.org/881b/625d2d036cdaa98a6ee332bd451ad403f3.pdf>
- [2]<http://www.ijetae.com/files/Volume2Issue2/IJETAE>
- [3] <http://www.ijsr.net/archive/v4i6/SUB156101.pdf>
- [4]<https://tcetmumbai.in/multicon/2017/ICCCV>.

E-Commerce Android Application For Solid Waste Management

Alireza Chaviwala,
Computer Engg.
MHSSCOE,
Mumbai, India

Huney Pandita,
Computer Engg.
MHSSCOE,
Mumbai, India

Vequar Ansari,
Computer Engg.
MHSSCOE,
Mumbai, India

Nazneen Pendhari
Computer Engg.,
MHSSCOE,
Mumbai, India

Uma Kale
Computer Engg.,
MHSSCOE,
Mumbai, India

Abstract- The generation of Municipal Solid Waste in huge quantities is a major problem in the fast growing cities as well as villages. Apart from that, the lack of waste management has further aggravated the problem of pollution and spread of diseases. The problems in question are: No segregation at source, dumping, land filling, broken bins, lack of awareness and expertise etc. With the widespread use of technology in today's times, these problems can be effectively addressed and awareness can be created on a mass scale. Techniques of waste segregation, composting both aerobic and vermicomposting, can be taught. Also keeping in mind the fast-paced lifestyle of the citizens of India, many may not be able to practice these techniques on an individual level, therefore a door-to-door waste collection and segregation service can further solve the problem. This paper proposes a mobile application for the Android platform that provides this service to counter waste mismanagement and throw light on its contents as well as implementation. The waste collected from the user of this application will be exchanged for cash that will be transferred electronically to him/her, thereby encouraging maximum participation.

Keywords-Solid Waste Management, e-Commerce, Android application, Door-to-door service, Segregation at source, Waste Management awareness, Composting at Home.

I. INTRODUCTION

The rapidly plummeting population in India has resulted in an exponential increase in the consumption rate of the masses.

With an ever-increasing population of 1.6 billion and a population density of approximately 382 per sq km[1], the population pressure makes its increase in consumer waste generation. With a growth rate of nearly 7 % in the last 20 years[2], the solid waste generation is expected to rise even more in the coming years. The solid waste generated on a micro-level is termed as Municipal Solid Waste (MSW). More than 90% of the MSW generated in India is dumped in landfilling areas and disposed in an improper manner. This has a direct impact on the environment and results in the spread of various diseases and health issues. The status report of Central Pollution Control Board [CPCB] states that India produced around 0.143 million tons per day from which only 23% was properly processed [2]. The MSW that is dumped in open grounds decomposes to form methane, a greenhouse gas which contributes to climate change. In India, the per capita of municipal waste (MSW) generated daily ranges from 500g in large towns to 100g in small towns. [3].

Before delving into details about MSW, we shall first list the different types of Solid wastes. This shall give us an insight in solving the problem about segregation at source.

Solid wastes, from a general perspective can be classified as follows:

Classification of Solid wastes:

A. Domestic/Residential waste:

This category of wastes comprises the solid waste that originate from households. These wastes are generated as a consequence of household's activities such as cooking, cleaning, repairs, hobbies and redecoration and contain empty containers, packaging, clothing, book, writing paper, and furnishings. Residents in developed countries sometimes discard bulky wastes such as furniture and large appliances which cannot be repaired and used.

B. Commercial waste:

Included in this category are solid wastes that originate from offices, wholesale and retail stores, restaurants, hotels, markets, warehouses and other commercial establishments.

C. Street Waste:

This term applies to waste that is collected from streets, walkaways, alleys, and vacant areas. Street waste includes paper, cardboard, plastic, dirt, leaves and other vegetable matter discarded by the road users. It also included the waste left by vehicles moving over the roads.

D. Ashes:

Ashes are used for cooking and other domestic uses, produced from the burning of charcoal, wooden articles, coal as well as coke. It is also produced as residue from institutions and small industrial organizations. It is classified as industrial waste when it is produced in large quantities at power plants.

E. Hazardous wastes:

The wastes which are dangerous physically, chemically and biologically and which origin from industries, institutes or Consumers are known as Hazardous wastes.. It is included in solid wastes since it is stored in solid containers, even if its active agents are liquid or gases. Common examples: solvents, pesticides or paints which mix with municipal waste and become part of urban waste.

A look at a few statistical records give further insights about waste generation in India.

Table 1 Percent-Wise Distribution Of Municipal Solid Waste In Kolkata City[4]

Sources of Wastage	Percentage (%)
Household waste	34.20
Street sweeping	22.80
Institutional waste	6.32
Commercial and market waste	36.37

Table 2: MSW COMPOSITION AT KOLKATA DURING 1970, 1995 & 2007[4]

Sl. No.	Parameters	1970	1995	2007
1.	Biodegradables	40.36	44.29	50.56
2.	Green coconut shells	4.95	8.51	4.5
3.	Paper	3.17	4.64	6.07
4.	Plastics	0.64	3.22	4.88
5.	Metals	0.66	0.43	0.19
6.	Glass & Crockery	0.38	1.72	0.34
7.	Coal	6.08	3.10	-
8.	Inert	40.76	26.82	29.6
9.	Others	3.00	7.27	3.86

One of the major challenges in Solid Waste Management is to address these problems in complete detail and come up with a practical solution that can be implemented on a mass scale and which is also feasible.

F. No Segregation at Source:

One of the major factors for garbage generation is that waste that is generated is not segregated. Segregation at source refers to separation of different kinds of waste such as plastic, glass, paper, kitchen waste etc. right after it is generated and before it is dumped into the garbage bin. Segregation primarily is categorized into Wet and Dry waste. Further, within the Dry waste category, each item is collected separately For eg: Glass is collected in a separate container, Paper in a separate, Dust in another etc.

However, this problem can be solved by implementing the Coloured Bins method of segregation. In this method, a unique colour code is used for each item and the bin is coded with that colour such that, that particular item is collected ONLY in that bin. Likewise, a colour scheme can be devised to implement

- 1) Garden waste are indicated by green bins

- 2) Food waste are indicated by picture signs.
- 3) Paper is indicated by blue bins.
- 4) Soft plastics indicated by orange bins
- 5) Hard items ranging from metal, plastic, unbroken glass etc. belong in red bins.
- 6) Cloth in lavender bins.
- 7) Sanitary in pink bins.
- 8) Sharp hazardous needles and broken glass in yellow bins.
- 9) Hair and dust and sand in beige bins.

G. Land filling:

In many metropolitan cities, open, uncontrolled and poorly managed dumping is commonly practiced, giving rise to serious environmental degradation. More than 90% of MSW in cities and towns are directly disposed of on land in an unsatisfactory manner. Such dumping activity in many coastal towns has led to heavy metals rapidly leaching into the coastal waters. In larger towns or cities like Delhi, the availability of land for waste disposal is very limited [5]. In the majority of urban centers, MSW is disposed of by depositing it in low-lying areas outside the city without following the principles of sanitary landfilling. Compaction and leveling of waste and final covering by earth are rarely observed practices at most disposal sites, and these low-lying disposal sites are devoid of a leachate collection system or landfill gas monitoring and collection equipment. All of the infectious waste disposed at hospitals is dumped at disposal sites since no segregation

H. Lack of awareness

This problem is at the core of all MSW mis-management. A major part of the Indian population does not have proper knowledge and skill to dispose every-day waste. Dumping in street niches, or on the road is a common habit. However, this attitude has to be changed as a primary measure to counter garbage generation. Firstly, the people need to be educated and made aware of the consequences of dumping waste and brought to their notice, the direct co-relation between their unhygienic habits/practices and the rampant diseases and damage to the environment. Outreach is a very important aspect of waste management in India, as the awareness on proper depositing of waste into receptacles is not clearly understood or widely practiced. Municipal dumpsters gives the impression that organized waste management is an unpleasant, unsightly and health

hazardous practice. This only intensifies the cultural concept that waste is something not to be associated with, or it is a problem that is too big to be tackled. This forces individuals to deal with their waste in extremely non eco-friendly ways, such as burning or dumping indiscriminately on land or in water. Thus, public health concerns become aggravated. Lack of awareness among individuals affects their whole community; as people leave their own devices unchecked. Waste management needs to be understood as resource management, protection, and utilization. [6]As a secondary measure, one can decompose their solid waste at home using various techniques like Aerobic composting and Vermicomposting. These techniques can be taught to the people as part of educating them with techniques to manage garbage.

t source takes place

I. Aerobic Composting

The organics present within the MSW, when decomposed in the presence of air under hot and moist conditions, undergoes bacterial conversion, this process is called composting, and the final product yield is called compost (humus) which is a very important agricultural resource. This compost or humus is nonodorous and is used as a fertilizer. Also, it is free of pathogens. The waste volume is reduced to 50-85% as a result of the composting process [5]. The composting processes are either manual or mechanical depending upon the use of manual or mechanical means. Manual composting processes are carried out in smaller urban centers and mechanical composting plants have been established in big Indian cities.

J. Vermi-composting

The stabilization of organic matter along with the joint action of earthworms and aerobic micro-organisms is called as Vermicomposting. Initially, microbial decomposition of biodegradable organic matter occurs through extra cellular enzymatic activity (primary decomposition). Earthworms consume five times their body weight of organic matter per day on partially decomposed matter. The ingested organic matter is further decomposed in the gut of the worms, resulting in further particle size reduction. This product can serve as a bio fertilizer in agriculture.

II. CONTROL MEASURES

The ever-increasing issue of garbage dumping and littering has been a major concern for urban residential settlements and corporate offices. The health and well-being of citizens is at risk since such dumping and littering causes breeding of mosquitos and spreads many diseases like dengue, malaria, etc. An ideal solid waste management system implements waste

picking and disposing in an organized and effective manner. Such an objective can be fulfilled by calling for active and direct interaction with the people. It helps in controlling the waste at the source. The confidence of the people is gained and they will be invited to participate. The residents communicate directly to the system which acts as a middleman between the residents and scrap companies.

The system functions through an Android application where the user of the application can sell his solid-waste to the organization running the application. This approach not only will boost the peoples participation but will also stop them from littering and take waste management seriously as it will help them make a small earning that too, at the convenience of their doorstep. Corporate companies, keeping in mind the well-being of their valuable employees, will volunteer to dispose waste via the medium if this application and further contribute to effective prevention of waste littering and pollution.

The afore mentioned problem is not only solved but is also expected to drastically reduce pollution and prevent further mismanagement of garbage/waste.

The android application shall run on a stable internet connection provided either by a Wi-Fi or a mobile data connection. The application will register the user on the first use and will ask the user to link his/her bank account to his account within the application. Also, the user is asked to enter his location/address which will be used for picking up solid-waste from the user. The user will be asked to choose a category between Solid waste, Compostable matter and Hazardous waste. After that, the user shall choose the items he wishes to sell from a drop-down list of waste materials that fall in the chosen category. Finally the user enters the quantity by weight. Also, the user specifies the timing preference for collection. The application gives a final review of the items the user wants to sell and asks for a confirmation. Once confirmed by the user, the application sends a request to the system which then looks for collectors within the area specified by the user for the given time period. The system then notifies the user within the application and through SMS acknowledging the pick-up request and the arrival time of the collector.

Once collected the user can track the progress of the garbage to the scrap company's warehouse. After the garbage is handed over to the company, the cash transfer process is started and reaches the user's account within 24 hours of pickup and after reaching the warehouse. The cash is transferred to the user and he/she is notified through the application as well as SMS text message.

The application also intends to share the work load of the Municipal Department and provide employment to many.

Payment scheme

Table 3 Rates of waste articles

Sr.no.	Waste articles	Rates per kg.(in Rs)
1	Paper	Rs.7
2	Glass	Rs.15
3	Plastic	Rs.10
4	Rubber	Rs.4
5	Metals (varies as per type)	Rs.10 - Rs.17
6	Hazardous waste	Rs.6

III. CONCLUSION

Thus, in conclusion of this paper we would like to cover all the points which are discussed in detail above. We first looked at the past and present situation of the Solid waste management in India, and established the major problems that are created and faced there-in. Then, we enlisted the major issues such as No Segregation at source, Land filling and Lack of Awareness and the solutions to those problems. Thirdly, we mentioned the decomposition methods that can be adopted on an individual level. Finally, we proposed the Android application, that can be used to provide door-to-door waste collection service which will further solve the problem of garbage generation on a mass scale.

REFERENCES

- [1] Census.(2011). Registrar General Of India. Retrieved from http://censusindia.gov.in/2011-prov-results/data_files/india/Final PPT 2011 chapter3.pdf
- [2] Ministry of Environment, Forest and Climate Change (2015). "Action Plan for Management of Municipal Solid Waste". Central Pollution Control Board (CPCB). Available from <http://www.cpcb.nic.in/NGTorderdatedFeb-05-2015.pdf>
- [3]. S. Singhal and S. Pandey , "Solid waste management in India: Status and future directions", TERI Information Monitor on Environmental Science (TIMES), vol. 6(1), pp. 1-4, 2001.
- [4] A Holistic Approach for Integrated Solid Waste Management System of Kolkata Municipality Corporation Area(IEEE 2011)
- [5] Municipal solid waste management in Indian cities – A review(ELSEVIER 2007)
- [6] A System Design for Solid Waste Management: A Case study of an implementation in Kerala(IEEE 2017)

BlockChainTechnology: Future Aspect

HetalAmrutia

Thakur College of Engineering & Technology, Mumbai
hetal.amrutia@thakureducation.org

Anil Vasoya

Thakur College of Engineering & Technology, Mumbai
anilk.vasoya@thakureducation.org

Kamal Shah

Thakur College of Engineering & Technology, Mumbai
kamal.shah@thakureducation.org

Abstract-Digital world has produced new advanced and efficient products, and close customer relationships globally by the use of smart phone, Cloud Computing, IoT (Internet of Things), social media and big data analytics to generate models for better decisions. Blockchain is a decentralized ledger used to securely exchange digital currency, perform deals and transactions. Each member of the network has access to the latest copy of encrypted ledger so that they can validate a new transaction. Blockchain ledger is a collection of all Bitcoin transactions executed in the past. Basically, it's a distributed database which maintains a continuously growing tamper proof data structure blocks which holds batches of individual transactions. The completed blocks are added in a linear and chronological order. Each block contains a timestamp and information link which points to a previous block. Bitcoin is peer-to-peer permission-less network which allows every user to connect to the network and send new transaction to verify and create new blocks. Satoshi Nakamoto described design of Bitcoin digital currency in his research paper posted to cryptography listserv in 2008. Nakamoto's suggestion has solved long pending problem of cryptographers and laid the foundation stone for digital currency. This paper explains the concept, need of Blockchain and how it works, blockchain as a cryptocurrency bitcoin. It attempts to highlights role of Blockchain in shaping the future of real estate and other use cases of blockchain.

Keywords: Blockchain Technology, Bitcoin, Real Estate, Use cases.

I. INTRODUCTION OF BLOCKCHAIN TECHNOLOGY

Historically, when it comes to transacting money or anything of value, people and businesses have relied heavily on intermediaries like banks and governments to ensure trust and certainty.[1] Middlemen perform a range of important tasks that help build trust into the transactional process like authentication & record keeping.[2].The need for intermediaries is especially acute when making a digital transaction. Because digital assets like money, stocks & intellectual property, are essentially files, they are incredibly easy to reproduce. This creates what's known as the double spending problem (the act of spending the same unit of value more than once) which until now has prevented the peer to peer transfer of digital assets.[3]But what if there was a way of conducting digital transactions without a third party intermediary? Here Answer is Blockchain.

A blockchain is a tamper-evident, shared digital ledger that records transactions in a public or private peer-to-peer network. Distributed to all member nodes in the network, the ledger permanently records, in a sequential chain of cryptographic hash-linked **blocks**, the history of asset exchanges that take place between the peers in the network. All the confirmed and validated transaction blocks are linked and chained from the beginning of the chain to the most current block, hence the name blockchain. The blockchain thus acts as a single source of truth, and members in a blockchain network can view only those transactions that are relevant to them.

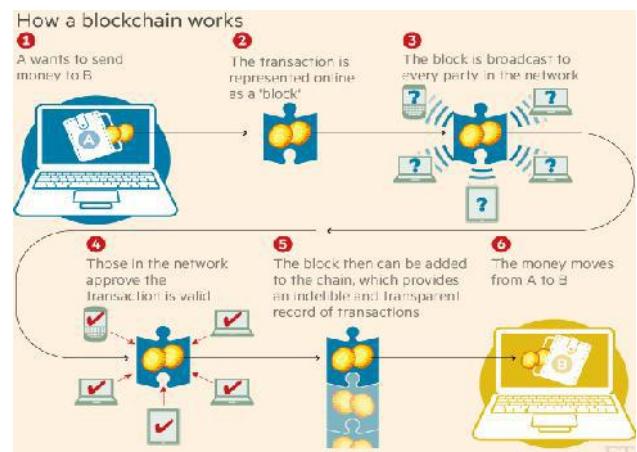
Blockchain allows users to manipulate the ledger in a secure way without the help of a third party.

A bank's ledger is connected to a centralized network. However, a blockchain is anonymous, protecting the identities of the users. This makes blockchain a more secure way to carry out transactions.

The algorithm used in blockchain reduces the dependence on people to verify the transactions. This technology used for recording various transactions has the potential to disrupt the financial system.

II. HOW BLOCKCHAIN TECHNOLOGY WORKS?

We explain the concept of the blockchain by explaining how Bitcoin works since it is intrinsically linked to the Bitcoin. However, the blockchain technology is applicable to any digital asset transaction exchanged online.



III. USE OF BLOCKCHAIN TECHNOLOGY IN CRYPTOCURRENCY: BITCOIN

Bitcoins are a crypto-currency and digital payment system invented by an unknown programmer, or a group of programmers, under the name Satoshi Nakamoto. That means they can be used like a usual currency, but don't physically exist like dollar bills. They are an online currency which can be used to buy things. These are similar to "digital cash" that exist as bits on people's computers. Bitcoins exist only in the cloud, like Paypal, Citrus or Paytm. Even though they are virtual, rather than physical, they are used like cash when transferred between people through the web.

The Bitcoin system is peer-to-peer network based and transactions take place between users directly, without an intermediary. These transactions are verified by network nodes and recorded in a public distributed ledger called a Blockchain. Since the system works without a central repository or single administrator, Bitcoin is called the first decentralized digital currency.

Bitcoin production makes them a unique currency. Unlike normal currencies, Bitcoins cannot be created as needed. Only 21 Million Bitcoins can be created, of which 17 million have already been created. Bitcoin gets created whenever a block containing valid transactions is added to the Blockchain. This is the only means for creating Bitcoins and through various mathematical and encryption algorithms we ensure no fake Bitcoins are created or circulated [1].

IV. BLOCKCHAIN COULD MAKE AN IMPACT ON THE REAL ESTATE MARKET

Real estate and technology seem to go hand in hand these days. With more advancements in technology being made, more options are available to realtors to sway buyers to purchase new homes. Blockchain is one such technology that could make an impact in the real estate market by offering real-time data that can help secure financial transactions.

1. Speed Up The Transaction Process

Brokers, lawyers, and banks have long been part of the real estate ecosystem. However, blockchain may soon usher in a shift in their roles and participation in real estate transactions. New platforms can eventually assume functions such as listings, payments, and legal documentation. Cutting out the intermediaries will result in buyers and sellers getting more out of their money as they save on commissions and fees charged by these

intermediaries. This also makes the process much quicker as the back-and-forth between these middlemen gets cut.

2. Increase Transparency

One of the largest costs associated with real estate transactions is the lack of transparency, and the second arises from transaction friction due to so many middle men. The blockchain allows all transactions to happen peer to peer, with no middle men and with all information recorded in a fully transparent manner.

3. Secure Online Data Storage

More than a cryptocurrency, blockchain technology allows for secure online data storage. This could be used in the future for storing of all property records, including deeds, plats, surveys, title chain, etc.

4. Be Cost Effective While Trusted

Blockchain will eliminate the need for title insurance and add efficiency and confidence to transactions. It will also add transparency across the entire purchaser platform, so that economics such as not being able to afford an attorney will not be a factor in understanding title and ownership issues.

5. Cross-border Standards

One more advantage associated with blockchain technology is geographical freedom. And it is something that comes in especially handy since there are so many expats facing challenges when applying for rentals outside their home countries. With international currencies like BERRY tokens and standardized rental procedure, the process of renting and renting out will be equally seamless and stress-free for people from all around the globe.

V. BLOCKCHAIN TECHNOLOGY FOR FUTURE ENTREPRENEURS



Fig.2 Use cases for blockchains

Blockchain can be used to provide immutable land titles the prominent challenge Indian real estate developer's face is finding a land with clear titles or ownership record. In property buying or selling, lots of documents need to be verified by authorized persons. In this technology, all these information are going to be recorded in numbers of block over a period and blocks are linked together like a chain. Each block is secured by password or key. Thus, deleting or tampering data in one block alone will not help the cause. This kind of immutable security is the fundamental difference between blockchain technology and others.

Blockchain technology could finally enable online voting systems and direct democracy to be created by ambitious social entrepreneurs to provide to governments because of the powerful cryptography built into the system that could be used to virtually eliminate voter fraud and massive increase civic engagement.

Blockchain technology can help bring people out of poverty. Globally people work in developed countries and send part of their wages back to their families back in their ancestral land. A Philippine McDonald's worker in the United States who sends her wages home to her sick grandma in Manilla will end up paying Western Union 10% just to deliver the money, a process which can take up to a week or more. With the use of blockchain technology, this process can not only be instant but can be

sent to local tellers with high ratings to deliver the cash, converted into pesos, directly to her door for likely as little as even 2 cents. (Tapscott, 2016)

Blockchain could help create true sharing economies, which today do not really exist. Services like Airbnb, while empowering individual property owners to rent them out, still aggregate those offerings through the Airbnb platform acting as an intermediary and taking a cut of every transaction.

Blockchain technology could enable true peer to peer transacting without the need for Airbnb to provide things like the payment processing, verification of the property, and more.

Blockchain applications can disrupt financial markets, particularly in things like stock trading, could be used to make the stock brokerages, which contributed to swelling wall street coffers and the billions paid out in bonuses to employees while everyone else lost their homes in the 2008/2009 stock market crash (Story & Dash, 2009), obsolete. If you can directly buy ownership in a company without requiring someone to verify that transaction and sent proof of ownership to you, then why would you need a broker? Not only would this, but no-commission trading allow millions of micro-investors from all over the world to begin participating in financial markets.

Blockchain technology provides end to end charity

Common complaints in the charity space include inefficiency and corruption, which prevent money from reaching those that are meant to have it. Using blockchain technology to track donations can let you be sure your money is going to end up in the right hands. Bitcoin-based charities like the BitGive Foundation use blockchain's secure and transparent distributed ledger to let donors see that the intended party has received the funds.

Blockchain technology provides cyber Security, Although , the blockchain ledger is public, the data is verified and encrypted using cryptography. This way data is less likely to being changed or hacked without authorization.

Blockchain technology help in Forecasting The blockchain is set to change the entire approach to research, consulting, analysis and forecasting. Online platforms like Augur are looking to create global decentralized prediction markets. These technologies can be used to place and monitor bets on anything from sports to stocks to elections in a decentralized way.

Blockchain technology help in Private Transport and Ride Sharing The blockchain can be used to create decentralized versions of peer-to-peer ridesharing apps, allowing both car owners and users to arrange terms and conditions in a secure way without third party providers. Startups working in this area include Arcade City and La'Zooz.

The use of built-in e-wallets can allow car owners to automatically pay for parking, highway tolls, and electricity top-ups for their vehicle. UBS, ZF and Innogy are some of the companies developing blockchain based e-wallets.

Blockchain technology help in Public Benefits, The public benefits system is another sector that suffers from slowness and bureaucracy. Blockchain technology can help assess, verify, and distribute welfare or unemployment benefits in a much more streamlined and secure way. GovCoin is a UK-based company that is helping the government to distribute public benefits using blockchain technology. The blockchain is also a good contender for implementing a basic income. Circles is a project working on developing a blockchain-based technology for implementing a universal basic income.

Blockchain technology help in Healthcare, Another industry that relies on many legacy systems and is ripe for disruption is healthcare. One of the challenges hospitals face is the lack of a secure platform to store and share data, and they are often victims of hacking because of outdated infrastructure. Blockchain technology can allow hospitals to safely store data like medical records and share it with authorized professionals or patients. This can improve data security and can even help with accuracy and speed of diagnosis. Gem and Tierion are two companies that are working on disrupting the current healthcare data space.

Blockchain technology help in Online Music Several startups are coming up with ways for musicians to get paid directly from their fans, without giving up large percentages of sales to platforms or record companies. Smart contracts can also be used to automatically solve licensing issues, and better catalog songs with their respective creators. Mycelia and Ujo Music are two startups creating blockchain-based solutions in the music industry.

Blockchain technology help in Supply Chain Management, With blockchain technology, transactions can be documented in a permanent decentralized record,

and monitored securely and transparently. This can greatly reduce time delays and human mistakes. It can also be used to monitor costs, labour, and even waste and emissions at every point of the supply chain. This has serious implications for understanding and controlling the real environmental impact of products. The distributed ledger can also be used to verify the authenticity or fair trade status of products by tracking them from their origin. Some blockchain companies working in this sector are Provenance, Fluent, SKUChain, and Blockverify.

Five facts entrepreneurs should remember about blockchain:

1. The platform of future: Blockchain is future, learn, adopted and apply.
2. Evolving Technology: The technology is still in its incubation stage, so it will change every day.
3. Risk of failure: Like in any other field; the startup build on blockchain will have the same odds of success or failure.
4. Builds Peer-to-Peer relationship: Blockchain will drive peer-to-peer interaction as no technology has done before.
5. Trust of Common man: Though the new technology has grabbed the world's attention, there is still lack of trust among common man when it comes to bitcoin there are not sure what it is?

VI. CONCLUSION

As mentioned above, most of these applications are still underdeveloped. The future potential of the blockchain applications is still unraveling. The next couples of years will be all about experimenting and applying to all aspects of society. Regardless of which application comes first on a global scale. The bottom line is, Blockchain is here to stay and is transforming how our society functions.

REFERENCES

- [1]. RESEARCH ARTICLE Where Is Current Research on Blockchain Technology?—A Systematic Review Jesse Yli-Huumo¹, Deokyoon Ko², Sujin Choi⁴ *, Sooyong Park², Kari Smolander³
- [2]. https://www.huffingtonpost.com/ameer-rosic-/5-blockchain-applications_b_13279010.html
- [3]. <http://futurethinkers.org/industries-blockchain-disrupt/>
- [4]. <https://www.shapingtomorrow.com/home/alert/665529-Future-of-Blockchain>

IoT & Cloud Based Vehicle Toll Collection System

Neha Kapadia

Assistant Professor, IT Dept.,

Thakur College of Engineering & Technology,
Mumbai, India

I. INTRODUCTION

Abstract- The most common approach for collecting tolls was to have the driver stop and pay a toll collector sitting in a tollbooth. However, it is now viewed as toll collection process is unfeasible principally due to its adverse impact on traffic flow and its high collection costs, not to mention its effects on the environment. Different problems associated with traditional toll collection methods urged sophistication in the approach. After progressive developments resulting mostly from flawed strategies, "Electronic" Toll Collection (ETC) systems proved to suitably deal with the shortcomings. Though many different ETC schemes are in operation across the globe, the fundamental is to be able to automate vehicle identification and assess tolls requiring no action by the driver. This research was, therefore, focused on studying the varied approaches to electronic tolling while working towards a feasible solution. The approach adopted for implementing the prototype employed Radio Frequency Identification (RFID). The basic idea was to work with RFID chips affixed at a corner of the vehicle's windshield. As the vehicle would pass through the toll junction, the chip would be scanned by RFID readers calibrated to the same frequency as the chips, mounted on either end, and an ID, unique to each chip, would be sent to the server via an on board WiFi module. This ID would be used as an index to look up the database, fetch the associated user's details and assess the toll accordingly. The user would also be notified of the transaction via SMS and/or email. This would allow speedy passage of vehicles eliminating the heavy congestion. As an added benefit, this would also eliminate the need for traditional book keeping by permitting authorized personnel to access daily logs anywhere, anytime. This would mean centralized control, improved audit, vehicle tracking and more. An online portal was also provisioned, allowing users to register themselves, check their billing history, choose appropriate payment methods, recharge their accounts or pay their dues.

Keywords- Internet of things(IoT),Automatic vehicle identification, Amazon web service ,GPS, RFID tag, RFID reader, Theft detection

Tolled highways are some of the busiest roads in the country, not only because of high volume of traffic but also due to the fact that they employ a tedious and an inefficient practice of manual toll collection. The existing system works on the principle of dividing the toll way into a number of lanes, each serviced by a toll booth. Each toll booth houses two operators, one to collect toll payments as vehicles enter the toll plaza and the other to print and hand in acknowledgement receipts as vehicles move out. Towards the end of each lane, where vehicles exit the junction and join the free flowing traffic, stands another operator who drives the barricade mechanism to obstruct the passage of vehicles until the toll is paid.

A. Case studies

Express Toll Route

ETC system used in Canada is known as the Canada 407 Express toll route (ETR). It is one of the most complicated toll roads in the world [3] [4]. The Canada 407 ETR is a closed-access toll road, which means that there are gantries placed at the entrance and exit points of each toll. In this system, Optical Character Recognition (OCR) is set with the cameras. The OCR cameras are used to take photographs of license plate numbers of vehicles that do not have transponders. The toll bill will then be sent directly to the registered address of the vehicle owners. Other than that, two laser beam scanners are placed above the roadway to detect the types of vehicles passing through the gantries. However, this toll road bears a very high infrastructure cost, and the users are the ones who help recover the cost through increments in their toll bills [3].

i. National Automatic Toll Collection System

The ETC system used in Poland has been proposed by the Motor Transport Institute along with the University of Technology in Warsaw and Dublin. This system is called the National Automatic Toll Collection System (NATCS), and consists of the National Automatic Toll Collection Center (NATCC), control gates, and on-board units (OBU). The NATCS uses a combination of mobile telecommunication technology (GSM) with satellite-based Global Positioning System (GPS). Using GPS technology, the OBUs determine the kilometers that have been driven, calculate the toll fees and rates, and then transmit the information to the NATCS computer center. Each vehicle will be charged from the highway entrance up until the end of the highway. In order to identify the plate numbers of trucks, the system has control gates equipped with digital short range communication

(DSRC) detection equipment and high resolution cameras [3] [5]. Due to the technical specifications, this system incurs a high cost for motorists.

ii. South Luzon Expressway

The ETC system used in the Philippines has been implemented at the South Luzon Expressway (SLEX) since August 2000. The ETC is referred to as the E-PASS system, which uses Transcode technology. Here, electronic transponders are placed in front of a vehicle's rearview mirror. Each time a vehicle enters the toll booth, the tag is read by the receiver, automatically identifying the account and debiting the toll fee amount from the corresponding account. Once the amount has been debited, the control gate will lift and the vehicle is allowed to pass through [3] [6].

Above case studies discussed here simply present a general idea of how ETC schemes have been implemented in different parts of the world.

II. PROPOSED SYSTEM

It is evident that toll plazas are usually bottlenecks for traffic flow and the situation is bound to worsen. The usually sought alternative to deal with this is to build additional lanes. However, this option requires a huge investment and it does not even attempt to tackle the major issue – limited throughput of manual transactions. Under such circumstances, automating the entire process, independent of human intervention, seems a much more viable alternative. Toll road operators can create conditions that are conducive to improving the flow of traffic along the road and reducing journey times for the customer by eliminating the need for toll road users to stop and pay at a tollbooth altogether. Such schemes are in operation across the world, with the majority operating in Europe and the United States. Building up on these, the study intends to propose a feasible scheme that can be put into effect resulting in hassle-free operation.

One aspect that demands attention here is whether to employ a hybrid or a fully automated approach. The term ‘fully automated’ is a catchall to describe a system that doesn’t involve manual payment, although not all users would be expected to opt into the system and provisions would have to be made for infrequent/occasional users. For such systems, the success rate would depend large extent on how vehicles that are not equipped and enrolled for electronic tolling are dealt with together with the onus of deploying mechanisms that accurately identify violations [1].

The current toll collection system requires each vehicle to stop at the toll junction and pay taxes, which results in long queues and heavy congestion all throughout the day. To avoid this, we need to automate the process so as to eliminate the delays involved. The idea is to use an RFID tag, say embedded in the license plates or affixed on a corner of the windshield, and as the vehicle passes through the toll junction, the tag is scanned by RFID readers mounted on either end and an ID, unique to each tag, is then sent to the server via an on board WiFi module.

This ID is used as a primary key to look up a database, also hosted on the server, to fetch the associated user’s account details and open a secure transaction in order to deduct the toll amount. All this happens in “real-time”, as the vehicle passes through the toll junction. This, therefore, allows speedy passage of vehicles eliminating the heavy congestion. Thus,

this project proposes the replacement of the currently employed tedious system with a low-cost, as well as a low-power consuming system, which not only speeds up the entire process but also aims to eliminate traditional book keeping and provision daily traffic analysis.

III. OBJECTIVES & SCOPE

The purpose of this research is to:

1. Identify and assess available technologies and methodologies which are in practice in and around the country for toll collection
2. Develop recommendations for the best way to automate the process.
3. Implement a prototype demonstrating the findings.
4. Identification of vehicle theft detection.

The study does not address any regulations concerning the technologies or methodologies proposed and those governing the implementation of tolls in the country. The scope is bound to the implementation of a prototype that aims to redress the impediments in the existing toll collection system.

IV. METHODOLOGY

Automated toll collection system is an adaptation which works in a manner similar to the traditional ETC except that it employs RFID for vehicle identification. The in-vehicle transponder consists of a passive RFID unit that operates in the 125 KHz radio frequency (RF) band using dedicated short-range communication (DSRC) protocols. Stored in this RFID transponder is a unique identification number, which is used as an index to a database containing the user’s details. The RCU emits radio frequencies in the same RF band as the transponder to communicate with it and use its identification number to assess the toll. The RCU emits radio frequencies in the 125 KHz RF band. As the transponder (vehicle) comes in the RCU’s proximity, it draws power from the signals emitted and reflects its identification code back to the RCU. The RCU decodes this information and stores it in the buffer. The identification code is then fetched from the buffer and used as an index to look up a central database.

The associated user’s details are retrieved and the toll is assessed accordingly. The user is immediately notified of the transaction via SMS and/or email. The system also determines whether the vehicles passing are enrolled in the program, and gathers information for further collection or enforcement. Suppose the transaction is successful, the barricade mechanism senses the presence of the vehicle and allows it to cross the junction. However, if the transaction fails due to insufficient balance or unregistered vehicle, the vehicle is directed towards a separate lane for manual collection of toll with a service penalty.

All the data is hosted on an Amazon Web Services (AWS) Relational Database Service (RDS) instance running MySQL. This database instance is linked to the web portal as well as ATCS (Automated toll collection system) thereby maintaining consistency and at the same time allowing concurrent access. Users register themselves for the ATCS program via the web portal. Once registered, the user’s details are verified, the user is assigned a unique identifier and an in-vehicle RFID unit is shipped to the user’s address. Hereupon, the user

simply has to maintain sufficient balance in their accounts and zip through the toll plazas without having to stop to pay the toll. The web portal also provisions the users to easily recharge their accounts, check their billing histories and add new vehicles. For the toll operators, the portal provisions centralized control, improved audits, daily logs and vehicle tracking.

The system also makes use of two other services from AWS: Amazon Simple Notification Service (SNS) and Amazon Simple Email Service (SES). Rather than integrating a GSM module into the system and programming it separately or dealing with the mechanics of SMTP, Amazon SNS and SES were resorted to as simple yet effective means of notifying users of their transactions.

A Raspberry Pi will serve as the heart of the system, handling all communications with the cloud server. A web portal will also be designed allowing users to view a summary of their monthly toll deductions, a facility to recharge their RFID prepaid cards, and, for the users who have opted for the 'postpaid' billing method, a provision to pay their toll bills via credit/debit cards, net banking, and a variety of different means.

This portal would also allow administrators to view daily logs and interactive analysis reports, and address user grievances. The users would also be notified immediately, via SMS and/or E-mail, of the toll amount deducted as they cross the junction.

V. RESULTS & SIMULATIONS

Fig. 1 Web Portal Login



Fig. 2 Theft Detection



Fig. 3 Vehicle location details

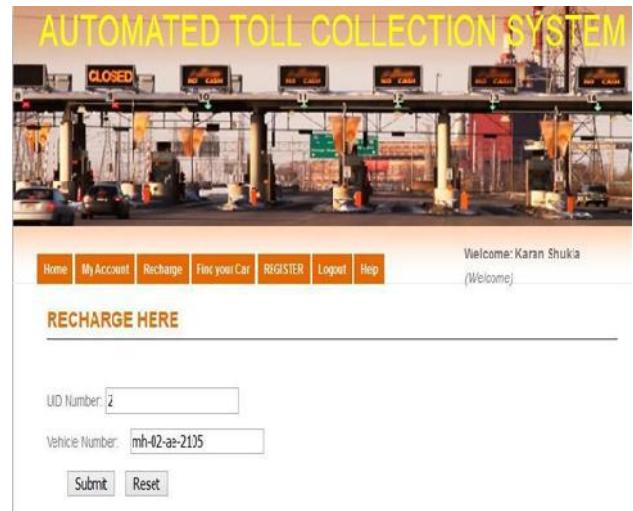


Figure 4: Recharge RFID tag



Fig. 5 Payment details



Fig. 6 ATCS Prototype

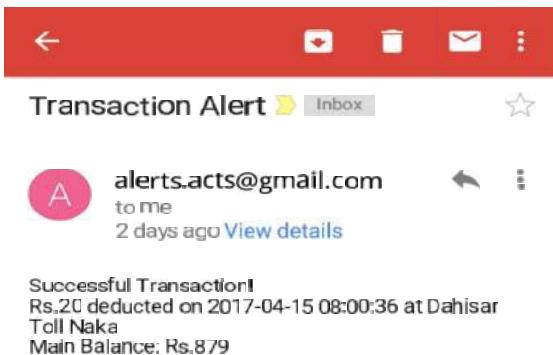


Figure 7: ATCS Transaction Alerts

VI. FUTURE SCOPE

1. GPS tracking of vehicles for theft detection
2. Traffic pattern analysis for intelligent re-routing
3. License plate readers employing OCR

VII. CONCLUSION

Electronic toll collection provides substantial advantages over manual toll collection, including reductions in transaction times, waiting times, fuel consumption, traffic congestion, air pollution and operational costs. The greatest benefits from ETC are achieved with open-road tolling, which minimizes traffic impact as well as the footprint of the toll plazas. This also allows leveraging the existing facilities while implementing ETC.

Another point of concern is the vulnerability of the system to skimming and eavesdropping if the communication channel between the RFID tag and the reader is not sufficiently encrypted. A major decision point, however, is determining how to deal with vehicles that are not equipped or enrolled for ETC. Either prohibit such vehicles from using this facility or set up the hybrid system for cash payments.

REFERENCES

- [1]. Joseph C., Candice W., Natasha M., "Technology Scan for Electronic Toll Collection", Kentucky Transport Center, Research Report KTC-08-15/SPRF359-08-IF, 2008
- [2]. Athanassiou, Bagchi, Bates, "The Role of Tolling Strategies in the Success of Electronic Toll Collection Systems", Association for European Transport and Contributors, 2005
- [3]. Khadijah Kamarulazizi, Dr. Widad Ismail, "Electronic Toll Collection System using Passive RFID Technology", Journal of Theoretical and Applied Information Technology 2005 JATIT & LLS
- [4]. P. Khalil, C.W. Michael, H. Shahriyar "Toll Collection Technology and Best Practices", Project 0-5217: Vehicle/License Plate Identification for Toll Collection Application, January 2007
- [5]. N. Gabriel, I. Mitaszewska, K. Tomasz, "The Polish Pilot Project of Automatic Toll Collection System", Proceedings of the 6th International Scientific Conference TRANSBALTICA, 2009.
- [6]. D.D.E. Crispin, U.M. Aileen, G.S. Ricardo, J.M Jim, S.P. Hilario, "Allocation of Electronic Toll Collection Lanes at Toll Plazas Considering Social Optimization of Service Times and Delays", Proceedings of the Eastern Asia Society for Transportation Studies, vol. 5, pp. 1496–1509, 2005.
- [7]. W.C Tang, T.V. Ho, "Electronic Toll Collection System", US Patent, US 7233260 B2, June 2007.
- [8]. Al-Deek, H. M., A. A. Mohamed, et al. "Operational benefits of electronic toll collection: case study", Journal Of Transportation Engineering 123(N6)
- [9]. Chaodhary, R. H., "A Model for the Benefits of Electronic Toll Collection System", Department of Industrial & Management Systems Engineering, University of South Florida.
- [10]. AungMyint Win, Chaw MyatNwe, KyawZinLatt, "RFID Based Automated Toll Plaza System", International Journal of Scientific and Research Publications, Volume 4, Issue 6, June 2014.
- [11]. David Fenno, "Electronic Toll Collection Systems", Department of Civil Engineering, Texas A&M University, August 1993.
- [12]. "FASTag to boost Safe riding in India" Prashant Nayakanti, LinkedIn, July 29, 2015: <https://www.linkedin.com/pulse/fastag-boost-safe-riding-india-prashant-nayakanti>.
- [13]. "Expressway e-toll collection has just 68 takers" Manish Umbrajkar, TNN, Apr 11, 2015: <http://timesofindia.indiatimes.com/city/pune/Expressway-e-toll-collection-has-just-68-takers/articleshow/46884067.cms>
- [14]. "Private banks fail toll scheme, PSUs in fray" Dipak K Dash, TNN, Aug 25, 2015.

A Proposal For Monitoring System For Secure Health Using IOT

VineshBagul

*Department of Information Technology
Thakur College of Engineering and Technology
Kandivali (E) Mumbai-400101 Maharashtra, India
vineshbagul6@gmail.com*

Rajesh Bansode

*Department of Information Technology
Thakur College of Engineering and Technology
Kandivali (E) Mumbai-400101 Maharashtra, India
rajesh.bansode@thakureducation.org*

Abstract: The idea of this research is to develop smart health centers in India. The proposal basically exploits the idea of Internet of things and aims to use the existing technology and background. The work has been motivated by the fact that our country is inadequate as far as number of comfort focuses relating to the number of inhabitants in nation and in towns particularly. More important if a health center is available, doctors generally either are not willing to serve the rural areas or they are not available 24x7 leading to rural population turning to cities even for casual illness. This adds to long queues in hospitals and lot of annoyance to patients. In order to avoid above mentioned limitations, the proposed model aims to address the patient without even physically visiting the health center and also prescription can be generated in the absence of doctor. It is expected that such a smart system would not only facilitate sick but also will add prosperity indirectly to villages. The proposed framework would definitely bring a significant change in the health care sector of India.

Keywords: IOT, RFID Tag, LoWPAN, CoAP Protocol, Contiki, Mobile Technology.

I. INTRODUCTION

Internet of Things (IOT) is procured widespread popularity among research community because of its potential to digitize real world physical objects around us. IOT has emerged due to present wireless telecommunication services and ubiquitous presence of Internet. Wireless sensor networks, RFID tags, actuators and various handheld intelligent devices such as mobile phones, PDAs, Tabs etc. are leading to the emergence of IOT. IOT seems appealing and its emergence is being accepted by research and industries, due to its impact on our day to day life both in production and consumption processes. It has opened up wide spectrum of innovative scenarios to improve quality of human life. Sensor systems have effectively demonstrated their greatness in making human life simple in routine undertakings, for example, controlling water tanks, to spare power at open spots like exhibition halls, libraries and additionally basic errands, for example, living space observing, controlling/helping modern

procedures. Intelligent devices such as mobile phones have been transformed from embedded keypad based traditional phones to lightweight touch screen based devices. Applications of sensors in our day to day life are numerous which indicates their importance, however they are constrained due to limited battery life which limits their use. Radio Frequency Identification Tags (RFID) innovation gave substitute to sensors as these labels can be utilized to distinguish, track and find any protest utilizing interesting Electronic item code (EPC) which is encoded in these labels.

RFID tag comprises of a small chip, an antenna and a cover for encapsulating chip and antenna. Antenna receives signal from RFID reader device and transmits the tag ID to it. These tags can be either active or passive based on power source. Dynamic labels are related with a battery life and in this way rely upon it for their lifetime, much the same as sensors. Be that as it may, inactive labels get vitality from per user gadget either through attractive enlistment or electromagnetic wave catch procedures [4]. Flag got by RFID receiving wire, delivers a current in it through enlistment, which is additionally used by the radio wire to return the label ID to the per user.

This technique can help transmit tag Id to a radio range of few kilometers. Thus RFID tags eliminate battery limitation of sensors, further being small in size they can be embedded in any real life object for its monitoring. Thus RFID tags are helping great deal to convert every real life physical thing into digital entity. Such RFID detecting items will shape their RFID sensor systems with per user gadgets as sinks of information produced. Rise of these RFID sensor systems, in our everyday life will fill the hole in inescapability of Internet and will enable IOT to spread its underlying foundations in our society. ‘Anytime, Anywhere, Any media’ computing has turned into reality with every object embedded with either RFID tag or sensors, these when combined with already existing wireless communication technologies make everything digitized and on Internet.

This gives avenue for large range of inventive applications such as keen homes, E-healthcare, activity observing and route management, resource management at retail stores, automated checkouts at shopping centers, condition based

maintain of vehicles are some possibilities. Applications of divided into four categories i.e.

- Smart environment (home, office, plant) domain
- Transportation and logistics domain
- Healthcare domain
- Personal and social domain

The current proposal focuses on developing smart healthcare centers and it outlines a framework for mobile based portable health care services for rural population in India. Next section provides an overview of relevant literature in this field. Section 3 provides proposed framework and section 4 finally concludes with Next section presents a proposal for Mobile E-care Health Services using IOT.

PORPOSED SMART HEALTH MODEL

Cisco Packet Tracer

Packet Tracer complements the Networking Academy curricula, allowing instructors to easily demonstrate complex technical concepts and networking systems design. Instructors can customize individual or multiuser activities. Helps to build, configure, and troubleshoot systems utilizing virtual equipment and recreated connections, alone or as a team with other. Packet Tracer offers an effective, interactive environment for learning networking concepts and protocols. Most importantly, Packet Tracer helps instructors create their own virtual “network worlds” for investigation, experimentation, and clarification of networking ideas and technologies. HP’s Earth initiative, just to list a couple. European commission and Chinese Government is also making efforts in this direction. However, advancements in IOT is likewise raising trust and security issues all the while. Institutionalized conventions and administration systems are required for IOT to work at worldwide level. Study on IOT introduced by McKinsey Global Institute [3] highlighted that most IOT information being caught today isn’t utilized at present. By and by the caught information is utilized just for irregularity discovery and control; however it may be used for optimization and prediction which is of more importance. Further, they pointed that there is large scope for IOT in developing economies such as India. The critical investigation of available literature clearly reflects that IOT is the demand and strong requirement of developing countries and there is a huge gap prevailing between theory and practice. The proposal submitted aims to fulfill this gap in one of the domains i.e. health care. advancement of IOT must be a consequence of synergistic in various fields such as telecommunications, informatics, electronics and social science. Coetzee and Eksteen [2] have expounded IOT space and underscored that different application areas such as Green IT, energy efficiency and logistics have already started gaining benefits from it. Because of vast capability of this area, IOT has grabbed higher priority on the research agenda of academia, industry and governments such as IBM’s Smarter Planet, Microsoft’s Eye-on-Earth platform and

IOT.

requirements to develop the proposed smart healthcare center and its future scope.

II. RELATED WORK

This section explores work already done in the field of IOT[1] presented a survey on Internet of Things highlighting the most engaging purpose of IOT which is the integration of several technologies and communication solutions. Their work emphasized that any contribution towards

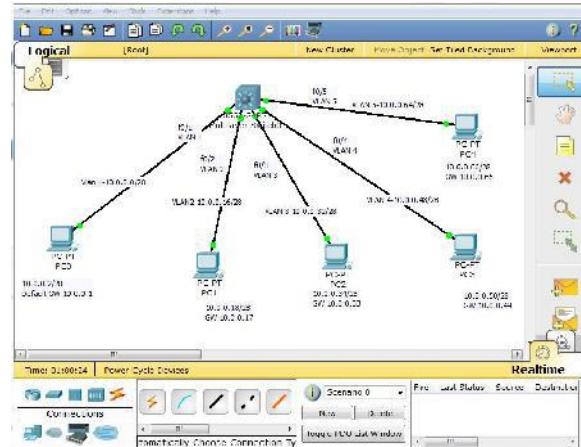


Fig. 1. Packet Tracer’s drag-and-drop interface allows instructor to configure and approve framework design

Cisco Packet Tracer provides two operating modes to visualize the behavior of a network—real-time mode and simulation mode. In real-time mode the network carries on as genuine devices do, with immediate real-time response for all network activities. The real-time mode gives students a viable other option to genuine equipment and allows them to gain configuration practice before working with genuine equipment. In simulation mode clients can see and control time intervals, the inner workings of data transfer, and the propagation of data across a network. Cisco Packet Tracer is a powerful network simulation program that allows instructors to experiment with network behavior and ask “what if” questions. Packet Tracer provides simulation, representation, writing, appraisal, and coordinated effort capacities to encourage the educating and learning of complex innovation ideas.

Table 1: Cisco Packet Tracer supports the following protocols:

Layer	Cisco Packet Tracer Supported Protocols
Application	FTP, SMTP, POP3, HTTP, TFTP, Telnet, SSH, DNS, DHCP, NTP, SNMP, AAA, ISR VOIP, SCCP config and calls ISR command support, Call Manager Express

Transport	<ul style="list-style-type: none"> TCP and UDP, TCP Nagle Algorithm & IP Fragmentation, RTP
Network	<ul style="list-style-type: none"> BGP, IPv4, ICMP, ARP, IPv6, ICMPv6, IPsec, RIPv1/ v2/ng, Multi-Area OSPF, EIGRP, Static Routing, Route Redistribution, Multilayer Switching, L3 QoS, NAT, CBAL , Zone-based policy firewall and Intrusion Protection System on the ISR, GRE VPN, IPsec VPN
Network Access/ Interface	<ul style="list-style-type: none"> Ethernet (802.3), 802.11, HDLC, Frame Relay, PPP, PPPoE, STP, RSTP, VTP, DTP, CDP, 802.1q, PAgP, L2 QoS, SLARP, Simple WEP, WPA, EAP

Benefits of using a packet tracer

Packet tracer provides network simulation and visualization. It can be used to enhance and improve the practical knowledge of computer networking principles among students. Moreover, students can design miniprojects with solutions with more innovation and creativity. As with other tools, students are able to understand the use of different networking protocols but they are not able to understand the application of these conventions in the genuine networks, along these lines packet tracer can be utilized to design and configure a network, and comprehend the use of different conventions. Packet Tracer can further be used, to understand the difference between different networking devices like hubs, switches, routers etc and their appropriate use while connecting various Computers to design a network. How to assign logical address to various networking devices like computers, routers etc appropriately. While moving from source to goal, which route is selected by a packet depending on various routing protocols? Kind of link to be utilized while associating networking devices. Checking connectivity between different networking devices by running different networking tests. Basic networking concepts like DNS, DHCP, NAT, routing etc can be easily explained by using packet tracer and students can build, configure and troubleshoot networks using packet tracer. It also makes teaching easier, students can create their own scenario based labs and provides real simulated and visualization environment.

FEATURES OF PACKET TRACER 6.2 (ADVANCED PACKET TRACER OPTIONS)

A. New devices and protocols In February 2015 Packet Tracer 6.2 was released in two versions - Student and Instructor. PT 6.2 includes an ASA 5505 firewall with CLI configuration (but no ASDM or CCP tools) [8,9]. It also includes a netflow collector as a desktop application in the server device, routing protocols for IPv6 (OSPFv3,

EIGRPv6, RIPng), DHCP snooping, IPv6 CEF and IPsec commands. This PT version includes a new Cisco 819 ISR router with a embedded wireless access point and some new devices: 3G/4G cell tower, Central Office (CO) server and 3G/4G support for end devices (smartphone, tablet, server etc.), [table 1].

B. Workspaces and operating modes PT has two types of workspace: logical and physical. Logical workspace allows students to build logical network topologies by placing, connecting, and clustering virtual devices. The physical workspace provides a graphical visualization of the logical network and represents how network devices would look in a real environment. The physical view enables geographic representations of networks, including multiple cities, buildings and wiring closets.

There are two operating modes to visual representation of a network behavior: real-time and simulation mode. The real-time mode enables students to gain configuration practice because the devices in network look and behave exactly the same as real Cisco devices. In simulation mode they can see, control and analyze time intervals and propagation of data across a network, and can learn how to troubleshoot network failures. This significantly helps most of them understand the fundamental concepts behind network operations. In Figure 1 the example of the PT network devices topology map is shown. This topology represents to how to make site-to-site IPsec VPN utilizing ASA 5505 firewall. A small branch office is securely connected to the enterprise campus over the internet using a broadband DSL modem connection [5].

C. Modular network devices Graphical representations visually simulate network equipment, and offer the capability to insert interface cards into Cisco modular routers and switches. Selecting switches or routers from the device-type selection box lists both Cisco devices and some devices labeled Generic. These are custom PT devices running on Cisco IOS, but the slots that hold the modules are different.

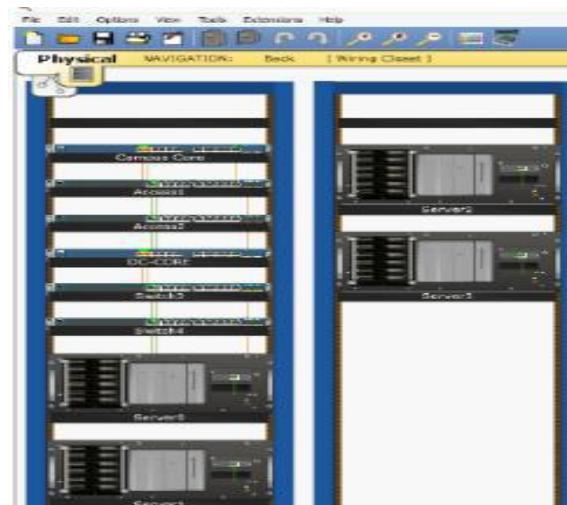


Fig.2. The physical workspace (Wiring Closet) with graphical view of network devices and connections

D. This new feature has opened the door to develop many interesting new activities such as interactive and dynamic troubleshooting and serious gaming for introductory networking classes. PT multi-user activities can make networking more interesting to learn and lead to greater student engagement. The multiuser capacity permits association of remote instances of PT on separate machines [4-5]. The Activity Wizard guides students through the creation of an assessment. It is made accessible by navigating to Extensions Activity Wizard. The Activity Wizard allows students to author their own learning activities by setting up scenarios using instructional text, and creating initial and final network topologies and predefined packets. Instructions in Activity Wizard are built in simple HTML format and also includes grading and feedback capabilities [6]. The physical workspace is divided into four layers to reflect the physical scale of real life environments: Intercity, City, Building and Wiring Closet. The Wiring Closet is the final layer that contains devices placed in the logical topology and it doesn't have any specified area (Figure 2)

Contiki

Contiki is an operating system for networked, memory-constrained systems with a focus on low-power wireless Internet of Things devices. Extant uses for Contiki include systems for street lighting, sound monitoring for smart cities, radiation monitoring, and alarms. It is open-source software released under a BSD license.

Contiki is designed to run on types of hardware devices that are severely constrained in memory, power, processing power, and communication bandwidth. A typical Contiki system has memory on the order of kilobytes, a power budget on the order of milliwatts, processing speed measured in megaHertz, and communication bandwidth on the order of hundreds of kilobits/second. Such systems include many types of embedded systems, and old 8-bit computers. Contiki provides three network mechanisms: the uIP TCP/IP stack,[6] which provides IPv4 networking, the uIPv6 stack,[7] which provides IPv6 networking, and the Rime stack, which is a set of custom lightweight networking protocols designed for low-power wireless networks. The IPv6 stack was contributed by Cisco and was, when released, the smallest IPv6 stack to receive the IPv6 Ready certification.[8] The IPv6 stack also contains the Routing Protocol for Low power and Lossy Networks (RPL) routing protocol for low-power lossy IPv6 networks and the 6LoWPAN header compression and adaptation layer[8] for IEEE 802.15.4 links. Rime is an alternative network stack, for use when the overhead of the IPv4 or IPv6 stacks is prohibitive. The Rime stack provides a set of communication primitives for low-power wireless systems. The default primitives are single-hop unicast, single-hop broadcast, multi-hop unicast, network flooding, and address-free data collection. The primitives can be used on their own or combined to form more complex protocols and mechanisms.[9] An autonomous discovery of devices and services inside a network domain is a central usability criterion. Applications and users need

mechanisms to identify new devices and gather relevant information to access offered services. For Internet of Things scenarios, this means that devices need to be discovered in wired and wireless networks, with support through pre-configured infrastructure as well as possibly without (ad hoc) any infrastructure support. Next to these requirements, a discovery solution should also work in compliance with current systems. We thus favor using IP-based solutions to facilitate an easy integration into current networks and systems. choosing our own working environment as a testbed for a typical IoT scenario. In our “smarter workplace” scenario [10], different devices (sensors as well as actuators) are distributed to support workers in their daily routines, as Figure 3 illustrates. This leads to a case comparable to common “smart home” scenarios. Smart homes are widely covered in research, although normally utilized technologies are restrictive and not IP-enabled [11], thus requiring gateways and protocol adapters. As we omit protocol gateways in favor of an end-to-end connection with common Internet protocols, we focus on setting up a test bed of different devices connected over IP links. We connect various stationary computers and workstations, mobile devices (e.g., smartphones, netbooks), as well as embedded devices, sensors, and actuators over their respective communication technologies. Figure 3 depicts the architecture of the system. It is important to note that we cannot exclude bridges to interconnect different technologies physically. Assume that one of the following bridging techniques is given to interconnect different communication technologies: interconnection via USB/serial port, support for multiple radio technologies within a single device, or embedding the sensor/device in power outlets for a connection over power line communication.

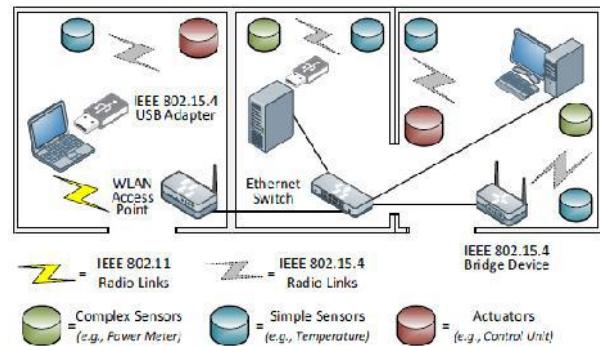


Figure 3-Systeme Architecture

Advantage of contiki

A.Better Encryption and security

Encryption and security support was always lacking in the previous versions of Contiki. Contiki 3.0 brings much needed support for 802.15.4 link layer encryption with AES128, available in software for all platforms and hardware-accelerated for those platforms that support it (primarily the TI CC2538).

B.New Network Protocols

The Contiki 3.0 release adds support for the IP64/NAT64/DNS64 set of protocols that make it feasible to connect Contiki IPv6 mesh networks directly to IPv4 networks, such as the Internet, without the need for a special translation proxy. The protocols operate directly at the IP layer to create a ordered translation between the mesh and the Internet. Alongside, device driver support for the ENC26j80 Ethernet chip has been added, to make it easy to connect Contiki mesh networks to Ethernet networks. For application layer connectivity, support for the MQTT protocol, a new HTTP socket module, and an updated CoAP module has been added. New socket APIs for TCP and UDP sockets make network programming easier than in the past.

The IPv6 mesh networking support has been massively reworked and improved. The RPL mesh routing protocol has seen many bugfixes and stability improvements. Support for new IPv6 mesh multicast mechanisms has been added. Many internal improvements to the IPv6 stack has been added.

C. Internal Changes

The structure of the code has seen a rework, with a new module system that makes it easier to selectively use only parts of the code. Instead of using C macros and #defines to select whether to use the IPv6 or IPv4 stack, the new module system uses Makefile settings, which makes compilation more efficient and the code easier to read. The radio API has been updated to better match the way the radio duty cycling protocols use the radio. For example, the previous radio API lacked a clean way to set the radio channel, which now is part of the new API. The idea is to update a number of old hardware platforms that were not used anymore and dropped a bunch of examples that weren't relevant any longer. We also added a bunch of new regression tests to ensure the continued stability of the system.

Applications

A. There are several applications that come packaged as part of Contiki. These include a small Web browser, Web server, calculator, shell, telnet client and daemon, email client, vnc viewer and ftp.

B. Developers can find tools such as Cooja simulator for application development. Contiki also provides mechanisms to measure power for various operations that help developers produce applications that are powersensitive. Software bundle also comes with a UNIX-style shell for interacting with OS and debugging.

Operating system functions include:

1. Process and memory management. Contiki supports memory block allocation and standard 'C' style allocation with malloc(). The concept and implementation of 'protothreads' has been introduced keeping the low system

requirements in mind. 'protothreads' are C-language-based implementation minimizing the overhead of multi-threaded programming.

2. Communication management.

Contiki supports both IPv4 and IPv6 stack implementations, which include TCP, UDP and HTTP protocols with the smallest footprints. The IP stack is standards-compliant. The OS also supports low-power variants such as 6LoWPAN.

3. File system management

Not all devices as part of IoT have luxury of persistent storage such as Flash. But for such devices, Contiki provides support through the Coffee Flash file system.

USING CoAP

Overview of CoAP (Constrained Application Protocol)

It is a software protocol intended to be used in very simple electronics devices that allows them to communicate interactively over the Internet. It is particularly targeted for small low power sensors, switches, valves and similar components that need to be controlled or supervised remotely, through standard Internet networks [12]. Therefore, efficiency is very important. CoAP can run on most devices that support UDP or a UDP analogue. Message types involved in CoAP are Confirmable, Non confirmable, Acknowledgement, Reset messages. Confirmable requires Acknowledgement whereas Non Confirmable doesn't require Acknowledgment. Reset message indicates missing of few contexts.

- GET: The GET method helps in retrieving information from that of server.
- POST: The POST method requests the server to provide data for the user.
- PUT: The PUT method helps in updating or creating the information in the corresponding URI
- DELETE: The DELETE method helps in erasing the information stored.

The usage of CoAP in patient monitoring application is to form a client-server relationship, whereas Doctors use the browser as client and the server program runs in the GUI for updating the browser dynamically. Thus the Method definitions are used efficiently for remote monitoring application.

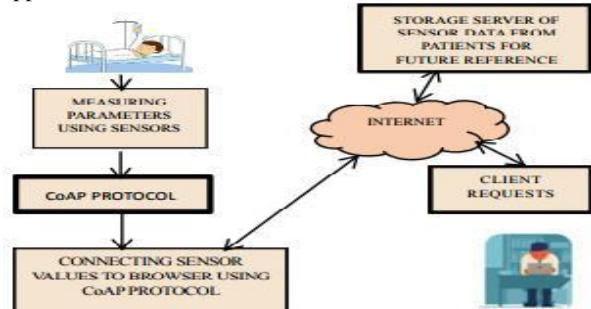


Fig. 4: Architecture of proposed system using CoAP

As per Figure 3, Patient is monitored by the sensors mounted on them. Servers will host the information on the network. The server collects the data from the sensors. On the other side, Doctors are acting as clients. If doctors request server for data, the CoAP protocol helps to connect server with client to display the current status and check for variation of sensor values [11]. If there is any variation, Doctors can give medication indicating emergency situation for the patient. This medication is done in browser using CoAP Protocol. The data of patient is stored in some data storage device for future purposes.

In multi-hop flat wireless topology as shown in figure 4, there are one or more intermediate nodes along the path that receive and forward packets via wireless links. During transmission and reception in multi-hop networks, nodes communicate with each other using wireless channels and do not have the need for common infrastructure or centralized control. The above Figure 4 explains the Multi-hop topology with Border router, Client and server.

In this paper we gave an overview of current CoAP implementations. Although CoAP is a very new protocol, over 20 different implementations in various programming languages are known for devices ranging from resource-constrained devices such as wireless sensor nodes up to smartphones and servers. This implies that CoAP is going to be used in many different applications and domains [13]

III. CONCLUSION

Packet Tracer has two workspaces—logical and physical. The logical workspace allows users to build virtual network topologies by placing, connecting, and clustering virtual network devices. The physical workspace provides a graphical physical dimension of the logical network, giving a sense of scale and placement in how network devices such as routers, switches, and hosts would look in a real environment. The physical view additionally gives geographic portrayals of systems, including numerous urban communities, structures, and wiring storage rooms. Contiki is an open source software project with the vision to create a minimalist and working OS for wide deployment. This project has broken many myths about the smallest footprint in which an OS can be stored and made to function. One such software is the uIPv6—world's smallest IPv6 stack—which is integrated into Contiki. While there are many similar OS such as TinyOS, what makes Contiki different is the completeness and flexibility it offers to the programmers. CoAP is a very new protocol, over 20 different implementations in various programming languages are known for devices ranging from resource-constrained devices such as wireless sensor nodes up to smartphones and servers. This implies that CoAP is going to be used in many different applications and domains. The proposed model can be easily implemented using existing RFID-LoWPAN, CoAP and an expert health care system such as Mycin. A new intelligent e-health system can also be designed considering the health issues and environments. The proposition would go about as a guide to de-limited and would likewise contribute towards Digital India.

REFERENCES

- [1] Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- [2] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things—promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE.
- [3] Babar, S., Stango, A., Prasad, N., Sen, J., & Prasad, R. (2011, February). Proposed embedded Security framework for internet of things (IOT). In Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on (pp. 1-5). IEEE.
- [4] Ant, R. (2006). An introduction to RFID technology. *Pervasive Computing*, IEEE, 5(1), 25-33, January-March, 2006.
- [5] Application_of_Cisco_Packet_Tracer_62_in_teaching_of_advanced_computer_networks
- [6] Jara, A. J., Zamora, M. A., & Skarmeta, A. F. (2011). An internet of things---based personal device for diabetes therapy management in ambient assisted living (AAL). *Personal and Ubiquitous Computing*, 15(4), 431-440.
- [7] The Internet of Things: Mapping the Value Beyond the Hype, Executive Summary by McKinsey Global Institute, June 2015.
- [8] Ukil, A., Sen, J., & Koilakonda, S. (2011, March). Embedded security for Internet of Things. In Emerging Trends and Applications in Computer Science (NCETACS), 2011 2nd National Conference on (pp. 1-6). IEEE.
- [9] Istepanian, R. S., Zitouni, K., Harry, D., Moutosammy, N., Sungoor, A., Tang, B., & Earle, K. A. (2009). Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. *Journal of Telemedicine and Telecare*, 15(3), 125-128.
- [10] Istepanian RSH, Jara A., Sungoor A, Philips N. (2010). Internet of Things for M-health applications (IoMT).
- [11] Chander RPV, Elias S, Shivashankar S, Manoj P (2012) A REST based design BonjourContiki: A Case Study of a DNS-based Discovery Service for the Internet of Things Distributed and Grid Computing pp: 337-342.
- [12] Dimcic T, Krco S, Gligoric N (2012) CoAP (Constrained Application Protocol) implementation in M2M Environmental Monitoring System. E-society Journal 21.
- [13]. Industry Adoption of the Internet of Things: A Constrained Application Protocol Survey.

Wireless Sensor Network for Environmental Monitoring and Control

Pankaj Malviya

*Department of Information Technology
Thakur College of Engineering & Technology Mumbai*

Kautilya Joshi

*Department of Information Technology
Thakur College of Engineering & Technology Mumbai*

Deep Modi

*Department of Information Technology
Thakur College of Engineering & Technology Mumbai*

Zahir Alam

*Department of Information Technology
Thakur College of Engineering & Technology Mumbai*

Abstract-Wireless Sensor Networks (WSNs) have attracted much attention in recent years. The potential applications of WSNs are immense. They are used for collecting, storing and sharing sensed data. WSNs have been used for various applications including habitat monitoring, agriculture, nuclear reactor control, security and tactical surveillance. The WSN system developed in this project is for use in environmental parameter monitoring such as temperature, humidity, and light intensity detection. The aim of this project is to create a wireless sensor network for environmental sensing and control. The IEEE 802.15.4 (ZigBee) standard has been implemented for wireless communication among the nodes with the network motes (sensing nodes) designed to include temperature sensing, humidity sensing, and light intensity sensing functionalities by using DHT11 sensor and BH1750 sensor respectively. The base station hosts a web application that enables the user to check real time sensor data as well as control the connected devices via the website. The website also displays data about the connected devices such as when it was last switched ON/OFF, how much time has passed since then. The sensor successfully controls the connected home appliances with the future scope of expanding it throughout the city to sense and measure environmental data such as air pollution or water pollution in local water bodies.

Keywords- Wireless Sensor Network, ZigBee, Arduino, DHT11, BH1750 sensor

I. INTRODUCTION

Wireless sensor network is a technology for wide range of wireless environments. Recently more research work has been done in direction to develop wireless network that works on low power, low data rate, low cost personal area network. Many organisation have developed WSNs for smart home, smart farm, smart hospital for patient monitoring, for traffic monitoring, fire monitoring in smart cities and home automation. The importance and application has been increased by the recent delivery of the IEEE 802.15.4 standard and the forthcoming ZigBee standard. The ZigBee Alliance has developed very low-cost, very low-power consumption,

wireless communications standard for network and application layer to fulfil the demand of automation and remote-control applications. IEEE 802.15.4 committee started working on a low data rate standard a short while later for physical and MAC sub layer. Then the ZigBee Alliance and the IEEE decided to join forces and ZigBee is the commercial name for this technology.

ZigBee is expected to provide low cost and low power connectivity for equipment that needs very long battery life as several months to several years but does not require data transfer rates as high as those enabled by Bluetooth. ZigBee can also be implemented larger networks than is possible with Bluetooth. ZigBee compliant wireless devices are operated in the unlicensed RF worldwide (2.4GHz global, 915MHz Americas or 868 MHz Europe). The data rate is 250kbps at 2.4GHz, 40kbps at 915MHz and 20kbps at 868MHz.

II. SCOPE

The network of wireless sensory nodes is able to measure temperature, humidity, and light intensity with the help of various sensors. The information collected by the nodes will be transmitted to a base station by the Digi XBee S2 ZigBee wireless modules. The base station will upload the data to a local database and to a web application through which a user can view real time data collected by the sensors independent on location. The user can control an electronic device by sending an SMS to a specific phone number through the web application

III. MOTIVATION

Wireless sensor networks are used in those harsh and hostile environments where wired networks can't be deployed. Environmental monitoring and controlling are some of the fields that have most benefited from the evolution of sensors, in particular bio-sensors, and sensor networks. They represent a group of applications that provide great development for scientific communities. The deployment of a network of this nature brings advantages many ways, from wire reduction, which represents a cost reduction, to

robustness, scalability and the data decentralization, circumventing the problem with linking failures.

IV. PROBLEM DEFINITION (PHASE WISE)

There are several WSN hardware platforms available. In our prototype system, we chose our hardware platform to base on an Arduino board as both end nodes and the base station and an XBee S2 modules for wireless communication. Our network consists of various motes with different sensors such as DHT11 for temperature and humidity sensing and BH1750 sensor for Light intensity sensing. Many organizations have developed WSNs for smart home, smart farm, smart hospital for patient monitoring, for traffic monitoring in VANET, fire monitoring in smart cities and home automation. The importance and application has been increased by the recent delivery of the IEEE 802.15.4 standard and the forthcoming ZigBee standard.

A. PHASE 1: (Planning, Analysis, Design, Coding)

Planning: Conceptualizing and designing the architecture for proposed solution. Choosing whether to use Raspberry Pi 3 or Arduino.

Analysis: Critical study, analysis and review of feasibility for proposed solution. Choosing the right type of components such as XBee s2 modules, Arduino Uno, Relay module, DHT11 temperature and humidity sensor, BH1750 photo-resistor for light sensitivity.

Coding: Creating a website to display real time environmental data gathered by the sensor nodes and data about current status of the connected devices. The websites will also have functionalities to control the connected devices.

Design: The application will display real time environmental data and device data with an option to automatically control the device after crossing a threshold value.

B. PHASE 2: (Integration, Testing, Deployment)

Integration – Integrating of various sensors such as temperature sensor (DHT11) and Photo-resistors like BH1750, notifying the user about the temperature or light intensity change in the room.

Testing – Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution.

Perform
alpha testing after completion of prototype.

Deployment – Give the completed prototype to the authorized person for evaluation purpose and beta testing.

V. METHODOLOGY

We present the application of developing a wireless sensor network with temperature, humidity, and light intensity measuring capabilities. In this study, we have developed a small scale wireless network for automating home appliances by sensing environmental data and both, giving the user the control

the devices according to his/her liking or putting the system to an auto mode which controls the connected devices in accordance with the data collected by the sensors by setting threshold values.

Firstly, we collect the environmental data by our sensor motes in areas we want to observe. The data collected by the sensor are then forwarded to the micro controller it's attached to (Arduino Uno in our case) which converts it into suitable form for wireless communication and forwards it to the wireless ZigBee modules. The reduced-function devices then send their respective data to the base station which uploads the received data to the website it hosts.

The website displays the real time data that the base station receives from various motes and it also display some information about the connected devices such as the last time the device was last switched ON/OFF and how much time has elapsed after that.

The website also contains an “AUTO” toggle button which enables the user to enable/disable the system to automate the connected devices according the environmental data received. When ON, the system has control over the devices too meaning that the system is able to turn ON/OFF the devices and notify the user when the observed environmental data values cross a threshold.

VI. FEASIBILITY STUDY

Technical Feasibility - Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

Economic Feasibility - As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

Legal Feasibility - Legal issues can affect a system’s acceptance by Bottom users, its performance, or the decisions on whether to use it in the first place so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any log analysing system.

Operational Feasibility - It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

Social Feasibility - The acceptability of a log analysis system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness.

VII. FLOW CHART

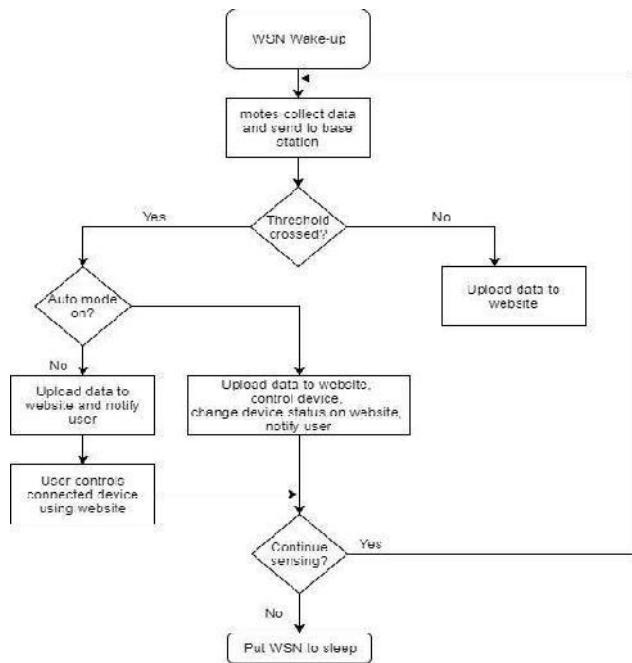


Fig.1 Flowchart of the system

A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to study, document, plan, improve, and communicate often-complex processes in clear, easy-to-understand diagrams. The below figure is the flowchart that explains the basic working of our Wireless Sensor Network.

The below flowchart depicts how our system keeps on monitoring the environment using various environmental sensors until it is switched off. One more thing about ZigBee modules is that they enter sleep mode whenever they are not required to sense or just forward any signal that leads to efficient energy consumption.

VIII. BLOCK DIAGRAM

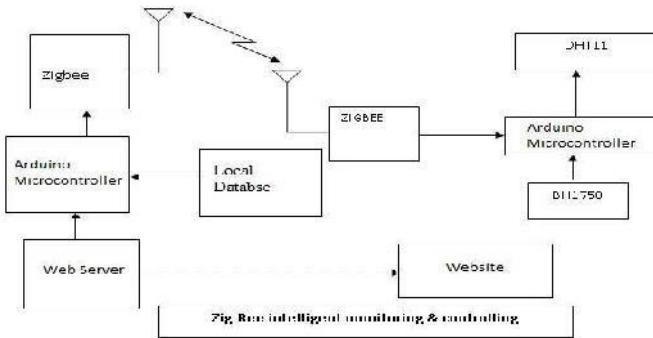


Fig.2. Block diagram of the system

- DHT11 and BH1750:

DHT11 is the temperature and humidity sensor used in our system while the BH175 is a photo-resistor used to detect and monitor light intensity.

- ZigBee modules:

The Digi XBee S2 modules have been used in our system for wireless communication using ZigBee (IEEE 802.15.4/LR WPAN) standard. All the Reduced Function Devices (end nodes) send their respective data to the base station to be uploaded onto the website.

- Base station:

The base station comprises of a ZigBee module attached onto an Arduino microcontroller board. This node also hosts a website which the user can access to check the real time environmental data.

- Website:

The website displays the real time environmental data collected by the sensory motes as well as current status of the connected devices. The website also has an auto button that enables/disables the system to control the connected devices by itself without any human intervention.

X. EXPECTED RESULT

Outputs:

The wireless system will help to monitor neighbouring environment depending on the sensors used. The data that will be collected by our system will be used to control the devices that will result in us being able to control the environment as well.

Outcomes:

Phase 1.1 Planning

In this phase Proposing statement of work, scope definition and scope boundary for planning the prototype from the definition to decide what needs to be done and what not to be done.

Phase 1.2 Analysis

Creating the synopsis, literature survey and feasibility study for proposed solution that will act as an input to the design phase.

Phase 1.3 Design

Proposing design architecture of sensor nodes (motes) for wireless sensor network (WSN) for implementing the surveyed case studies and synopsis Creating a visually appealing and easy to use front-end of the web application deployed by the base station.

Phase 1.4 Coding

Used available open source libraries for respective sensors to write our own code to use the environmental sensors via Arduino.

Phase 2.1 Integration

Integrating of various sensors such as temperature sensor (DHT11), notifying the user about the temperature change in the room.

Phase 2.2 Testing

Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.

Phase 2.3 Deployment

Give the completed prototype to the authorized person for evaluation purpose and beta testing.

c. Percentage improvement in results (comparison of results reported & results achieved)

Identification accuracy in sensors = 71%

XI. FUTURE SCOPE

GSM Module

For future use, a GSM module can also be implemented in our project for the user to control the connected devices directly via Short Message Service (SMS) without logging onto the website. This will further increase the usability of our system by allowing the user to interact with the devices in multiple ways in case of any uncertainty.

Wider range of network coverage

The coverage area is currently restricted due to ZigBee's poor coverage distance. Other wireless technologies like WIFI modules can be used instead of ZigBee to cover a wider range of area covered under the network which may allow the system to be deployed over an entire city for smart use of government controlled devices such as street lights for more efficient energy consumption or fog machines used in polluted cities to be automatically enabled when needed.

XII. CONCLUSION

The wireless network constructed solves the problem of human error in switching OFF appliance by automatically determining when and how to control electronic devices by measuring certain environment parameters. The wireless network constructed provides a way to automate home appliances with the help of sensory motes connected to each by the wireless ZigBee protocol which supports low power consumption, low cost and low but reliable data rate for transmission of data packets. Thus, the proposed solution will be implemented.

REFERENCES

[1]Valdo Henriques and Reza Malekian (Member, IEEE), "Mine Safety System Using Wireless Sensor Network", "Institute of Electrical and Electronics Engineering Access", ISSN: "2169-3536" on Volume: 4, 16 June 2016, Pages: 3511 - 3521.

[2]Omar S. Alwan, K. Prahalad Rao, "Dedicated real-time monitoring system for health care using ZigBee", "Institute of Electrical and Electronics Engineering Access", ISSN:" 2053-3713" on Volume: 4, Issue: 4, 8 2017, Pages: 142 – 144.

[3]S.S.Sarade, Prof.A.C.Joshi, Sachin S. Patil, A.N.Shinde, "Wireless Temperature Monitoring System Using Wireless Sensor Networks", "International Journal of Advanced Research in Electronics and Communication Engineering", ISSN: 2278 – 909X on Volume 1, Issue 4, October 2012.

[4]Murat Dener, "Intelligent Home System Design Using Wireless Sensor Networks", "Research World International Conference, Istanbul, Turkey", ISBN: 978-93-86291-69-1 on Vol. 3, Issue. 2, 27th December 2016.

Design of Guest Tagging System and Surveillance Using RFID

Satyendra Yadav

*Thakur College of Engineering and Technology
University of Mumbai, India*

Lokesh Ucheniya

*Thakur College of Engineering and Technology
University of Mumbai, India*

Akash Singh

*Thakur College of Engineering and Technology
University of Mumbai, India*

Deven Shah

*Thakur College of Engineering and Technology
University of Mumbai, India*

Abstract — Considering the present scenario at Govardhan Eco Village, it has a manual way of authorizing the visitors by providing them simple Identity cards for authentication purposes. Being a 70-acre sustainable farming community, Govardhan Eco Village (GEV), doesn't have an adequate way of controlling the access of the visitors due to the large area and require something which can automatically do it. Around a few hundred guests visit daily and the village is not so secured. To overcome this real-world issue proposed solution is "IoT Based Guest Tagging System". RFID Reader only reads the data through RFID tag but through IoT all data entries will be collected & it will be stored online on remote server. Analysis will be performed on it so as to manage the flow of guests in different zones effectively, reduce the waiting time and serve the visitors in a more convenient manner. The proposed system uses a RFID passive tag and RFID reader to record the current location and timestamp of the visitors (guests). The main idea behind implementing IoT Based Guest Tagging System is to automate the guest monitoring process. It will enhance the existing security system of the Govardhan Eco-Village and also be able track down visitors who are crossing the zones where they are not allowed to go.

Keywords: Radio Frequency ID, RFID Reader, Mass Gatherings, Prepared Plan Evacuation, Precaution Plan Evacuation, Rescue Plan Evacuation, Simulation.

I. INTRODUCTION

Govardhan Eco-village needs to find an eco-friendly cost-effective solution to tag the different categories of guests arriving, and granting them various permission access to different zones in the village. Every type of guest needs to be served in a more convenient manner so as to maximize the level of user experience at the village which is an attraction place for hundreds of devotees. After solving the above problem, we realized Crowd management is a very complex, challenging and costly exercise. If not managed or controlled effectively, it can result into crowd disasters aka stampede like situations. The recent Elphinstone Rail

Stampede in Mumbai, 29 September 2017 is an apt example to illustrate the same.

A radio frequency identification reader (RFID reader) is a device used to gather information from an RFID tag, which is used to track individual objects. Radio waves are used to transfer data from the tag to a reader.

TYPES OF RFID TAGS:

1. Active RFID Tags

Active RFID tags are used on large assets, such as cargo containers, rail cars and large reusable containers, which need to be tracked over distances such as a distribution yard. These usually operate at 455 MHz, 2.45 GHz, or 5.8 GHz frequencies and they typically have a read range of 20-100 m, costing from 10 to 50 US dollars (2005), depending on the amount of memory, the battery life required, whether the tag includes an on board temperature sensor or other sensors and the ruggedness required.

2. Passive RFID Tags

Passive RFID tags contain no power source and no active transmitter, their power to transmit their information, typically between 10 mW and 1 mW, comes from the reader. They are cheaper than active tags, currently costing about 20-40 US cents. They are lower maintenance and much shorter (read access) range than active tags, typically from a few cm to 10 m.

II. PROBLEM DEFINITION

The objective of the problem definition is to find an eco-friendly cost-effective solution to tag the different categories of guests that come to Govardhan Eco Village, namely daily visitors, residential guests, unregistered day visitors, villagers and VIPs. Each category of guests will have access to different zones in the village.

III. PROPOSED SOLUTION

A. Design of Proposed Solution

The proposed solution uses a RFID tag and Long-Range UHF Passive reader to record the current location and timestamp of the guests (visitors). RFID Reader with a range of about 25 m will be present at a center of each prominent location of the Govardhan Eco-Village, thus covering a diameter of about 50 m. Each RFID Tag will be mapped with the details of the Visitor to whom it is assigned.

RFID reader will be used to record the RFID Tag's identity and current timestamp. RFID Reader will send this data along with its deployed location to a centralized database (Server) so that any kind of user can be tracked and their access can be controlled to different zones of the area. Long range communication will be implanted by line of sight communication whereas short range by using the radio frequency communication in walky-talkies which is already used in the village.

For e.g., if the number of users is more in the Prasad Ghar of the village than its holding capacity we can move them to another area for other services.

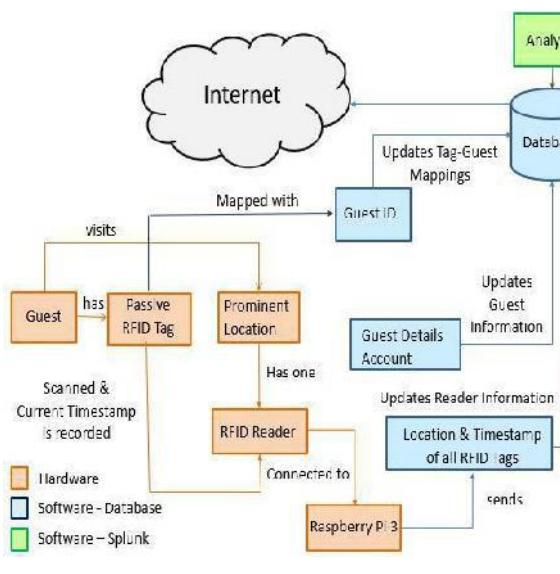


Fig. 1. 500 Guest Details namely Name, Address, Gender, Age mapped with 500 RFID identities in csf.

TABLE I. LOG FILES AND ITS ATTRIBUTES GENERATED WHILE

1	ARUN KHARA	AKOLA	M	54
2	ASHOK MARU	MUMBAI	M	49
3	RAMESH PARDESI	KHANDWA	M	31
4	SANJAY SHAH	KOPARGAON	M	15
5	AYUSH DOSHI	DHULE	M	9
6	MIHIR SAWANT	HARDA	M	22
7	OMKAR JANGAM	MADRAS	M	33
8	SHALINI AYYAR	KOPARGAON	F	29
9	NAMITA JANI	HYDRABAD	F	9
10	PRIYA MADNANI	MANGLORE	F	22
11	SHREYA PATKAR	AGRA	F	50
12	DAMINI ROHTE	INDORE	F	61
13	JANAK BHIMANI	DEWAS	M	34
14	NARAY ANAND	UJJAIN	M	51
15	SANJEEV BHASIN	NIMACH	M	40
16	RAJAK KHAN	RATLAM	M	16
17	MILAN DESPANDEY	NAGDA	M	42
18	NAMAN GOUR	SURAT	M	49
19	AJAY JAIN	BARODA	M	11
20	RAJESH BALAI	BARUCH	M	13
21	NITIN NANE	PALI	M	25
22	SUDHIR MANIK	JAMSEDPUR	M	27
23	PRAKASH REDKAR	BHILWARA	M	58
24	AJAY JAIN	BHIWANDI	M	46
25	RASHMI KAPDIYA	PALGHAR	F	24
26	RAMILA BALI	JALGAON	F	30
27	PUSHPA MANI	BHUSAVAL	F	20

```

S504 X
1 16#-1.181288692471365#48.39722589721657#D#0.0957306666666666
2 16#-4.484288692474365#48.39722589721657#D#1.1282306666666662
3 16#-4.484288692474365#48.39722589721657#D#2.1607306666666666
4 16#-4.484288692474365#48.39722589721657#D#3.1932306666666665
5 16#-4.484288692474365#48.39722589721657#D#4.2257306666666664
6 20#-4.470600240001205#40.39571907C49995#D#5.2225730666666664
7 16#-4.484288692474365#48.39722589721657#D#6.2257306666666664
8 20#-4.47868882400512695#48.39571907C49995#D#7.2257306666666663
9 16#-4.484288692474365#48.39722589721657#D#8.2257306666666664
10 20#-4.47868882400512695#48.39571907C49995#D#9.2257306666666665
11 2#-4.4886880634307861#48.39761056713476#D#10.2257306666666657
12 16#-4.484288692474365#48.39722589721657#D#11.2257306666666666
13 2#-4.482443332672119#48.390215879708624#D#12.22573066666666
14 20#-4.47868881400512695#48.39761056713476#D#13.225730666666666
15 2#-4.484288692400512695#48.39761056713476#D#14.2257306666666664
16 16#-4.484288692474365#48.39722589721657#D#15.2257306666666664
17 2#-4.482443332672119#48.390215879708624#D#16.2257306666666664
18 20#-4.4786882400512695#48.39571907C49995#D#17.2257306666666664
19 2#-4.400000634007061#40.39761056713476#D#18.2257306666666664
20 6#-4.494009017944336#48.3952597610596#D#19.225730666666666
21 2#-4.482443332672119#48.390215879708624#D#20.2257306666666653
22 20#-4.47868882400512695#48.39571907C49995#D#21.2257306666666653
23 2#-4.488880634307861#48.39761056713476#D#22.225730666666665
24 6#-4.494009017944336#48.3952597610596#D#23.2257306666666653
25 2#-4.482443332672119#48.390215879708624#D#24.2257306666666657
26 2#-4.4786882400512695#48.39571907C49995#D#25.2257306666666657
27 2#-4.488880634307861#48.39761056713476#D#26.225730666666665
28 6#-4.494009017944336#48.3952597610596#D#27.225730666666665
29 2#-4.482443332672119#48.390215879708624#D#28.2257306666666657
30 2#-4.488880634307861#48.39761056713476#D#29.22573066666666
31 2#-4.482443332672119#48.390215879708624#D#30.22573066666666
32 7#-4.4920249900964#40.39802221737764#D#31.225730666666667
33 7#-4.488880634307861#48.39761056713476#D#32.225730666666664
34 2#-4.482443332672119#48.390215879708624#D#33.225730666666664
35 7#-4.48920249938964#48.398082221737764#D#34.225730666666664
36 2#-4.488880634307861#48.39761056713476#D#35.225730666666667
37 2#-4.482443332672119#48.390215879708624#D#36.225730666666668
38 7#-4.48920249938964#48.398082221737764#D#37.225730666666667

```

Fig. 2. Workflow of Guest Tagging System Design for Surveillance using IoT.

A Guest visits a prominent location. Each guest is provided with the Passive RFID tag which is mapped with guest details. RFID reader operated via a Raspberry Pi at every prominent location of the village reads each tag and sends the tag's identity, current timestamp recorded and location to the database.

B. Importance of the Proposed Solution Proposed solution incorporates the following features:

Crowd Management and Control

Crowd management refers to situations where the evacuations places for the crowd is relatively more and therefore the crowd gets various options to ease itself. Crowd Control refers to scenarios where evacuations places for the crowd is relatively less and therefore the crowd has few or one option(s) to ease itself. Hence, crowd is forced to follow a particular plan.

1. Prevention, Avoidance and Detection of Crowd Disasters (Stampede like phenomenon). Track or Monitor Individual User Movements At the granular level, with the support of high computational power we even can able to track individual person movements as well.
2. High Scalability and Interoperability

The proposed application can be integrated to any RFID System which track its users automatically or needs to swipe its RFID card such as Railway Stations which uses NFC cards, Attendee Tracking in any mass-gathering event etc. The proposed application can also be integrated to location aware mobile system, Wi-Fi membership etc.

IV. METHODOLOGY

The aim of our proposed innovation is to provide a crowd management and Implantation of Access Control System, making the Village more Secured.

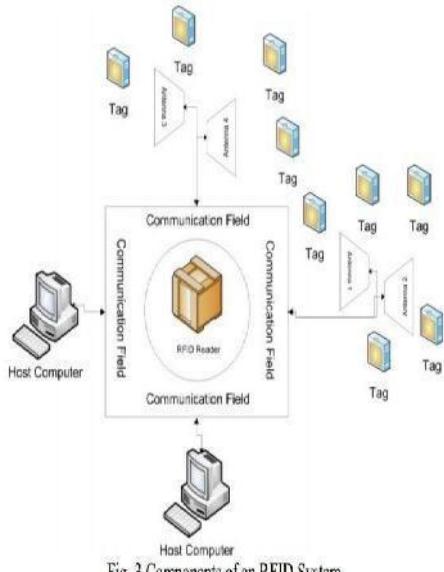


Fig. 3. RFID Reader and tags Communication

The proposed application will have following features:

1. In public places, we need to get the information about number of people in particular time and collect it by various means, like RFID (as part of Smart card based ticket), location aware mobile system, Wi-Fi
2. Multidimensional input data from any RFID system or other means as discussed, about its user, location, current timestamp and RFID tag/card or other means as discussed above is mapped with user details.

At the granular level, with the support of high computational power we even can able to track individual person movements.

- The project maintains two levels of users:-

1. Administrator Level- Owner or Staff at GEV
2. User Level- the Guests/ Visitors and the data-entry level operators

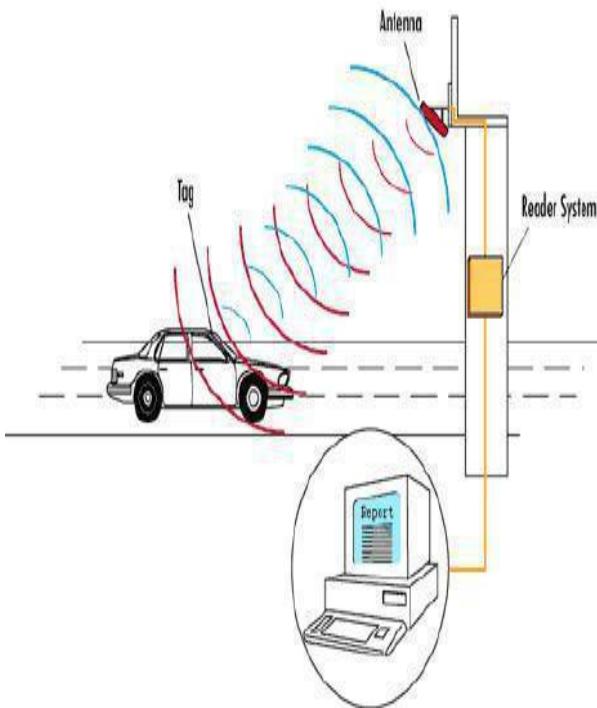
How RFID System Works

Most RFID systems consist of tags that are attached to the objects to be identified. Each tag has its own "read-only" or "rewrite" internal memory depending on the type and application. Typical configuration of this memory is to store product information, such as an object's unique ID manufactured date, etc. The RFID reader generates magnetic fields that enable the RFID system to locate objects (via the tags) that are within its range.

The high-frequency electromagnetic energy and query signal generated by the reader triggers the tags to reply to the query; the query frequency could be up to 50 times per second. As a result communication between the main components of the system i.e. tags and reader is established. As a result large quantities of data are generated. Supply chain industries control this problem by using filters that are routed to the backend information systems. In other words, in order to control this problem, software such as Savant is used. This software acts as a buffer between the Information Technology and RFID reader.

Several protocols manage the communication process between the reader and tag. These protocols (ISO 15693 and ISO 18000-3 for HF or the ISO 18000-6, and EPC for UHF) begin the identification process when the reader is switched on. These protocol works on selected frequency bands (e.g. 860 – 915 MHz for UHF or 13.56 MHz for HF). If the reader is on and the tag arrives in the reader fields, then it automatically wakes-up and decodes the signal and replies to the reader by modulating the reader's field.

All the tags in the reader range may reply at the same time, in this case the reader must detect signal collision (indication of multiple tags). Signal collision is resolved by applying anti-collision algorithm which enables the reader to sort tags and select/handle each tag based on the frequency range (between 50 tags to 200 tags) and the protocol used. In this connection the reader can perform certain operations on the tags such as reading the tag's identifier number and writing data into a tag.



Tags

Tags contain microchips that store the unique identification (ID) of each object. The ID is a serial number stored in the RFID memory. The chip is made up of integrated circuit and embedded in a silicon chip. RFID memory chip can be permanent or changeable depending on the read/write characteristics. Read-only and rewrite circuits are different as read-only tag contain fixed data and cannot be changed without re-program electronically. On the other hand, re-write tags can be programmed through the reader at any time without any limit. RFID tags can be different sizes and shapes depending on the application and the environment at which it will be used. A variety of materials are integrated on these tags. For example, in the case of the credit cards, small plastic pieces are stuck on various objects, and the labels are also embedded in a variety of objects such as documents, cloths, manufacturing materials etc. Figure 4 demonstrates the different sizes and shapes of the RFID tags.

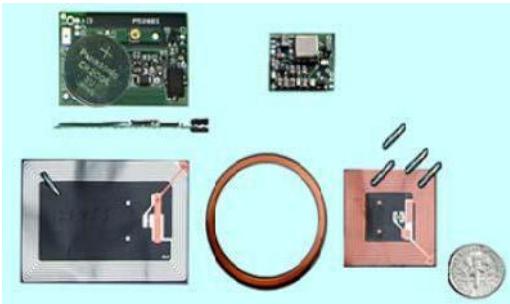


Fig. 4. RFID Tags

V. RESULTS AND DISCUSSION

ID	Position x	Position y	Locations (A, B, C, D, E)	Current Timestamp
1	-4.45	48.39	A	37.094064

SIMULATION OF SMART CITY ENVIRONMENT FROM READERS

S501	
1	#-4.488880634307861#48.39761056713476#A#0.0940640000000001
2	9#-4.485867687225342#48.39865058605447#A#1.094063999999993
3	7#-4.485202499389648#48.398052221737764#A#2.094063999999995
4	2#-4.488880634307861#48.39761056713476#A#3.094063999999998
5	9#-4.485867687225342#48.39865058605447#A#4.094063999999998
6	21#-4.489116668701172#48.39977606202664#A#5.094063999999997
7	7#-4.485202499389648#48.398052221737764#A#6.094063999999997
8	2#-4.488880634307861#48.39761056713476#A#7.094063999999995
9	9#-4.485867687225342#48.39865058605447#A#8.094063999999996
10	21#-4.489116668701172#48.39977606202664#A#9.094063999999992
11	9#-4.485867687225342#48.39865058605447#A#10.094063999999993
12	21#-4.489116668701172#48.39977606202664#A#11.094063999999992
13	5#-4.493322372436523#48.39433365600335#A#12.094063999999993
14	9#-4.485867687225342#48.39865058605447#A#13.094063999999992
15	21#-4.489116668701172#48.39977606202664#A#14.09406399999994
16	37#-4.491391181945801#48.39923469694555#A#15.09406399999996
17	5#-4.493322372436523#48.39433365600335#A#16.09406399999995
18	9#-4.485867687225342#48.39865058605447#A#17.094063999999995
19	21#-4.489116668701172#48.39977606202664#A#18.09406399999996
20	37#-4.491391181945801#48.39923469694555#A#19.09406399999992
21	100#-4.487872123718262#48.40057385269775#A#20.09406399999985
22	11#-4.495133372184131#48.396599021488214#A#21.09406399999995
23	9#-4.485867687225342#48.39865058605447#A#22.09406399999995
24	10#-4.495103359222412#48.39739666198373#A#23.09406399999985
25	21#-4.489116668701172#48.39977606202664#A#24.09406399999999
26	37#-4.491391181945801#48.39923469694555#A#25.09406399999975
27	66#-4.453987560222174#48.39039975693221#A#26.09406399999993
28	92#-4.495879502105713#48.3930940739231#A#27.0940639999998
29	100#-4.487872123718262#48.40057385269775#A#28.09406399999995
30	11#-4.495133372184131#48.396599021488214#A#29.09406399999992
31	89#-4.453279457092265#48.389018917235715#A#30.09406399999992
32	9#-4.485867687225342#48.39865058605447#A#31.09406399999996
33	10#-4.455103359222412#48.3973966198373#A#32.09406399999996
34	21#-4.489116668701172#48.39977606202664#A#33.09406399999996
35	37#-4.491391181945801#48.39923469694555#A#34.09406399999996
36	1#-4.4883012771606445#48.3913842558447#A#35.09406399999996
37	66#-4.453987560222174#48.39039975693221#A#36.094064
38	92#-4.495879502105713#48.3930940739231#A#37.094064

Fig. 5. Log file Generated from READER 1

```

S502
1 5#-4.493322372436523#48.39433365600335#B#0.0940640000000001
2 6#-4.49400901794433648#48.39525976109966#B#1.1573973333333333
3 6#-4.49400901794433648#48.39525976109966#B#2.1573973333333325
4 26#-4.49297904962617#48.38041070655736#B#3.1573973333333316
5 6#-4.49400901794433648#48.39525976109966#B#4.1573973333333311
6 26#-4.49297904962617#48.38041070655736#B#5.1573973333333313
7 29#-4.494051933288574#48.3909282946295#B#6.1573973333333333
8 32#-4.492206573486328#48.39359276009081#B#7.1573973333333325
9 59#-4.49040412902832#48.38958888498521#B#8.157397333333329
10 6#-4.49400901794433648#48.39525976109966#B#9.157397333333325
11 26#-4.49297904962617#48.38041070655736#B#10.157397333333321
12 29#-4.494051933288574#48.3909282946295#B#11.157397333333325
13 32#-4.492206573486328#48.39359276009081#B#12.157397333333325
14 49#-4.4880515853881836#48.3915979862513#B#13.157397333333325
15 58#-4.49039756011963#48.389603134209636#B#14.157397333333325
16 81#-4.493386745452881#48.391256017169496#B#15.157397333333329
17 59#-4.49040412902832#48.38958888498521#B#16.157397333333318
18 26#-4.49297904962617#48.38041070655736#B#17.157397333333333
19 29#-4.494051933288574#48.3909282946295#B#18.157397333333333
20 32#-4.492206573486328#48.39359276009081#B#19.1573973333333325
21 49#-4.488515853881836#48.3915979862513#B#20.1573973333333318
22 43#-4.489803314208984#48.393806481221105#B#21.157397333333318
23 58#-4.490339756011963#48.389603134209636#B#22.157397333333314
24 71#-4.490489959716797#48.389161406396575#B#23.157397333333318
25 81#-4.493386745452881#48.391256017169496#B#24.15739733333332
26 59#-4.49040412902832#48.38958888498521#B#25.1573973333333307
27 26#-4.49297904962617#48.38041070655736#B#26.1573973333333314
28 29#-4.494051933288574#48.3909282946295#B#27.1573973333333314
29 32#-4.492206573486328#48.39359276009081#B#28.157397333333332
30 49#-4.488515853881836#48.3915979862513#B#29.1573973333333325
31 43#-4.489803314208984#48.393806481221105#B#30.157397333333325
32 45#-4.499474613189697#48.391298763430406#B#31.157397333333332
33 58#-4.49039756011963#48.389603134209636#B#32.157397333333336
34 71#-4.490489959716797#48.389161406396575#B#33.157397333333336
35 81#-4.493386745452881#48.391256017169496#B#34.15739733333333
36 59#-4.49040412902832#48.38958888498521#B#35.15739733333333
37 88#-4.493150711059578#48.3875797043953#B#36.157397333333336
38 26#-4.49297904962617#48.38041070655736#B#37.157397333333336

```

```

S503
1 49#-4.4880515853881836#48.3915979862513#C#0.09573066666666666
2 7#-4.4892062499389648#48.398052221737764#C#2.16073066666666666
3 61#-4.4913482666015625#48.38891916693571#C#3.16073066666666665
4 7#-4.489202499389648#48.398052221737764#C#4.16073066666666664
5 61#-4.4913482666015625#48.38891916693571#C#5.16073066666666664
6 71#-4.490498995716797#48.389161406396575#C#6.16073066666666664
7 7#-4.489202499389648#48.398052221737764#C#7.16073066666666662
8 61#-4.4913482666015625#48.38891916693571#C#8.16073066666666662
9 73#-4.48915958404541#48.38903316211975#C#9.160730666666665
10 28#-4.4877648353576664#48.3893038965842#C#10.1607306666666655
11 7#-4.489202499389648#48.398052221737764#C#11.1607306666666655
12 8#-4.494030475616455#48.398057050336145#C#12.1607306666666655
13 56#-4.49588737487793#48.393293549000234#C#13.1607306666666659
14 61#-4.4913482666015625#48.38891916693571#C#14.1607306666666662
15 73#-4.48915958404541#48.38903316211975#C#15.16073066666666662
16 28#-4.4877648353576664#48.3893038965842#C#16.1607306666666662
17 7#-4.489202499389648#48.398052221737764#C#17.1607306666666662
18 8#-4.494030475616455#48.398057050336145#C#18.1607306666666662
19 56#-4.49588737487793#48.393293549000234#C#19.1607306666666662
20 61#-4.4913482666015625#48.38891916693571#C#20.1607306666666665
21 5#-4.493322372436523#48.39433365600335#C#21.160730666666665
22 73#-4.4891391181945801#48.38903316211975#C#22.160730666666648
23 7#-4.486541748046875#48.3885629303397#C#23.160730666666665
24 15#-4.493966102600098#48.39525976109966#C#24.1607306666666655
25 28#-4.4877648353576664#48.3893038965842#C#25.160730666666664
26 3#-4.493945369720459#48.388487919815994#C#26.1607306666666648
27 7#-4.489202499389648#48.398052221737764#C#27.1607306666666648
28 8#-4.494030475616455#48.398057050336145#C#28.1607306666666655
29 56#-4.49588737487793#48.393293549000234#C#29.160730666666666
30 61#-4.4913482666015625#48.38891916693571#C#30.160730666666666
31 5#-4.493322372436523#48.39433365600335#C#31.16073066666666666
32 73#-4.48915958404541#48.38903316211975#C#32.16073066666666666
33 76#-4.486541748046875#48.3885629303397#C#33.16073066666666666
34 15#-4.493966102600098#48.39525976109966#C#34.16073066666666666
35 28#-4.4877648353576664#48.3893038965842#C#35.16073066666666666
36 3#-4.493945369720459#48.388487919815994#C#36.16073066666666666
37 7#-4.489202499389648#48.398052221737764#C#37.16073066666666666
38 8#-4.494030475616455#48.390857050336145#C#38.16073066666666666

```

Fig. 7. Log file Generated from READER 2

Fig. 8. Log file Generated from READER 3

- a. Sample of 501, 502, 503, 504, 505 Readers dataset layout.
 (Table footnote)

S502
1 5#-4.493322372436523#48.39433365600335#B#0.094064000000000001
2 6#-4.494009017944336#48.39525976109966#B#1.1573066666666666
3 6#-4.494009017944336#48.39525976109966#B#2.157397333333325
4 26#-4.492979049682617#48.38041070655736#B#3.157397333333316
5 6#-4.494009017944336#48.39525976109966#B#4.157397333333331
6 26#-4.492979049682617#48.38041070655736#B#5.15739733333333
7 29#-4.494051933288574#48.39092829464295#B#6.15739733333333
8 32#-4.49220657348632#48.3935927609081#B#7.1573973333333285
9 59#-4.49040412902832#48.3895888490521#B#8.157397333333329
10 6#-4.494009017944336#48.39525976109966#B#9.157397333333325
11 26#-4.492979049682617#48.38041070655736#B#10.157397333333321
12 29#-4.494051933288574#48.39092829464295#B#11.157397333333325
13 32#-4.49220657348632#48.3935927609081#B#12.157397333333325
14 49#-4.488515853881836#48.3915979862513#B#13.157397333333325
15 50#-4.490339756011963#48.389603134209636#B#14.157397333333329
16 81#-4.49338674545281#48.391256017169496#B#15.157397333333329
17 59#-4.49040412902832#48.3895888490521#B#16.1573973333333
18 26#-4.492979049682617#48.38041070655736#B#17.1573973333333
19 29#-4.494051933288574#48.39092829464295#B#18.15739733333333
20 32#-4.49220657348632#48.3935927609081#B#19.157397333333325
21 49#-4.488515853881836#48.3915979862513#B#20.157397333333318
22 43#-4.49803314209894#48.393806481221105#B#21.157397333333318
23 58#-4.490339756011963#48.389603134209636#B#22.157397333333314
24 71#-4.490489959716797#48.389161406396575#B#23.157397333333318
25 81#-4.49338674545281#48.391256017169496#B#24.15739733333332
26 59#-4.49040412902832#48.3895888490521#B#25.157397333333307
27 26#-4.492979049682617#48.38041070655736#B#26.157397333333314
28 29#-4.494051933288574#48.39092829464295#B#27.157397333333314
29 32#-4.49220657348632#48.3935927609081#B#28.1573973333332
30 49#-4.488515853881836#48.3915979862513#B#29.157397333333325
31 43#-4.49803314209894#48.393806481221105#B#30.157397333333325
32 45#-4.494974613189697#48.391298763430406#B#31.15739733333332
33 58#-4.490339756011963#48.389603134209636#B#32.15739733333336
34 71#-4.490489959716797#48.389161406396575#B#33.15739733333336
35 81#-4.49338674545281#48.391256017169496#B#34.1573973333333
36 59#-4.49040412902832#48.3895888490521#B#35.1573973333333
37 88#-4.49315071105957#48.38757970439532#B#36.15739733333336
38 26#-4.492979049682617#48.38041070655736#B#37.15739733333336

S503
1 49#-4.488515853881836#48.39159798625123#C#0.09573066666666667
2 7#-4.489202499389648#48.389052217377764#C#2.1607306666666666
3 61#-4.4913482666015625#48.38891916693571#C#3.1607306666666655
4 7#-4.489202499389648#48.389052217377764#C#4.1607306666666664
5 61#-4.4913482666015625#48.38891916693571#C#5.1607306666666664
6 71#-4.490489959716797#48.389161406396575#C#6.1607306666666664
7 7#-4.489202499389648#48.389052217377764#C#7.1607306666666662
8 61#-4.4913482666015625#48.38891916693571#C#8.1607306666666662
9 73#-4.48915958404541#48.38903316211975#C#9.160730666666659
10 28#-4.487764835357666#48.3893038996584#C#10.1607306666666655
11 7#-4.489202499389648#48.389052217377764#C#11.160730666666659
12 8#-4.494030475616455#48.389057050336145#C#12.160730666666659
13 56#-4.489588737487793#48.393293549000234#C#13.1607306666666659
14 61#-4.4913482666015625#48.38891916693571#C#14.160730666666662
15 73#-4.48915958404541#48.38903316211975#C#15.160730666666662
16 28#-4.487764835357666#48.3893038996584#C#16.160730666666662
17 7#-4.489202499389648#48.389052217377764#C#17.160730666666662
18 8#-4.494030475616455#48.389057050336145#C#18.160730666666662
19 56#-4.489588737487793#48.393293549000234#C#19.16073066666666
20 61#-4.4913482666015625#48.38891916693571#C#20.160730666666665
21 58#-4.493322372436523#48.39433365600335#C#21.160730666666665
22 73#-4.48915958404541#48.38903316211975#C#22.160730666666648
23 76#-4.486541748046875#48.3885629303357#C#23.160730666666665
24 15#-4.493966102600098#48.39525976109966#C#24.1607306666666655
25 28#-4.487764835357666#48.3893038996584#C#25.160730666666664
26 38#-4.483945369720459#48.388847919815994#C#26.160730666666648
27 7#-4.489202499389648#48.389052217377764#C#27.160730666666648
28 8#-4.494030475616455#48.389057050336145#C#28.160730666666655
29 56#-4.489588737487793#48.393293549000234#C#29.16073066666666
30 61#-4.4913482666015625#48.38891916693571#C#30.16073066666666
31 58#-4.493322372436523#48.39433365600335#C#31.1607306666666666
32 73#-4.48915958404541#48.38903316211975#C#32.1607306666666666
33 76#-4.486541748046875#48.3885629303357#C#33.1607306666666666
34 15#-4.493966102600098#48.39525976109966#C#34.1607306666666666
35 28#-4.487764835357666#48.3893038996584#C#35.1607306666666666
36 3#-4.483945369720459#48.388847919815994#C#36.1607306666666666
37 7#-4.489202499389648#48.389052217377764#C#37.1607306666666666
38 8#-4.494030475616455#48.389057050336145#C#38.1607306666666666

Fig. 9. Log file Generated from READER 4

Fig. 10. Log file Generated from READER 5

VII. CONCLUSION AND FUTURE WORK

Crowd Monitoring and Management in large scale gathering are complex and multifactorial phenomenon which are not so easy to solve. We need automated systems as companions to help solve these issues in the most optimized manner. And save the great havoc- loss or damage to human life.

- The proposed solution tries to achieve both by providing assistance to the people serving and managing the crowd with accurate statistical knowledge and most optimized decision.
- In terms of **performance** neural networks provide accurate results with proper initialization of thresholds on the nodes, effective and accurate study of conditions constrained on the neural network.
- In terms of **security**, this method collects all guest entries who visit the village on remote server, so any theft or stranger will be detected.
- Manual entry work will be **automated**, so time, cost & man power will be reduced. All guest's entries will be forwarded online through internet. Previously entry work was in an offline mode where simple guest ids were given and noted in a register.
- Analysis** will be performed on data acquired so as to manage the flow of guests in different zones effectively, **reduce the waiting time and serve the visitors in a more convenient manner**.

Thus, the proposed solution will be implemented to maximize the user experience during crowd management at large scale events and prevent stampede like phenomenon.

The proposed solution is more inclined towards the social and functional requirements of **crowd management strategic approaches- preventing, detecting and avoiding stampede like phenomenon** which creates a great havoc – loss of human life in large scale gathering events / areas. Also, it would be an added feature to **maximize the user experience during crowd management at large scale events**

The system can be scaled to various Crowd Management phenomenon like Stampede which can occur in any densely crowded areas such as **Mumbai Local Stations**,

Pilgrimages like Hajj, Kumbh Mela, Festivals like Ganpati, Durga Puja, Mass Gatherings, Live Shows etc.

We know it's not that easy to manage a crowd of say 1 million, but when it comes to saving lives, some measures need to be taken.

VIII. ACKNOWLEDGMENT

We would like to express our heartfelt gratitude towards our mentor Dr. Deven Shah for his unconditional guidance and support throughout the development of the proposed solution and for always being there for us. It has been a pleasure experience working with him.

Special thanks to Govardhan Eco-Village for providing us the problem definition and allowing us to visit your place for requirement gathering and pursue our innovative idea. A big thank you to all our friends, family, teachers and well-wishers notably Mrs. Neha Kapadia, Mr. Aaditya Desai, Shubham, Atul and Rahul for providing their invaluable feedback and warm support always.

REFERENCES

Following is the list of references text-based, journal based or web – based used to propose the solution:

Textbook References:

- [1] RFID Handbook: Fundamentals And Applications In Contactless Smart Cards, Radio Frequency Identification And Near-Field Communication, Third Edition WILEY Publications
- [2] Designing the Internet of Things: Adrian McEwen, Hakim Cassimally WILEY Publications

Journal References:

- [3] Identification and tracking of persons using RFID-tagged items: United States Patent Application 20020165758 Kind Code A1 Hind, John R. ; et al. November 7, 2002
- [4] <http://www.spychips.com/documents/ATT00075.pdf>
- [5] [4]. The Use of RFID for Human Identification: A DRAFT REPORT from DHS Emerging Applications and Technology Subcommittee to the Full Data Privacy and Integrity Advisory Committee Version 1.0
- [6] https://www.dhs.gov/xlibrary/assets/privacy/privacy_adv_com_rpt_rfid_draft.pdf
- [7] <http://www.ijeee.org/papers/306-E794.pdf>

Web References:

- [8] <http://www.corerfid.com/Files/White%20Papers/033%20Health%20Fact%20Sheet.pdf>
- [9] <https://www.ecovillage.org.in>
- [10] https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
- [11] <https://www.instructables.com/RFID-RC522>
- [12] <https://electrosome.com/em-18-rfid-reader-raspberry-pi/>
- [13] <https://raspberrypi.org/forum>

To Analyze Student's Learning Experience in social media

Swathi Yadav

*Thakur College of Engineering & Technology
Kandivali East, Mumbai.*

Abstract-Student's posts and remarks via social media (informal organization destinations, for example, twitter) centers into their educational experiences, for example, their issues, comments which focus on students learning process. Some important data about student taking in encounters can be construed from the information assembled from such situations. It is difficult to dissect that data. Since the online networking information continues expanding in measure it demands automation in data analysis. At that point likewise the data induced from those information needs human interpretation since it is the impression of students' emergency. A work process that joins quantitative analysis and large scale data mining systems is created. We concentrated on student's posts on comprehend issues and issues in their instructive encounters. Substantial work stack, absence of awareness of social exercises, and restlessness are a few issues that student's look as they experience circular activities. In light of these outcomes, we began to actualize a classification algorithm to group posts mirroring student's issues.

Keywords

social mining, text classification, naïve Bayes classifier.

I.INTRODUCTION

Online networking destinations for example, twitter, Facebook gatherings give great venues for students to impart their experiences, vent feeling and stress, and look for social support. Around Different Online networking sites, scholars talk about activities in casual manner . The plenitude of Online networking information gives chances understand all the students' experiences, as well as raises methodological challenges in making sense of Online networking information to instructive purposes. Just envision those sheer information volumes, the assorted qualities of web slang, the flightiness of area Furthermore timing from claiming student's posts on the web, and additionally those complexities about students' experiences. Pure manual examination can't manage those ever-growing scale of data, at the same time pure automatic algorithms typically can't catch in-depth importance inside the information.

Traditionally, educational scientists have been utilizing techniques like surveys, interviews, centering groups, Furthermore classroom exercises to gather information identified by students'. These method consume lot of time, Subsequently can't be duplicated alternately repeated with high frequency. Those scale about such investigations may be also limited. To addition, when provoked regarding their experiences, scholars necessity to reflect around the thing that they were thinking in the past, which might have turned obscured overtime.

Kamal Shah

*Thakur College of Engineering & Technology
Kandivali East, Mumbai.*

II.TEXT MINING

Text mining can avail an organization derive potentially valuable business insights from text-predicated content such as word documents, email and postings on gregarious media streams like Facebook, Twitter and LinkedIn. Mining unstructured data with natural language processing (NLP), statistical modeling and machine learning techniques can be arduous, however, because natural language text is often inconsistently erratic [1]. It contains ambiguities caused by inconsistently erratic syntax and semantics, including slang, language categorical to vertical industries and age groups, double entendres and mordancy.

Text analytics software can avail by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. With an iterative approach, an organization can prosperously use text analytics to gain insight into content-categorical values such as sentiment, emotion, intensity and pertinence. Because text analytics technology is still considered to be an emerging technology, however, results and depth of analysis can vary wildly from vendor to vendor.

III.NAÏVE BAYES

Naïve bayes expects a specific probabilistic generative model for text. Those model is a specialization of the mixture model, and hence also makes the two presumptions examined there. Additionally, naïve bayes makes expression independence presumptions that permits the generative model should be portrayed with a extraordinarily diminished amount of parameters. Whatever remains of this subsection portrays the generative model All the more formally, giving an exact determination of the model parameters, Furthermore inferring those likelihood that An specific record may be produced provided for its class name.

First let us introduce some notations to describe text.Consider A document, d_i , is an ordered list of word events, $w_{di,1}, w_{di,2}, \dots$.then We can write $w_{di,k}$ for the word wt in position k of document d_i , where wt is a word in the vocabulary $V = w_1, w_2, \dots, w|V|$.

When a document is to be created by a specific mixture component, c_j , a document length, $|d_i|$, is selected independently of the component. Then, the chosen mixture component creates a word sequence of the specified length. We furthermore assume it creates each word independently of the length.

Hence, the likelihood of a archive provided for a mixture part As far as its constituent Characteristics camwood be communicated Similarly as those record length and the

expressions in the archive. Those likelihood of a expressions occasion must make molded once every last one of expressions that precede it.

$$P(d_j|c_j; \theta) = P((w_{d_i,1}, \dots, w_{d_i,|d_i|})|c_j; \theta) = P(|d_i|)$$

$$\prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j, \theta, w_{d_i,q}, q < k) \quad (1)$$

Following we make those standard credulous bayes assumption: that those expressions of a record are produced freely for context, that is, freely of the other expressions in the same report provided for those population mark. We further expect that those likelihood of a expression is autonomous from claiming its position inside the document; thus, to example, those likelihood about seeing the statement "homework" in the Initially position of a archive is the same Concerning illustration seeing it to whatever viable position. We could express these presumptions as:.

$$P(w_{d_i,k} | c_j; \theta ; w_{d_i,q}, q < k) - P(w_{d_i,k} | c_j; \theta) \quad (2)$$

Joining together these A two equations provides for those naive bayes statement for the likelihood of a report provided for its class.

$$P(w_{d_i,k} | c_j; \theta) - P(|d_i|) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j; \theta) \quad (3)$$

Thus the parameters of an particular mixture component are a multinomial distribution over words, i.e. the gathered of word probabilities, each written $\theta_{wt|c_j}$, such that $\theta_{wt|c_j} = P(wt|c_j; \theta)$, where $t = \{1, \dots, |V|\}$ and $\sum P(wt|c_j; \theta) = 1$. because we assume that for all classes, document length is distributed same, there is no need to parameterize for classification.

IV.RELATED WORKS

The authors, Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, Tom Mitchell, Says that those exactness about figured out how quick classifiers might be enhanced Toward augmenting An little number from claiming marked preparing documents with an extensive pool of unlabeled documents. This may be paramount Since in large portions quick arrangement issues acquiring preparing labels may be expensive, same time expansive amounts about unlabeled documents would promptly accessible. He introduces an calculation for Taking in starting with marked Furthermore unlabeled documents In light of the blending for Expectation-Maximization (EM) Furthermore An naive bayes classifier. The algorithm to start with trains a classifier utilizing the accessible marked documents, and probabilistically labels those unlabeled documents. It At that point trains another classifier utilizing those labels for every last one of documents, Furthermore iterates should joining. This fundamental em system meets expectations great At the information fit in with those generative presumptions of the model [2]. Nonetheless morals these presumptions are frequently all the disregarded done act Also poor execution

camwood result. These writers introduce two extensions of the calculation that enhance order correctness under these conditions: An weighting element will regulate the commitment of the unlabeled data, Furthermore the utilization about numerous mixture parts for every population. Test results, acquired utilizing quick starting with three separate real-world tasks, demonstrate that the utilization of unlabeled information lessens arrangement lapse Toward up to 30%.

Bo Pang and Lillian Lee Concerning illustration the creators inspected those connection the middle of subjectivity identification Also polarity classification, demonstrating that subjectivity identification could layer reviews under substantially shorter extracts that at present hold polarity majority of the data toward a level tantamount to that of the full survey. On fact, to those naive bayes polarity classifier, those subjectivity extracts would demonstrated to make more viable in-put over those starting document, which prescribes that they are not main shorter, as well as "cleaner" representations of the planned polarity. They need indicated that utilizing those minimum-cut skeleton brings about the improvement from claiming productive calculations for assumption Investigation. Using relevant majority of the data through this frame-work might prompt statistically huge change over polarity-classification precision [3]. Currently these creators finish up expressing the naïve bayes may be relatively best.

Tina R. Patil, Mrs. S. S. Sherekar Sets out to make similar assessment of classifiers naïve bayes and J48 in the connection of bank dataset will expand valid certain rate Furthermore minimize false sure rate from claiming defaulters instead of accomplishing main higher arrangement correctness utilizing WEKA apparatus. Those investigations outcomes demonstrated in this paper need aid around order accuracy, affectability Furthermore specificity. Those brings about those paper with respect to this dataset likewise demonstrate that those effectiveness and correctness from claiming j48 will be superior to that of Naïve bayes. Yet our issue need the multi level classifier. The framework need various information qualities. Thereabouts we utilize naïve bayes with handle the ordinary developing dataset with those numerous quality values.

V.PROPOSED METHOD

Traditionally there is no such researches and implementation to analyze the posted data for the educational purpose in the social media by the students and to merge both analysis and mining techniques. The survey that has been proposed is to monitor the student activity by permitting the necessary privacy. The issues and problems of the students can be analyzed out through their posts and conversations in the social media. If the issues are negative and serious then the notice has to be sent to the concerned staff and their parents. The student's problem and the feedback can be considered in the higher level decision making The proposed method has seven steps to get the appropriate result are listed below,

1. Collecting Data
2. Sampling
3. Analysing Data Qualitatively
4. Result
5. Training And Evaluation of model
6. Model Adaption
7. Large Scale Data Analysis Result

In the Fig. 1. Data is gathered as a data set (i.e.,) the posts from the social media is collected as a sample and analyzed. The issues and problems of the students can be analyzed out through their post. If the issues are negative and serious then the message has to be sent to the concerned staff and their parents. The student's issues , problems and the feedback can be considered in the higher level decision making.The proposed method has following modules,

Computational Attribute Management

Those quality era module completely works on the director end Also it gathers the quality information starting with admin for administering the dataset to future motivation. Once those attributes are produced it will make simple to dissect those considerations of the student dependent upon their provided for input starting with their portal. To every last one of quality management is A dataset oversaw economy procedure which holds those data in regards the keywords and the particular qualities for mining the students performance and learning experience.

Qualitative Attribute Analysis

Once the attributes are produced the student effective mining portal is in other hand to gather the information from the student's end. This port can be analysed only by the administrator to know the thoughts of the students. Every time when the student tweet with some data that all data posted by them will be compared with the created attribute set and then the mining process takes place for analyzing the qualitative summary of the student [4]. Specifically the qualitative analysis module helps the administrator to analyze the qualitative summary of the student performance.

Public Web Conversation Port

The hypothetical establishment for the worth of casual information on the web can be drawn from Goffman's hypothesis of social execution. Despite formed to clarify face to face interactions, Goffman's hypothesis of social execution may be generally used to demonstrate mediated collaborations on the web today [5]. A standout amongst the practically essential viewpoints from claiming this hypothesis will be those thought from claiming front-stage and back-stage from claiming people's social performance. If a social setting is front-stage or back-stage may be a relative matter. To students, compared for formal classroom settings, Online networking is a relative casual and unwinding back-stage. When student's post content with respect to Online networking sites, they normally post the things that they feel during that moment.

Mining Student Conversation

From the different fields about A large number existing meets expectations specialists bring broke down the mining content with produce particular information to their particular liable domains. To example, Gaffney analyzes tweets for hash tag iran race utilizing histograms, client networks, Furthermore frequencies from claiming highest point keywords will quantify on the web activism. Comparative investigations bring been led over other fields including healthcare, promoting Also athletics, only on sake a couple. Investigation systems utilized in these investigations Typically incorporate

qualitative substance analysis, semantic analysis, organize analysis, and a few oversimplified systems for example, such that saying clouds and histograms. In this module, An order model may be manufactured In light of inductive substance dissection. This model might have been that point connected Also approved for a brand new dataset. Therefore, not main the insights picked up from particular case dataset need aid emphasized, as well as those provision of the order algorithm will other datasets for identifying person issues. The human exert will be In this way increased for extensive scale information investigation.

Text Preprocessing

Many social mining users may use some special symbols to convey their messages. For example, # is used to show a hashtag, is used to show an account of the user , and RT is used to show a re-tweet. Social Mining users may also repeat letters in words so that to emphasize the words, for example, "huungryyy", "soooo muuuchh", and "Monnnndayyy". Common stopwords such as "a, an, and, of, he, she, it, the", nonletter symbols, and punctuation also bring noise to the text. Therefore the pre-processed texts are needed completely for analysing the student mining data and through this module it would be simple to attain.

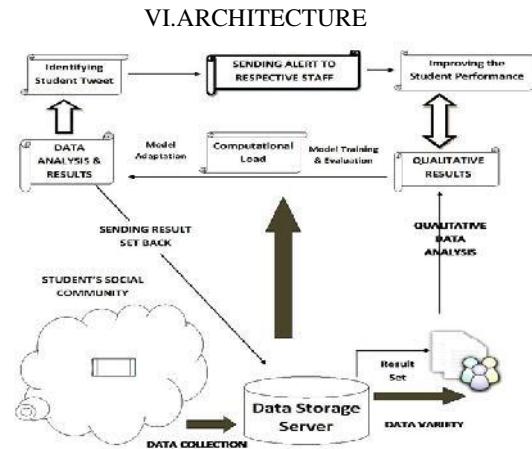


Fig 1 :Architecture

The above Fig 1 , shows the architecture of the prposed method . The architecture shows the overall flow of the proposed method.The flow starts with gathering of the students tweet and identifying those tweets which shows the student's learning experiences.After gathering those tweets the emotions of the student is analyzed by Naïve bayes algorithm. According to the analysis the result which is generated is send as an alert message to the respective teacher. Then the teacher will try to improve the students performance.

VII.CONCLUSION

This paper provides a flow for analyzingthe contents posted by the student in social media for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user generated textual content. And also it can give essential facts to the educational administrators, practitioners and other relevant decision makers to gain further insights of

engineering student's college experiences. In this project, the community like structure is planned between students and staff. Manually the admin has to gather the data from the student's post and then the alert message has to be send to the staff, if the posts are negative.

REFERENCES

- [1] Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, " Mining social media data for understanding students learning experience" , IEEE transaction on Learning Technologies, vol.7, no.3, Pp.16-22 July - September 2014
- [2] Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, Tom Mitchell, "Text Classification From Labeled And Unlabeled Documents Using EM", Machine Learning, 39, Kluwer Academic Publishers. Printed In The Netherlands, Pp. 103–134, 2000.
- [3] Bo Pang And Lillian Lee,"A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based On Minimum Cuts", Morgan & Clay Pool Publishers, Pp. 54-58, 2008.
- [4] Tina R. Patil, Mrs. S. S. Sherekar,"Performance Analysis Of Naive Bayes And J48 Classification Algorithm For Data Classification", J.sci.Education, Vol.86, No.1, Pp.7-15, 2000.
- [5] Swathi Yadav, Shwetha Yadav,"Text Mining of Voot Application Reviews on Google Play Store ",IRJET, Vol.05, Issue.1,Pp.56-72,2018.

To Study and Implement IEEE 802.11g Security Aspects Using Index Policy Method

Yamisha Savalia

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Jasmina Ambani

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Ashwini Nakil

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Rajesh Bansode

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Abstract—Wireless local area networks (WLANs) are beginning to play a much larger role in corporate network environments. The 802.11g specification is a standard for wireless local area networks (WLANs) that offers transmission over relatively short distances at up to 54 megabits per second (Mbps), compared with the 11 Mbps theoretical maximum with the earlier 802.11b standard. In this paper, focus is on WLAN security technologies designed to improve 802.11 standard by applying security policies. Policy indexing model are implemented using mesh topology (five nodes).Total packet transmission time for authentication and encryption of 802.1X model is evaluated with fixed packet size. The evaluation is done based on policy index in model implementation .The total number of iteration in complete file transmission is fixed. The use of different security configuration will provide both the necessary flexibility to network operators and high level of confidence to end users.

Keywords: 802.1X, RADIUS, security policy index, VPN, WEP

I.INTRODUCTION

The 802.11 wireless networks are fast becoming the preferred choice for LAN environments. Given their limited bandwidth (54 Mbps in 802.11g)[1] and the need for security in wireless standards, it is necessary to understand the relative overhead of different security protocols .The MAC sub-layer provides reliable data transmission for the IEEE 802.11 standard similar to a wired network. The IEEE 802.11g WLAN standard can be thought of as an intersection between the 802.11b and 802.11a standards. Like 802.11b, 802.11g operates in the same 2.4-GHz[2]portion of the radio frequency spectrum that allows for license-free operation on a nearly worldwide basis.An important mandatory requirement of 802.11g is full backward compatibility with 802.11b, which both provides investment protection for the installed base of 802.11b clients and extracts a substantial performance penalty when operating in this mode. Firewall hardware and software can be configured to allow safe Internet traffic to enter a network and block unsafe Internet traffic. It uses a series of rules to intercept and analyze data to determine whether the traffic is safe to enter either wired or wireless networks. The purpose of a firewall is to prevent unsafe traffic from entering to the network .If there is a small group of users, the network can be limited to assigning a certain number of DHCP addresses equal to the number of possible users. RADIUS is a protocol for authentication, authorization and accounting of remote access connections. The user inputs his name and password and submits it to the RADIUS server. [3] The RADIUS server determines if the user is authorized to use the network. Communication between the user and the RADIUS server is encrypted; however, the RADIUS protocol does not provide data encryption. RADIUS is often used with a VPN.

The paper is divided into ten sections were section II discusses the overview , section III describes details of IEEE 802.11 standards ,section IV describes the motivation in the field of wireless technology , whereas section V discusses the work carried out in the related fields of wireless security as literature review ,section VI details the problem definition ,section VII describes feasibility study ,section VIII describes the various techniques and methodologies ,section IX depicts the experimental results .The conclusion based on the results achieved is stated in section IX and X followed by acknowledgement and references in other sections.

II.OVERVIEW

The project is to provide organizations with guidance for establishing secure wireless networks. Wireless carriers and network access providers must meet consumer demands for: easier access , better Quality of Service, Better security ,More network management tools ,Flexible pricing. The bit rate per user must be increased to support higher-end application, wireless streaming media, Video conferencing. IT should also fulfill the following challenges: Security, Multicast support, Location management support, Scalability. The parameters that are measured here are Encryption time, Packet transmission time, Computational complexity, Space complexity.

III. DETAILS OF IEEE 802.11 VARIOUS STANDARDS

The IEEE802.11 has released multiple set of standards for various operating frequency, and ranges specification. The first release was IEEE802.11 original standard release that was defined 1997 and clarified 1999. Some of these old standards are now obsolete, and some are still active. One would be interested in investigating in details the available standards to determine a suitable standard for the intended application of the WLAN network.

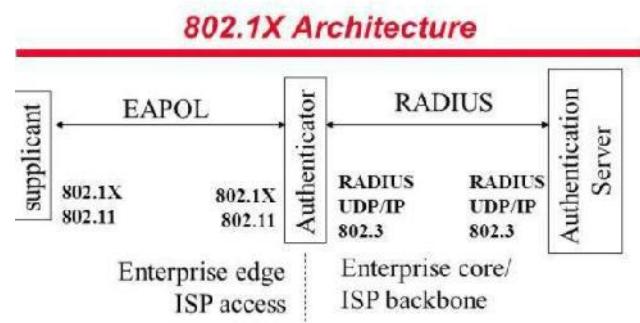


Fig.1 802.1X Architecture

Authenticator acts as a bridge.802.1x is a framework, not a complete specification in and of itself. The actual authentication mechanism is implemented by the authentication server as shown in fig.1.

1. IEEE 802.11a- The IEEE802.11a standard was released on September 1999. Networks using 802.11a operate at radio frequency of 5GHz or 3.7GHz and a bandwidth of 20MHz. The specification uses a modulation scheme known as orthogonal frequency-division multiplexing (OFDM) that is especially well suited to use in office settings. In 802.11a, data speeds as high as 54 Mbps are possible. This standard employ the single input, single output (SISO) antenna technologies, and the indoor/outdoor ranges from 35m to 125m for 5GHz operating frequency. The outdoor range goes to 5Km for operating frequency of 3.7G. The IEEE802.11a is less prone to interference compared to with 802.11b due to the high operating frequency of 5GHz.

2. IEEE 802.11b- IEEE 802.11b standard was released on September 1999 as well. This standard provides 11 Mbps transmission (with a fallback to 5.5, 2 and 1 Mbps) in the 2.4GHz operating frequency and bandwidth of 22MHz. The 802.11b uses only DSSS (Direct Sequence Spread Spectrum) modulation technique. This standard also employs the SISO antenna technology as in the IEEE802.11a standard. The IEEE802.11b standard was ratified on 1999 from the original IEEE802.11 standard which allowed wireless functionality comparable to Ethernet. The IEEE802.11b standard is prone to higher interference due to the fact that the 2.4GHz frequency range is becoming crowded with carriers, hence increased interference risk. The indoor and outdoor ranges for this standard is 35m to 140m[2].

3. IEEE 802.11g -The standard 802.11g was ratified in 2003 as an IEEE standard for Wi-Fi wireless networking and it supports maximum network bandwidth of 54 Mbps compared to 11 Mbps for 802.11b. This standard operates at 2.4GHz frequency and bandwidth of 20MHz. This standard uses the OFDM or DSSS modulation schemes. This standard employ the SISO antenna technologies, and its indoor/outdoor range are from 38m to 140m respectively.

4. IEEE 802.11n -The 802.11n standard was ratified in 2009 and it utilizes multiple wireless antennas in tandem to transmit and receive data[3-4]. The IEEE802.11n standard employs OFDM modulation technique. The antenna technology used with the IEEE802.11n standard is known as Multiple Input, Multiple Output (MIMO). This technology refers to the ability of 802.11n and similar technologies to coordinate multiple simultaneous radio signals. The MIMO increases both the range and throughput of a wireless network. An additional technique employed by 802.11n involves increasing the channel bandwidth from 20MHz to 40MHz. The 802.11n standard support maximum theoretical network bandwidth up to 300 Mbps. The IEEE802.11n indoor/outdoor ranges are 75m, and 250m respectively.

5. IEEE 802.11ac- IEEE 802.11ac is the fifth generation in Wi-Fi networking standards released December 2013[5-6]. This standard operating frequency is 5GHz, and bandwidth of 20, 40, 80, 160MHz sectors. The stream rates ranges for these bandwidth sectors are 7.2 - 96.3Mbps for 20MHz, and 15 – 200Mbps for 40MHz, 32.5 - 433.3Mbps for 80MHz, and 65 -

866. 7Mbps for 160MHz. This standard exhibits better performance, and better coverage compared to IEEE

802. 11a,b,g and n standards. The 802.11ac standard uses a wider channel and an improved modulation scheme that also supports more clients. The IEEE 802.11ac standard utilizes a modulation

technique known as multi-user MIMO. This technique allows a set of users or wireless terminals, each with one or more antennas, to communicate with each other. The indoor range is 35m, and there is no recorded max for outdoor range.

IV.MOTIVATION

The convenience of wireless LANs bring benefit to lives. They keep people connected beyond cubicle or office. People can access the Internet as easily from a coffee shop or our couches at home as while sitting at desks. An entry-level wireless Access Point (AP) costs less than \$75, a price that makes these devices one of the best-selling computer peripherals since the iPod. Wi-Fi-enabled laptops can connect to your network without any wires. This convenience brings with it a huge risk to your network and data, because when you accept the configuration defaults of many of these inexpensive wireless APs, you run the risk of others connecting to your network or snooping on your data just as easily. Also the 802.11g standard devices are compatible with 802.11a and 802.11b devices. Hence, to provide more security to 802.11g devices and to take advantages of its low hardware cost we have undertaken the project.

V. LITERATURE SURVEY

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they can move forward.

1.The solution found is Authentication using VPN is better and Encryption using 802.11 is better.The key findings are File Size=10KB,Packet transmission time 802.1x =17.37 sec VPN=21.99sec,ComputationalComplexity

802.1x=41.16sec,VPN=78.66sec,Space Complexity

802.1x=7.85 KB VPN=8.09KB and the Research gap found is to use 802.1x for encryption and VPN for authentication.[1]

2.The solution found is When more clients were added to the network the author found that overall throughput decreased by roughly 50% with additional client and 67% with two additional clients.The key findings are Transmission rates lowered to 500kbps,Bandwidth=12mbps,Packet sizes= 100,500,1000,1500 bytes and the research gap found is to Implement for 5 nodes.[2]

3.The solution found is the flaws lead to a number of practical attacks that demonstrate WEP fails to achieve its security goals.The key findings are the use of CRC-32 and RC4 can be justified by their speed and ease of implementation,they have discovered several security flaws in the protocol ,stemming from misapplication of cryptographic primitives and the research gap is to Use 802.1x for encryption and VPN for authentication.[3]

4.The solution found is WEP has minor impact on FTP Throughput decreases by 7.5% and throughput for HTTP decreases by 50%.The key findings are the combined effects of encryption and authentication for which response time increases by 268% and throughput decreases by 73% and the research gap is to use 802.1x for encryption and VPN for authentication.[4]

5.The solution found is Open VPN overcomes the weaknesses of WEP and there is an increase in performance of 802.11g with the

use of compression. The key findings Throughput Max decrease 15.84%(UDP), Throughput Max decrease 16.91% (TCP),

Latency increases for both TCP and UDP traffic as we increase the frame size, Frame loss percentage increases when we increase the frame size (up to 90%-92%).[5]

6. The solution found is EAP-TLS provides stronger authentication than EAP-MD5, They found that using small data amounts resulted in no visible differences between encrypted and unencrypted stream throughput. The key findings are in terms of Authentication Time for various Policies IPSEC=1.432(max) 1.405(min), 802.1x EAP(MD5)

without	IPSEC=1.749(Max)	0.427(min), 802.1x
EAP(MD5)with	IPSEC=1.722(min)	1.749(max), 802.1x
EAP(TLS)without	IPSEC=1.822(min)	3.144(max), 802.1x
EAP(TLS)with	IPSEC=3.117(min)	3.144(max) and the research gap is to use EAP-TLS For authentication[6]

VI. PROBLEM DEFINITION

The experiment is simulated with 5 nodes. Initially a Network controller (server) is created with five nodes. Nodes are connected in a mesh network with each other. The further work was done as follows:

- Phase I Selecting a source node and text file to transmit. In this work file size is limited to 12kB.
- Phase II Selecting the Encryption method from the options for encrypting the file.
- Phase III On selecting the desired encryption method, time to encrypt and key length is calculated for that specific file.
- Phase IV After Encryption, authentication method was selected for concerned security level
- Phase V The decryption key was encrypted again using Diffie-Hellman protocol key exchange protocol for secure transfer of file from source to destination.
- Phase VI Random path was generated in the simulation for every file transfer.
- Phase VII On reaching the destination safely computational complexity, space complexity and packet transmission is calculated for security level 2, 3, 4, 5 and 6.

VII .FEASIBILITY STUDY

1. Technical Feasibility – Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

2. Economic Feasibility - As part of this, the costs and benefits associated with the proposed system compared and the project is

economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

3. Legal Feasibility- Legal issues can affect a system's acceptance by users, its performance, or the decisions on whether to use it in the first place—so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any security system.

4. Operational Feasibility - It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

5. Social Feasibility-The acceptability of a security system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness.

VIII. METHODOLOGY

The Methodology is used to compare security index policies of 802.1X model with VPN model for response time, computational complexity, space complexity, key length and time taken to encrypt in simulator. Although there are a number of combinations that could be chosen for security configurations utilized for the testing of response time, computational complexity, space complexity, key length

and time taken to encrypt. This research focuses on those most likely to be present in a corporate network environment. Proposed system architecture is shown in the fig.2 below. The authentication protocols that are examined are MAC address, PPTP, CHAP, HMAC-MD5. However, all levels of encryption, RSA, DES, WEP with 40 bit key are included at some point in the trials. Below is an overview of the security combinations selected for study:

- i. Security Level 1 – This entails open association with no encryption on the data flow. This is the base linesecurity scheme used as the starting point for all data comparisons with encryption and authentication.
- ii. Security Level 2 – For 802.1X model encryption used is RSA and authentication is MAC address authentication. For VPN model RSA encryption and PPTP authentication were implemented.
- iii. Security Level 3 – Open association with a 40-bit WEP key for encryption and HMAC-MD5authentication for 802.1X model. DES encryption and CHAP (handshake procedure each time theclient re-associated with the access point) authentication for VPN model.

IX. RESULTS

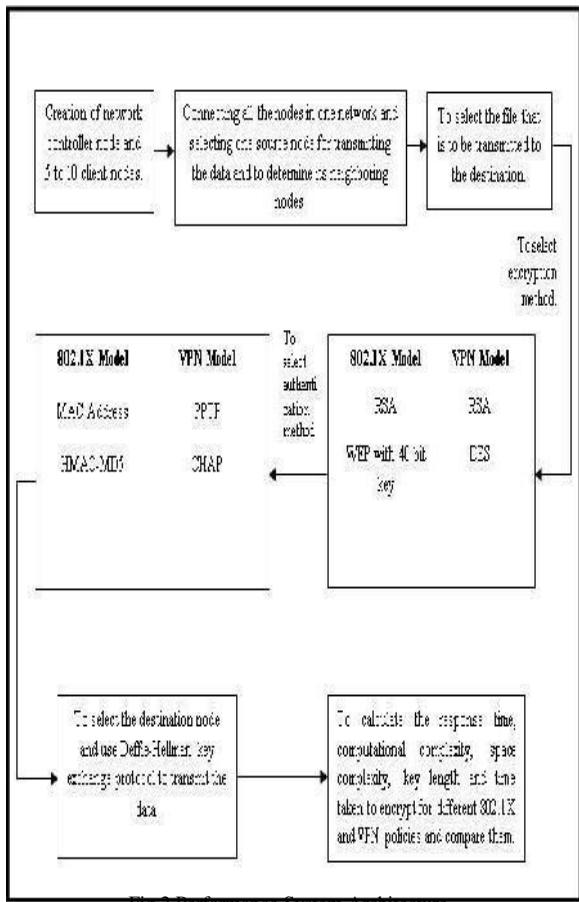


Fig.2 Performance System Architecture

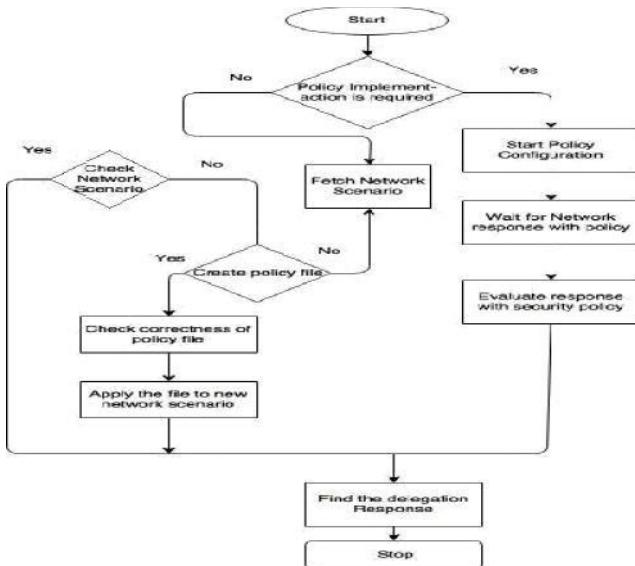


Fig.3 Flowchart for policy selection

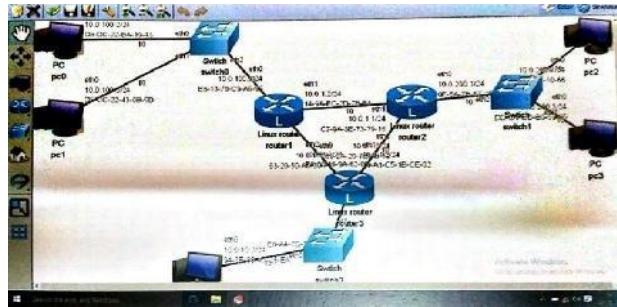


Fig.4 Mesh Network

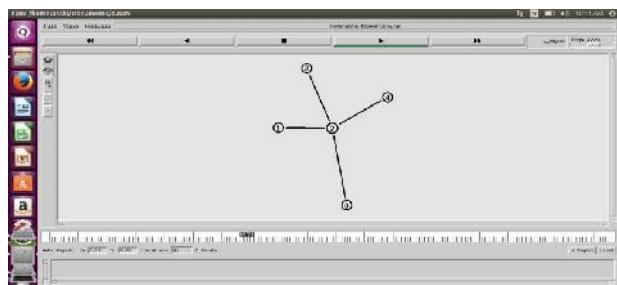


Fig.5 First Index

Policy The outcomes achieved:

- 1) Random path is generated in mesh topology.
 - 2) Source node selected and file with size 20 Mb selected
 - 3) For first index policy no algorithm is used for encryption
 - 4) Total Packet transmission time is 0.324 sec
 - 5) Packet size=1000 bytes
 - 6) Total number of iteration=10
- Thus, the proposed solution is implemented.

X. CONCLUSION

As the proposed system gives security to network and users data for the users who are using 802.11g standard. If the system recognized that the user is authorized person than the necessary features will be provided. If the system recognized that the user is unauthorized then it will not allow the users to connect to network. Features provided for authorized person:

- By using username and password, the system will know that driver is authorized and can connect the internet.
- Appropriate authentication and encryption protocols as per policy will be applied for the user to ensure security of data.

ACKNOWLEDGMENT

The people we would like to thank for their help in presenting this project seminar. We specially want to thank our Principal Dr. B.K. Mishra, our HOD-IT and our project guide Dr. Rajesh Bansode whose advice have inspired and shaped much of the content of this project also guiding us on this project work and spending hours working to improve the content of this project. Also we want to thank all the staff members who help us to improve. So majority of our credit goes to all staff members.

REFERENCES

- [1] S. Kakade, R..Bansode ,—Performance Evaluation of 802.11g Architecture Using Security Protocols Based on Index Policy Method, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), vol.11,Issue 4, Jul . 2016
- [2] N. Baghaei , IEEE 802.11 Wireless LAN Security Performance Using Multiple Clients, Department of Computer Science and Software Engineering University of Canterbury,2003
- [3] N. Borisov, I. Goldberg & D. Wagner,— Intercepting Mobile Communications: The Insecurity of 802.11,” Dept. of Computer Sci. & Software Eng., Canterbury Univ., Christchurch, New Zealand, Jul.2004
- [4] D.Nayak ,D.B.Phatak&A.Saxena,— Evaluation of Security Architecture for Wireless Local Area Networks by Indexed Based Policy Method: A Novel Approach, International Journal of Network Security, Vol.7, No.1,July 2008
- [5] A.TahaZamani, J.Ahmad , Wireless LAN Security : IEEE 802.11g &VPN, International Journal of Advanced Research in Computer Science and Software Engineering Vol.4, No. 2, Feb 2014
- [6] A. K. Agarwal, J. S. Gill & W. Wang ,— An Experimental Study on Wireless Security Protocols over Mobile IP Networks, Technology conference,2004

Livestock Management Using IoT

Veena Kamat

Department of Information Technology

Thakur College of Engineering & Technology Mumbai

Pooja Mistry

Department of Information Technology

Thakur College of Engineering & Technology Mumbai

Nemish Gandhi

Department of Information Technology

Thakur College of Engineering & Technology Mumbai

Namdeo Badhe

Department of Information Technology

Thakur College of Engineering & Technology Mumbai

Abstract- Internet of Things and sensor technology has been widely used nowadays, it can also be used efficiently in managing livestock, and farmers now can embed internet connected sensors on their livestock that does not cause them discomfort. Newer sensors are being created which are smaller in size such as wearables. The problem occurs when the animals don't easily let the sensors get embedded into them. If the sensors are not water-proof or dust proof then they are less efficient. Another challenge is to embed the sensors properly, because sometimes the sensors might fall off.

The regular temperature of cow is around 38.5-39.5 °C. When the temperature is below that the diseases arises are indigestion, milk infection etc. and when the temperature is above 41 °C the diseases arises are influenza and anthrax. When the temperature of the animal is very high on that time it may die. Humidity also affects the cattle health. If it is too humid then the cow can undergo serious stress like situations .hence optimum humidity should always be maintained. Humidity can reduce heat exchange and have enervating impact on the cattle. The adult cow has a heart rate of between 48 and 84 beats per minute. An elevated heart rate may be a sign of pain, and is seen in conjunction with several diseases.

The literature survey created by referring the papers led us to the conclusion that they have used Arduino, which has less accuracy. The sensors were not connected properly to the Arduino, and so the results were inefficient. In an effort to ensure animals do not contract diseases, farmers use antibiotics which in the long-term lead to the evolution of bacteria and the rise of drug-resistant pathogens.

The sensors are used to collect and transmit the signs, such as sensations, and then transfer it to the microcontroller. The Raspberry Pi receives the content from the sensor and development the content accordingly. After determining all the results from the literature survey, the expected correct results are of about 90% in the project. The project scope is planned and decided on the components and hardware to be used.In this project all the hardware components and sensors will read the results through the computer monitor.

Keywords- Heartbeat, Temperature, Sensors.

Agriculture is the major source of income in countryside areas. Animals like cow, buffalo, sheep, goat etc. play an important role in life of rural. They are used as a source of income. Hence animal husbandry becomes a most important concern. In this project it is illustrated that Wireless Sensor Network technology to the farmers This Project illustrate Wireless Sensor Network technology to the farmers. It proposes that a Wireless Sensor Networks should be installed on farms to gather ecological which shall then facilitate farmers not only in monitoring the animals via the Web from outside the farm, but also help the control of farm environmental in remote locations. Electronic means of monitoring cattle health or animal health were developed, technology can go even further in livestock management, and farmers now can embed internet connected sensors on their livestock that does not cause them discomfort. Using information from these sensors, farmers are able to monitor the overall health of the animal.

II. LITERATURE SURVEY

In the paper [1], the proposed solution refers to an automatic temperature and heartbeat measuring system for cattle based on their specific traits. Paper [2] refers to analysis of cattle based on their specific traits. In paper [3] it refers to Sensor Based Animal health Monitoring system, important parameters affecting cattle health such as body temperature along with humidity are continuously monitored. [4] It refers to Arduino UNO microcontroller is utilized to sense the various activities of animals like body temperature, respiration, humidity, heartbeat and rumination.

III. METHODOLOGY USED

The proposed solution, for our project is mainly used to monitor cattle health. If farmer wants to track ad measure his cattle health then this system can be used. The results can be collected using various sensors such as :

- LM35 Sensor: The LM35 series are precision integrated-circuit temperature devices with an output voltage linearly- proportional to the Centigrade temperature. The. LM35 device has an advantage over linear temperature sensors calibrated

in Kelvin, as the user is not required to subtract a large constant voltage.

- Heart rate Pulse Sensor Module: Heartbeat Sensor is an electronic device that is used to measure the heart rate i.e. speed of the heartbeat.
- LCD1602A : These LCDs can be used to display information from the Arduino or any sensor connected to it. For example, you can create a temperature monitoring system

Method:

Phase-1:

Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation.

Phase-2:

After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project.

Phase-3:

Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project.

Implementation: After acquiring the requirements we will proceed to the construction of the application

Phase-4:

After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again.

Phase-5:

After, complete integration and testing of project real time running and operation of the system will be done. Students are expected to validate against their previously stored data in the database to obtain their mark sheets and transcripts.

Algorithm:

1. Start
2. Connect sensors to Raspberry Pi
3. Initialize the components
4. Collect data from sensors
5. Transmit data using Wi-Fi module
6. Display on monitor
7. End

IV. RESULTS

A simulator software: Proteus is used by us for the simulation of a Raspberry Pi and temperature and humidity sensor. The accuracy of most of the systems was around 80%, because it was observed that there was lack of testing of the system and hence the overall efficiency turned out to be low. This project proposed an efficient IOT project based on Livestock Management. Using the simulator app the project has achieved 95-100% accurate results and improved upon the research gaps.

The results have been supported by the research done on this topic. Valid and efficient Livestock Management system can be developed using IOT.

V. DISCUSSION

After analyzing the IEEE Journals and Web references we have come up with the basic flow of the project. After the analysis, in the designing phase of the project we have come up with methodologies and steps to implement the project properly. The first step and the most important step would be the actual assembly of the components. The components should be in proper working condition and the wires should be connected to proper terminals. The flowchart for this project was designed. From the start to the connection of components, the collection of actual data from sensors, transmitting data through the Wi-Fi module and the final displaying of data.

A simulator software: Proteus is used by us for the simulation of a Raspberry Pi and temperature and humidity sensor is.

Referring to the literature survey performed by us and the research gaps observed we found out that the accuracy of most of the systems was around 80%, because it was observed that there was lack of testing of the system and hence the overall efficiency turned out to be low.

Another reason was that the data of the sensor was missing the sensors were turned out to be corrupt or not recognized by the system. But, after performing the simulation using Proteus with the Raspberry Pi and sensors the accuracy was improved up to 95-100%. Hence, improving the efficiency.

VI. CONCLUSION

This project proposed an efficient IOT project based on Livestock Management. Using the simulator app the project has achieved 95-100% accurate results and improved upon the research gaps. Implementation of the same on the hardware components will be done and physically built. This project will accurately track the changes in the environment of the Livestock.

Using sensors the farmers will be able to manage their livestock more efficiently. They will not have to wait for serious illnesses to identify sickness. As agriculture and farming is a very important aspect and a source of livelihood for majority of people in India, this can make a huge difference to a farmers life.

Accurate tracking of the cattle's body temperature and heartbeat can give early indications of any sort of sickness. Which can be avoided later on.

REFERENCES

- [1] Meenakshi, Snehal. S. Kharde Advance Cattle Health Monitoring System Using Arduino and IOT .International Journal of Advanced Research in Electrical, Electronics and Instrumentation
- [2] Rahul Kaushik International Journal Of Scientific Research And Education. Volume 3,Issue 5,Pages-3386-3392,May-2015,| ISSN (e): 2321-7545
- [3] Leena Narayana IJSRE Volume 3 Issue 5 May 2015 Page 3386.e-ISSN: 2395 -0056 Volume: 04 Issue: 03 | Mar -2017
- [4] Tejaswinee A. Shinde, Dr. Jayashree R. Prasad" IoT based Animal Health Monitoring with NaiveBayes Classification

Digitalizing Health Care System Using Aadhar Card

Seema Jamal

Department of Information Technology

*Thakur College of Engineering & Technology
Mumbai, India*

Sangeeta Vhatkar

Department of Information Technology

*Thakur College of Engineering & Technology
Mumbai, India*

Abstract- Technology is changing every industry in significant ways. Integration of digital technologies into everyday life by the digitization of everything that can be digitized. In India healthcare is the primary responsibility as it have less infrastructure and lack of doctors in rural areas. According to the statistics, 75% of doctor's work in urban sectors, 23% works in semi urban sector and only 2% work in rural sector to provide the services to both urban and rural areas, E-health care was introduced in India using Internet Technology, the authentication and authorization model of E-health care using Aadhaar card. This paper discusses development of an efficient electronic medical record with a unique patient identifier. This is to enable hospitals, physicians and individuals in India to efficiently share their records. The Doctor is needed to give there license number and pathology labs need to give there registration number and other details mentioned in the form. The patient just need to give there Aadhar card number and other basic detail like mail id, contact number, address etc. As we see electronic medical record keeping system coming into the market, two problems still remain outstanding, one of them being a need to connect all this data to a unique patient identifier and second problem is a need for a unified data. With Aadhar linked patient, data is available readily and there is no need to depend on patients to bring their previous records. All the relevant medical history is recorded and sorted by the software and is available at a single screen all linked to the patient's unique identifier.

Keywords—electronic medical record, unified patient identifier, unified data, digitizing, Aadhar number, health care

I. INTRODUCTION

As we are in the modern era, the field of medicine is additionally making an attempt to use the data technology to enhance patients' care. One among the foremost important IT applications in health care is the electronic medical record (EMR). The main motivation of this paper is to explore EMR in terms of data retrieval. The broad areas discussed during this paper are the model of medical data and therefore the challenges in indexing, sharing and linking of medical records

Proper delivery of patient care is important and depends on the flexibility to retrieve relevant data regarding the patient as quickly as possible. The technology behind product of the service will be used are a dynamic website is being developed for all the platform. The data are stored on cloud so that storage can be expandable as data keeps on increasing. It is mainly to solve real world problems. The security is taken into consideration so that unauthorized user will not be allowed to access the reports. For security reasons different access right is given to different users

II. SHIFTING TO ELECTRONIC MEDICAL RECORD SYSTEM AND ITS ADVANTAGES

Researchers all over the world have many different definition of Electronic Medical Records (EMR) in their studies. According to Haslina and Sharifah Mastura [1] and Boonstra and Broekhuis [2], EMR plays an important role in providing patients' medical histories, which include one or more computerised clinical information system that collects, stores, and displays patients' information. Furthermore, EMR consists of clinical information repository, order entry, processed provider order entry, pharmacy, and clinical documentation applications. The adoption of EMR system brings each positive and negative problem that requires to be mentioned. It is necessary to analyse the problems to confirm that the adoption of the EMR system process is successful and economical. Few of the advantages of this system are, it is simple and provides remote access to patients' information. Also it improves administrative tasks and is time and space saving. Lastly it brings legibility, is up-to-date and conjointly decreases medical errors. According to Md. Zan [3], the general public hospitals are typically related to slow and inefficient services. Thus, many of us don't seem to be comfortable using the services in public hospitals because the time to get the treatment is long. The projected system looks forward to eradicate issues like these.

III. LITERATURE SURVEY

A. Model of medical information

The work done by rector et al. (1991), [4] permits us to know the essential model for a medical record. Rector and his teammates outline vital ideas like quality to refer that whatever goes within the medical record should be a trustworthy record of

what clinicians have heard, seen, thought, and done. This shows that a medical record not solely contains observations but includes a better level of observation like speculation, thinking method, and action done by clinicians so as to support patient care. Rector et al. additionally introduced information model idea that incorporates a direct relationship with process model and inferred model. Process model of clinical care is regarding "what need to be done or recorded" whereas inferred model of patient explains what really happens to the patient. Thus, concluding that medical record ought to be a descriptive data model of what are often said instead of a prescriptive process model of what it is correct to say or do. This is a very important distinction between a knowledge based data and patient based data, where what is contained in a casehistory isn't always the reality regarding an individual instead is a set of thoughts and observations of an individual at a given time and place.

B. FIRD Framework

The work done by D. C. Leonard et al. (2009), [5] proposes the fingerprint, iris, retina scan, and DNA (FIRD) framework that utilize a patient's biometric characteristics to unambiguously associate them to their medical information. The framework establishes an infrastructure that may distinctively determine a patient to his or her complete electronic healthcare record (EHCR) with actual exactness and accuracy. The framework's inner workings can collect records that aren't properly allotted to the universal patient identifier (UPI), remove records that don't belong to the patient, and correct errors and omissions among the patient's EHCR. The creation of a uniform nationwide electronic healthcare record system within the united states would need some way to match a composite of an individual's recorded healthcare data to a known individual patient out of roughly 300 million individuals to a 1:1 match, leading to a final information compilation that has a whole healthcare history to the healthcare supplier, whereas reducing medical errors and lowering healthcare price. D. C. Leonard [5] proposed this system for implementation in USA in combination with the SSN.c. Unique ID number for patient's The wall street journal in the year 2012 published an article under the HealthCare column questioning the need for patients' of having a Unique ID Number for All Medical Records

RFIDSamanehMadanian [6], in his work proposed the utilization of RFID as the unique patient identifier. In his study RFID has shown its potential to avail healthcare systems enhance patient safety; as [7] reported, around 70% of the respondents to a Bearing Point Study believed that the most paramount implementation of RFID benefits patients by incrementing their safety. In another study [8], a coalescence of bar coding and RFID was examined to verify the patients' medical and identification data

All these study point towards a need for a unique patient identifier for a 1:1 match when an individual's details are searched in the big data pool and shows great concern for the security of the data.

IV. UNIQUE PATIENT IDENTIFIER OPTIONS

Patient Identifiers are paramount for health care organization's day to day operations like the distribution of care, administrative processes, support accommodations, record keeping, data management, and follow-up. The revolution, presently transpiring in our national health care distribution system and within the computer and telecommunication technologies, has increase the scope of these functions across multiple organizations ranging nationwide. Additionally, patients are mobile, visit multiple service providers and treated by multiple organizations. One of the most difficult queries inthis case is however would such system uniquely determine every patient and link him/her to composite medical records in one-to-one match [9]. Since this paper is focused totally on a nationwide unified healthcare database system for the purpose of sharing personal health data through a national health network that can improve the quality of health care. Thus first there is a need to analyze the options for identifying the patients uniquely. There are two approaches to accomplish this task

(1) Statistical Matching and (2) Unique Patient identifier (UPI) [10].

1. Statistical Matching: Statistical matching tries to integrate enough data regarding an individual to form a unique key used to locate his/her electronic health record. It strings attributes such as: name, date of birth (DOB), contact number, address, and gender. The drawback in such key is that some attributes, like name, DOB, and postal code, don't seem to be unique to the individual; others, like address, could change overtime. As the database of records gets larger, additional personal attributes should be superimposed to keep the key distinctive. Searching algorithms utilized in this approach vary from requiring a precise match on a selected set of attributes or to more advance probabilistic pattern matching. The development of statistical matching depends on human to clarify queries and reduce ambiguity this is known as disambiguation.

2. Unique Patient Identifier: Unique patient identification is a technique for linking patients to their electronic medical records that exist globally in a domain (state, country, region, or world). The American Society for Testing and Materials (ASTM, 2000) Standard Guide lists desirable attributes of a UPI, including that it be:

- Unique- Each UPI should be linked to only one individual i.e. different individualcannot have the same UPI.
- Non disclosing- The UPI should not incorporate any personal information for instance name or DOB to prevent revealing the confidential info of the patient.
- Invariable- The UPI should not change in the lifetime of an individual. Except in some exclusive cases like identity theft.
- Canonical- Each person should be linked with only one UPI in order to prevent fragmentation of patient's healthcare data.
- Ubiquitous- Every individual should definitely have one UPI irrespective of age, gender, community or qualification. This is tough to attain, notably if participation is voluntary, however the alternative is a hybrid system, within which some patient information cannot be found employing a UPI.

Various government issued picture IDs available in India today are Voter ID card, Ration card, Driver's License, Passport and

UID (Aadhar card). Following is the result when these identifiers are analyzed for the above mentioned UPI standards:

- Voter ID Card: Voter IDs are not issued to an individual under the age of 18 years, this if used as an UPI limits inclusion of healthcare database for the minors.
- Ration Card: Ration cards are generally issued for a family as a whole, and thus it violates the uniqueness of the UPI.
- Driver's License: Similar to the voter ID card, a driver's license is issued on the basis of age and driving skills, hence not making it available for every single individual.
- Passport: Passport complies with all the above mentioned UPI except being ubiquitous, which suggest an alternate hybrid system, within which some patient information cannot be found using a UPI.
- UID (Aadhar card): Though passport is in line with all the suggested UPI standards, the UID gives additional benefits of biometric details and QR scanning options for faster and more accurate results.

V. PROPOSED ARCHITECTURE

We do have EMR record concepts in India, but the concept is not widely adopted. More and more hospitals are adopting it. Most of the Tier 1 cities based tertiary care and specialty hospitals have Patient and Healthcare Management's systems - but not a comprehensive EMR. The complete adoption process of the EMR system relies on these few factors: [10]

1. Digitalization of all the medical records maintained by the healthcare centers at all level.
2. A high level of data integration, data Interoperability, and data sharing among healthcare professionals and health care institutions are required in order to deliver high-quality healthcare to the patient it serves.
3. Development of a unique patient identifier, to enable medicos, hospitals, patients and other sanctioned users to apportion clinical and administrative records more efficiently.
4. The highest caliber of security needed in context to the sensitivity of the data in concern.

To engender comprehensive database of medical information on the website that contains the medical and other health cognate history and the healthcare information of the patients in one practice utilizing UID. For we require to understand the genuine flow of the information in the system.

A Problem Definition

Currently, the medical patient records available are not opportunely stored and organized in many hospitals. In case of emergency it becomes authentically cumbersome for the medico to retrieve the required patient medical record immediately. There still subsists the challenge in storing these data and sharing them across different hospital networks in case of emergency. So taking the above-mentioned issues into consideration, we are developing a dynamic website which will store the reports of the patient on the cloud.

In the Figure 1 architecture there is a front end and back end operations are performed. In Digitization of health care system using UID, we are making a Dynamic website which is supported by different platforms. In this project we are using the cloud service for the storage of reports. Database is also there which helps in to store the users and their password. Database

also having the different access permission to various different users.

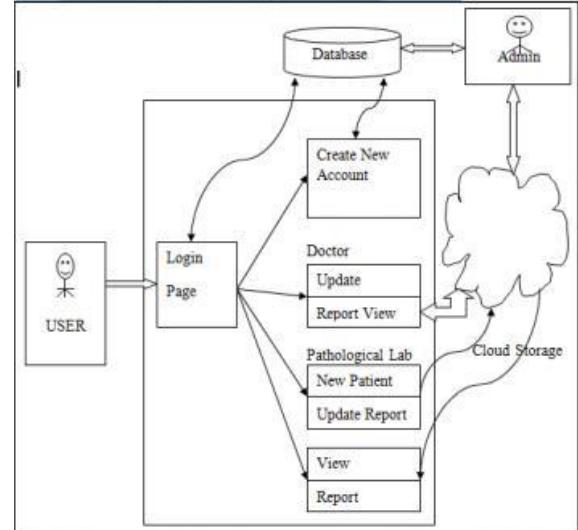


Fig. 1: System Architecture for Digitization of Health Care System

The home page of the website is the Login page. When the user first put the address of our website into browser then the login page will open in that if the user is already has an account they need to just put their login id and password. But if the users don't have the user name and password then for them there is an option link is there by clicking on that link the new user can get register for their user name and password. By clicking on the create new account link the next page will appear which give three options to choose on the basis of different users.

For example if doctor wants to create new account into the website then the doctor will click on create new account link and then he/she will chose the option doctor out of given three options. After this the doctor has to fill some information like their name, email id, address and license number etc. After submitting the information doctor will get their user id and password within three working days after verifying all the information by the admin.[11]

On the off chance that the client is as of now have a record then the client id and watchword is should be put into the shape. After tap on sign in the data is confirmed by database on the backend procedure. Presently if client is specialist then the specialist's landing page is get open where he can refresh the record and also read the report. On the off chance that the pathology lab has login then the pathology lab landing page is open on which they can refresh the reports and also they can include new patient. And if the patient is login then the patient's home page is get open in that patient can only read the reports of theirs. The updated reports will get stored on the cloud.

VI. IMPLEMENTATION METHODOLOGY

The tools will be used are PhpMyAdmin, XAMPP. The refreshed reports will get put away on the cloud. The reports on the cloud are constantly accessible at whenever and anyplace at whatever point required. So when pathology research facility include the reports which will be in pdf shape is put away on

the cloud. In the event that the patient need the report then he can without much of a stretch read the report by simply getting the reports from cloud.Server and MySql. For storage purpose we are using Cloud service. In the project the Data will be flow in following way. Here we have explained the Use Case Diagram of the project.[11]

VII. PROPOSED METHODOLOGY

This analysis aims to highlight one of the mostnecessary challenges that doctors face to have valuatedata regarding their patients, that consequently aid themto give higher service. Patients might have more than onemedical history in differentinstitutes that use differentmedical care systems. the need to connect the practitioner tojust one shared system that holds all the patient'smedical records has been inflated as a muststeptoimprove medical care services. This paperaims to shedlight on the importance of medical records and thereforethe need to improve the medical services provided to thepatient in the country.The proposed system first looks for the best availableunique identification system that already exists withcomplete structured personnel details of the user. For thispurpose UID (Aadhar card) issued by the government is thecurrently best available option. Alsothe UID databaseenables us to integrate the clinical database with thebiometric verification. Here the databases in use are

- 1.TheUID database as maintained by the government
- 2.Thenation wide unified clinical database and
- 3.IndividualDoctor, Admin and Pharmacy database for each hospital using our service. Using the above mentioned databases the overlaying relational database is formed for the system
The importance of biometric verification system can be seen in various cases of emergency, especially when the patient under observation is not in the condition to give outhis information in order to get his medical history on time.In such cases the doctor will have an option of accessing thepatients' vital information using his finger print or his cornea image. This will speed up the diagnosis and treatment hence enabling the practitioners to give a better service.

VIII. FUTUREWORK

Future aspects of the proposed system include using thedatabase for other greater purposes. One of this can bepopulation based disease analysis; the database can be usedto study the effects of various treatment methods, drugs andvaccines on very large data set, hence, providing precise andefficient study results. Other view in which the record canbe used is Health Insurance monitoring and authentication:A module of the Software bundle can offer tools to use themedical history of the patients to monitor their previousrecord and authenticate insurance claims. But most of all itcan be used to aid Crime Investigation. This database can bestructured and utilised to meet the need of certaininvestigations where the medical history of suspects orvictims might be required.

In views of an individual the further enhancement of theproposed system can be integrated with an RFID (RadioFrequency Identification).By using RFID, healthcare will beable to attain value saving and reduce patient death rate as

aresult of the human errors not only by real-time and precisedata but conjointly through perceptive right service at righttime for the correct person and with best quality.

IX. CONCLUSION

On adoption digital health care system at large hospitals in developing countries. a large range of potential edges may be attributed to thosesystem . It will be useful for human life in critical situation when he/she not able to show the hardcopy of the original report. It will be useful in analyzing and monitoring genetic diseases.

The time is also an very important factor now in our life. This project is very much inspired to save the time of people. As the India is developing in digital way, the DIGITAL INDIA Campaign is supported with this project.

REFERENCES

- [1]Haslina and Sharifah Mastura. Acceptance model of Electronic Medical Record. Retrieved January, 3, 2011,at<http://www.docstoc.com/docs/39354269/JAIMS>
- [2]Boonstra, A. and Broekhuis, M. (2010, 6 August). Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Services Research 2010. Published online 2010 August 6. Retrieved January, 3, 2011 at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2924334/>.
- [3] Md. ZanSaari. (2007, 5 December). Kos Rawatan Mahal: Pilihan di TanganPengguna. Retrieved April, 21, 2011, at <http://www.yadim.com.my/Kesihatan/KesihatanFull.asp?Id=175>
- [4] Rector, A.L., No lan, W.A., Kay ,S. "Foundations for an Electronic Medical Record." Methods of Information in Medicine .No 30,pp. 179- 186,1991.
- [5] D. C. Leonard et al. Realization of a Universal Patient Identifier for Electronic Medical Records Through Biometric Technology Ieee Transactions On Information Technology In Biomedicine, Vol. 13, No. 4, July 20
- [6] SamanehMadanian "The Use of E-Health Technology in Healthcare Environment: The Role of RFID Technology", 10th International conference on e-commerce with focus on e-tourism 2016.
- [7] H. J. Yazici, "An exploratory analysis of hospital perspectives on real time information requirements andperceived benefits of RFID technology for future adoption," International Journal of Information Management, vol. 34, no. 5, pp. 603-621, 2014.
- [8] J. L. Wilkerson, and C. L. McDonald Jr, "RFID in healthcare," The Journal of Organizational Leadership & Business Texas a& M University Tex Ark Ana, pp. 1-8, 2007.
- [9] Ahmed Emam, Ahmed Youssef, Samir EL-Masri, Mohammed Alnuem, "Integrating Electronic Health Records Using Universal Patient Identifiers KSA" 11th International Conference on Cyber Warfare and Security
- [10] RAND Corporation, Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the U.S. Health care System, 2008. RAND Health, www.rand.org.
- [11] International Journal of Digital Communication and Networks (IJDCN)Volume 5, Issue 4, May 2017
- [12] International Journal of Digital Communication and Networks (IJDCN)Volume 5, Issue 4, May 2017

Timing Side Channel Attack Implementation And Countermeasure For An AES-ECC Hybrid System

Nishtha Mathur

Department of Information

Technology,

Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Purvi Sankhe

Department of Information

Technology,

Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Shruti Mathur

Department of Information

Technology,

Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Neha Patwari

Department of Information Technology,
Thakur College of Engineering & Technology
Kandivali East, Mumbai.

Ratna Nayak

Department of Information Technology,
Thakur College of Engineering & Technology
Kandivali East, Mumbai.

Abstract—The objective to investigate a solution based on random loop and thread sleep for the timing side channel attack on a hybrid system of AES-ECC, taking advantage of both the virtues of a symmetric encryption algorithm and asymmetric encryption algorithm. The project methodology involves two phases. In phase one, the encryption and decryption process is carried out using the hybrid mechanism where AES algorithm is used to encrypt plaintext and ECC algorithm is then applied to encrypt the AES key, followed by phase two which focuses on implementing software-based countermeasure against the timing channel attack. The intention behind this research was to address and prevent possible vulnerabilities posed by the timing side channel attack. In this study the authors found that the hybrid system based on AES and ECC increases the time complexity of the system by 20% thereby increasing its security. The efficiency of the system is further increased as it requires $2^{126.2}$ operations to recover AES-128 bit key while $2^{190.2}$ operations are needed for AES-192 bit key. It was also found that the countermeasure based on random for loop and thread sleep proves to be successful against the timing side channel attack.. The authors have used the key length of 192 bit AES as compared to a more general 128 bit AES so as to increase the level of security of the system. The encryption process is based on the amalgam of both public and private keys on to which the timing side channel attack has been performed which has not been done in the previous literature.

Keywords- AES, Countermeasures, Cryptography, ECC, Timing Side channel attack, Hybrid algorithm, Encryption, Decryption.

I. INTRODUCTION

The ability to protect and secure information is vital to the growth of electronic commerce and data security. The information being transmitted is vulnerable to various passive and active attacks. Therefore, the information security is one of the most challenging aspects of communication. Cryptography has a significant role in secure communication and it provides an excellent solution to offer the necessary protection against the data intruders. Over a significant time, data encryption techniques took a massive leap from simple methods to complicated mathematical calculations to achieve secure communication.

It hence provides different techniques and tools that is used in most modern security protocols. It is probably the key enabling technology for protecting distributed systems. It not only protects the information

but also provides authentication to the user. Here the original information and encrypted information are referred as plaintext and cipher text respectively. The transformation of plaintext into unintelligible data known as cipher text is the process of encryption. Decryption is the reverse process of encryption i.e. conversion of cipher text into plain text. During communication, the sender performs the encryption with the help of a shared secret key and the receiver performs the decryption.

Cryptographic algorithms are broadly classified as Symmetric key cryptography and Asymmetric key cryptography which are discussed in brief below:

Symmetric Key cryptography: Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be the same or there may be a complex transformation applied on the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link. This need, that both users have access to the secret key is one of the main drawbacks of symmetric key encryption, in comparison to public-key encryption

Asymmetric Key cryptography: Public-key cryptography, also known as asymmetric cryptography, is a class of cryptographic protocols based on algorithms that require two separate keys, one of which is secret (or private) and one of which is public. Although different, the two parts of this key pair are linked mathematically. The public key is used, for example, to encrypt plaintext or to verify a digital signature; whereas the private key is used for the opposite operation, in these examples it can either be used to decrypt cipher text or for creating a digital signature.

Thus, cryptography serves the following purpose:

- i. Confidentiality: The principle of confidentiality specifies that only the sender and the intended recipient should be able to access the contents of a message.
- ii. Authentication: Authentication mechanisms help to establish proof of identities. This process ensures that the origin of the message is correctly identified.
- iii. Integrity: The integrity mechanism ensures that the contents of the message remain the same when it reaches the intended recipient as sent by the sender.
- iv. Non-repudiation: Non-repudiation does not allow the sender of a message to refute the claim of not sending the message.

- v. Access Control: Access Control specifies and controls who can access what.
- vi. Availability: The principle of availability states that resources should be available to authorized parties all the times.

The side-channel attacks are a class of physical attacks in which an adversary tries to exploit physical information leakages such as timing information, power consumption, or electromagnetic radiation. Since they are non-invasive, passive and they can generally be performed using relatively cheap equipment, they pose a serious threat to the security of most cryptographic hardware devices.

A timing attack observes data movement into and out of the CPU or memory on the hardware running the cryptosystem or algorithm. Simply by observing variations in how long it takes to perform cryptographic operations, it might be possible to determine the entire secret key. Such attacks involve arithmetic analysis of timing measurements and have been demonstrated across networks.

Because side-channel attacks rely on the relationship between information emitted (leaked) through a side channel and the secret data, countermeasures fall into two main categories:

- a. Eliminate or reduce the release of such information and
- b. Eliminate the relationship between the leaked information and the secret data, that is, make the leaked information unrelated, or rather uncorrelated, to the secret data, typically through some form of randomization of the cipher text that transforms the data in a way that can be undone after the cryptographic operation (e.g., decryption) is completed.

The first category of attacks displays with exclusive shielding to lessen electromagnetic emissions, reducing susceptibility to attacks, are now commercially available. Power line conditioning and filtering can help deter power-monitoring attacks, although such measures must be used cautiously, since even very small correlations can remain and compromise security. Physical enclosures can reduce the risk of surreptitious installation of microphones (to counter acoustic attacks) and other micro-monitoring devices (against CPU power-draw or thermal-imaging attacks).

Another countermeasure (still in the first category) is to jam the emitted channel with noise. For example, a random delay can be added to deter timing attacks, although attackers can compensate for these delays by averaging multiple measurements together (or, more generally, using more measurements in the analysis). As the amount of noise in the side channel increases, the adversary needs to collect more measurements.

In the case of timing attacks against targets whose encryption times are calculated into discrete clock cycle counts, an effective countermeasure against is to design the software to be isochronous, that is to run in an exactly constant amount of time, independently of secret values. This makes timing attacks impossible. Such countermeasures can be problematic to implement in practice, since even individual instructions can have variable timing on some CPUs.

II. LITERATURE SURVEY

S Guo, X Zhao, F Zhang, T Wang and Z J Shi (2014) research work can calculate the reduced key search space of AES for the given number of side-channel leaks [1]. IDASCA can also interpret the mechanism behind previous ASCAs on AES from a quantitative perspective, such as why ASCA can work under unknown

plaintext/cipher text scenarios and what are the extreme cases in ASCAs. 100 attacks are run for $T = 0.95$ on the first AES round.

D Jayasinghe, J Fernando, R Herath and R Ragel (2010), have investigated the applicability of Bernstein's attack and have implemented many countermeasures and have evaluated their performance and soundness[2]. They have compared the performance impact of many countermeasures and have concluded that the random sleep or loop will be a good countermeasure. These countermeasures fall into one of three categories: software improvements, hardware improvements, and operating system support.

C JunLi, Q Dinghu, Y Haifeng, Z Hao and M Nie (2011), research paper analyses both AES algorithm and ECC algorithm. By mixing the features of AES and ECC, a mixed email encryption system is designed, which can solve the problem such as password system speed and security, which can't efficiently realize the information, data encryption and identity verification[3]. The algorithm can not only enhance the speed of data encryption and decryption, but also solve the problem of key distribution.

X Li, J Chen, D Qin and W Wan (2010), work analyses the AES algorithm and S-box structure, then the replace plan based on S-box structure is proposed to improve AES encryption algorithm, secondly the ECC algorithm is been discussed. Based on this it put forward a mix encryption scheme of improved AES and ECC [4]. This plan has high operation speed, high security performance and strong usability. In this paper, a mix of improved AES and ECC encryption algorithm has been proposed. This algorithm can not only enhance the speed of data encryption and decryption, but also solve the problem of key distribution. At the same time using this plan it can complete the question of authentication, and it also has high computing speed and anti-attack capability, improve the security of data transmission process effectively.

H. Tange and B. Andersen (2013), work in IEEE-WPMC reflects an ongoing research in the field of countermeasures against the attacks mentioned above. The countermeasures are divided into two groups: Some of the attack types can be prevented by a careful implementation and some attack types can only be prevented by a change in algorithms [5].

M. León, R. Aldeco and S. Merino (2005), presents the time required for encrypting the frame body field of the IEEE 802.11 MAC frame using WEP, AES-CCM, and ECC for several data sizes and comparable key sizes. Results confirm that symmetric crypto algorithms are more efficient than the asymmetric for providing the confidentiality service, with WEP being fastest followed by AES-CCM [6]. Time required (in milliseconds) for encrypting the frame body field of the IEEE 802.11 mac frame by RC, AES-CCM and ECC is 0.006, 0.252, 20 respectively.

N R Potlapally, A Raghunathan, S Ravi, N K Jha, and R B Lee (2007), proposed a novel framework for implementing side-channel attacks where the attack is modelled as a search problem which takes the leaked information as its input, and deduces the secret key by using a solver, a powerful Boolean reasoning technique [7]. This approach can substantially enhance the scope of side-channel attacks by allowing a potentially wide range of internal variables to be exploited (not just those that are trivially related to the key). The

proposed technique is particularly suited for attacking cryptographic software implementations which may inadvertently expose the values of intermediate variables in their computations. It demonstrates this attack on standard software implementations of three popular cryptographic algorithms: DES, 3DES, and AES. This attack technique is automated and does not require mathematical expertise on the part of the attacker. An enabling set for AES consists of 128-bit input and output of any of the rounds. This pair is sufficient for breaking the 128-bit key of AE.

D Jayasinghe, R Ragel, J A Ambrose, A Ignatovi and Sri Parameswaran (2014), work investigates the vulnerabilities against power analysis based side channel attacks of all modes of operations, implemented on hardware circuits for low power and high speed embedded systems. Through such an investigation, it shows that AES is vulnerable in all modes of operations against Correlation Power Analysis (CPA) attack, one of the strongest power analysis-based side channel attacks. It also quantifies the level of difficulty in breaking AES in different modes by calculating the number of power traces needed to arrive at the complete secret key. Thus, it concludes that the counter mode of operation provides a balance in between area and power while maintaining adequate resistance for power analysis attacks than when used with other modes of operations. This paper shows that the previous recommendations for the rate of change in the keys and vectors is grossly inadequate and suggest that it must be changed at least every 2^{10} encryptions in CBC mode and 212 encryptions in CFB, OFB and CTR modes in order to resist power analysis attacks. Both the ECB and the CBC modes have to refresh their keys within 210 encryptions to be safe against CPA attack. The keys of modes CFB, OFB and CTR are to be changed within 212 encryptions [8].

S Bulygin, M Zohner, A Heuser, M Walter, J Buchman (2012), presents a practical algebraic side-channel attack: the IASCA. The enhancements have been derived in two ways: First, they have reduced the information required for solving the algebraic system by improving the algebraic representation of AES as well as Hamming weight information. Secondly, they have considered the application of IASCA under the assumption of erroneous information. This work managed a single-trace template attack and then examined the error distribution in detail, to obtain a more practical view on the error rate [9].

III. RESEARCH METHODOLOGY

The proposed system first inputs data from the user that is taken in the form of data blocks and then the key is encrypted by ECC algorithm which is followed by the storing of encryption time in the database that would be used for performing attacks and eventually to implement the countermeasure. The flow of proposed methodology is shown in Figure 1

1. The data block which the user wants to send is encrypted for security purposes.
2. The algorithm used for encryption of the data is AES.
3. The AES key generated is further encrypted with Elliptic Curve Cryptography (ECC).
4. The generated key is provided to the user which will be used to decrypt the AES key block at the time of decryption.

5. The total time of encryption of the data block is calculated and stored.
6. After decrypting the AES key, the encrypted data is further decrypted by the AES key block into its original format for the user to access it.
7. The attacker module will calculate the time for response of encrypted output from server by using various random keys along with a valid key.
8. Correlation program in the attacker module compares the timing details for both the cases and it will generate the possible key space according to the timing details which will be used to determine the correct key combination.
 - i. Store the time for AES encryption in database by a unique file name.
 - ii. Develop a desktop-based application for the attacker where in the attacker can give any key including the one used for the encryption originally.
 - iii. The time for encryption of all the given keys with the same file will be taken and its absolute difference from the encryption time of the original key will be computed.
 - iv. The keys with the minimum difference in threshold will be shown as the valid key thereby reducing the key space.
 - v. As a countermeasure against cache-timing attacks, a random for loop was included in the AES implementation.
 - vi. Before every encryption, a random number is generated, and random loop is run. This changes the encryption time significantly.
 - vii. Because of incorrect timing information received from the server, the attacker could not correctly identify the actual timing pattern.

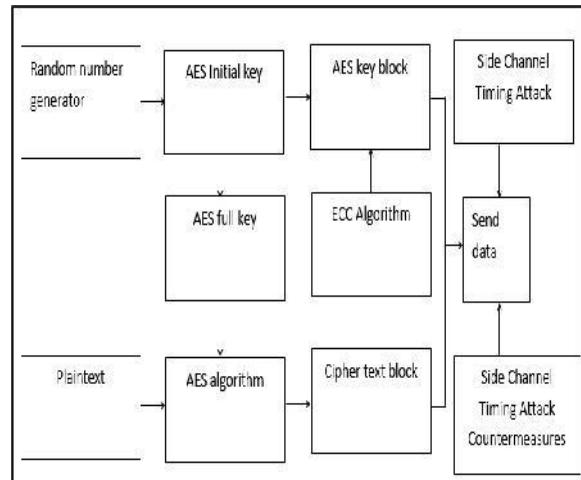


Fig. 1 Hybrid algorithms of AES and ECC

IV. PROPOSED ALGORITHM

The following is the algorithm used for the system:

- i. The user is required to register into the system (database) to get the login id and password for using the software product after which the user login id and password, the user should log into the system using the credentials provided (login id and password).
- ii. After successfully logging into the system, the user can click on the option for “Encryption” of any file that he/she wishes to encrypt.
- iii. A user interface will then ask for the dynamic key to be used for AES encryption and to select the file to be encrypted.
- iv. The software product then computes the encrypted data along with the encrypted key and displays it on the screen for the user to view the details of the block wise encryption done by the software.
- v. For the decryption process, the user will then click on the “Decryption” tab to get the decrypted data and key which can be used to check the correlation between the encrypted and decrypted data modules.
- vi. To carry out the process of side channel attack based on the timing information, a separate desktop application will be used by the attacker.
- vii. This desktop application is linked with the web browser module to get the timing details from the database which is being used to store the encryption time of unique files. attacker
- viii. The attacker module will require different keys to be given as input by the attacker to check for the encryption time for each key on the same data file.
- ix. The time for encryption of all the given keys with the same file will be taken and its absolute difference from the encryption time of the original key will be computed.
- x. The key with the minimum difference in threshold will be shown as the “Valid Key” thereby reducing the overall key space.
- xi. As a countermeasure against cache-timing side channel attacks, a random for loop along with thread sleep is included in the AES implementation.
- xii. Before every encryption, random number is generated, and a random loop is run. This changes the encryption time significantly.
- xiii. Because of the incorrect timing information received from the server, the attacker will not be able to correctly identify the actual timing pattern.
- xiv. Hence for every key given, the attacker will get the information as “Invalid Key”.

V. TOOLS REQUIRED

The software requirements to carry out this research are NetBeans IDE 8.0.2, Flexi core 1.7, Base64 Codec, common-io-2.4, Codec-build17jdk, windows 8 with 64 bit operating system and Intel(R) Core i7-3537U CPU @2.00Ghz 2.50Ghz .

VI. RESULTS AND ANALYSIS

TABLE I. COMPARATIVE SUMMARY OF EXISTING AND REPORTED RESULTS

Parameter	Existing Outcomes	Proposed as per Project Requirement	Improvement
Key length	The key length of AES used for encrypting the plaintext 128-bit.	For increasing the efficiency of the encryption, the key length of AES taken is 192 bit.	Possible key combinations for 128-bit is 3.4×10^{38} while for 192-bit it is 6.2×10^{57} .
Number of Rounds	The work used 10 rounds technique for implementing AES 128-bit algorithm.	For 192-bit key length the numbers of rounds have been increased to 12 rounds from 10.	CPU processing overhead of 5 times.
Key Matrix Used	The Key matrix used for 128-bit key size is 4x4.	The Key matrix used for 192-bit key size is 6x4.	Time complexity increases by 5.16 times.
Algorithm	AES is prone to side channel attacks and leaks information about the key and the key can be deduced exactly.	The AES encryption key will be encrypted with ECC, thereby increasing data security in the network transmission.	Data security is increased by 50% as ECC keys are harder to break due to algorithm complexity
Attack Performed	These attacks represent both a target algorithm and its physical information leakages.	This work will be based on performing Cache Timing Side Channel Attacks that focuses not on breaking the underlying cipher directly but on exploiting weaknesses found in implementations of a cipher.	This work thus focuses on increasing the algebraic complexity of the leakages by countermeasure implementations of random for loop and thread sleep.

VII. CONCLUSION

The project was aimed to provide a mixed AES and ECC encryption algorithm on a standalone system with a much faster throughput as compared to all other existing techniques and still maintaining a trade-off between the memory and time complexities as well as the security of the confidential data. The algorithm thus can not only enhance the speed of data encryption and decryption, but also solve the problem of key distribution. For better efficiency of the system AES-192 bit key has been used which requires 2190.2 operations to recover the key as compared to 2126.2 operations to recover AES-128 key. This increases the CPU processing overhead by 20% thereby making the system more complex to crack by an attacker. The system can handle a variety of file sizes and formats and produces results with 100% accuracy within a short span of time. The results obtained after implementing the proposed hybrid algorithm of AES and ECC were overwhelming and gave a hundred percent accuracy in terms of formats such as images, texts. The memory and time considerations were also fulfilled. Text encryption was found to be the fastest out of all the other formats due to small sized input. On to this hybrid AES-ECC encryption system, a timing side channel attack was performed that uses the stored encryption times to get the desired key space. As a countermeasure to this attack, a random for loop and thread sleep were used. Encryption time is masked so that the attacker will get wrong timing information. This will lead to a wrong key space. Running a dummy for-loop or a thread sleep for a random time is enough to mask the timing data. To sum up, the system fulfills all the objectives mentioned in the scope and lives up to the expectations.

REFERENCES

- [1]. S G'uo, X Zhao, F Zhang, T Wang and Z J Shi, "Exploiting the Incomplete Diffusion Feature:A Specialized Analytical Side-Channel Attack Against the AES and Its Application to Microcontroller Implementations," IEEE Transactions on information forensics and security, vol. 9, no. 6, pp. 999-1011, Jun. 2014
- [2]. D Jayasinghe, J Fernando, R Herath and R Ragel, "Remote Cache Timing Attack on Advanced Encryption Standard and Countermeasure," in IEEE International Conference on Information and Automation for Sustainability (ICIAFs),pp. 177-182, Dec. 2010.
- [3]. C JunLi, Q Dinghu, Y Haifeng, Z Hao and M Nie, "Email encryption system based on hybrid AES and ECC," in IET International Communication Conference on Wireless Mobile and Computing (CCWMC2011), pp. 347 - 350, Nov. 2011.
- [4]. X Li, J Chen, D Qin and W Wan, "Research and Realization based on hybrid encryption algorithm of improved AES and ECC," in IEEE International Conference on Audio Language and Image Processing (ICALIP2010), pp. 396-400, Nov. 2010.
- [5]. H. Tange and B. Andersen, "Attacks and Countermeasures on AES and ECC," in IEEE International Symposium on Wireless Personal Multimedia Communications (WPMC), pp. 1-5, Jun. 2013
- [6] M. León, R. Aldeco and S. Merino, "Performance Analysis of the Confidentiality Security Service in the IEEE 802.11 using WEP, AES-CCM, and ECC," in IEEE International Conference on Electrical and Electronics Engineering, pp. 52-55, Sept. 2005.
- [7] N R Potlapally, A Raghunathan, S Ravi, N K Jha, and R B Lee, "Aiding Side-Channel Attacks on Cryptographic Software With Satisfiability-Based Analysis," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 15, no. 4, pp. 465-470, Apr. 2007
- [8] D Jayasinghe, R Ragel, J A Ambrose, A Ignatovi and Sri Parameswaran, "Advanced Modes in AES: Are they Safe from Power Analysis based Side Channel Attacks" in IEEE International Conference on Computer Design (ICCD), pp. 173-180, Oct. 2014.
- [9] S Bulygin, M Zohner, A Heuser,M Walter, J Buchman, "Improved Algebraic Side-Channel Attack on AES," in IEEE International Symposium on Hardware-Oriented

Accessing Operating System Using Finger Gesture

Parth Parikh

*Thakur college of Engineering and Technology
Mumbai*

Parthparikh06@gmail.com

Bijith Marakarkandy

*Thakur college of Engineering and Technology
Mumbai*

bijit.marakarkandy@thakureducation.org

Abstract: With the massive use of computational devices human computer interaction has become a vital part of our standard of living. Gesture recognition allows humans to communicate with the machine and interact with none peripheral like mouse, keyboard etc. Compare to several existing interfaces, hand gestures have the benefits of being straightforward to use and intuitive. Gestures area unit used widely completely different for various applications on different domains. This includes human mechanism interaction, sign language recognition, interactive games etc. The essential aim of building hand gestures recognition system is to make interaction between human and pc wherever the recognized gestures are often used for dominant computer. With the assistance of this one will cause hand gesture within the vision range of device and desired action is performed by the system. Camera is employed for computer vision that helps in observance in gesture presentation.

I. INTRODUCTION

Computers turn into a key component of our general public. Surfing the web, writing a letter, playing a computer game are only a couple of the cases of the employments of computer. Furthermore, because of the nonstop abatement in cost of computer, they will much more impact our regular day to day existence sooner rather than later. For decades, motion controls have held a persistent place in our visions of the future. We've watch the super heroes, mad scientists, and space cowboys of popular media control digital experiences with just a wave of their hands. We've been captivated by these powerful, natural, and intuitive interactions; imagining what it would be like to have that power at our own fingertips. Whether it may be Tony Stark's workshop, Star Trek's 'Holodeck' all these experiences simply feel magical.

Gesture-based finger gesture recognition is an active area of research in human-computer interaction (HCI), as direct use of fingers is a natural means for humans to communicate with each other and more recently, with devices in intelligent environments. The trend in HCI is moving towards real-time finger gesture recognition and

tracking for use in interacting with video games, remoteless control of television sets, and interacting with other

similar environments. Given the ubiquity of mobile devices such as smart phones and notebooks with embedded cameras, a finger gesture recognition system can serve as an important way of using these camera-enabled devices to interact more intuitively than traditional interfaces [1]

To effectively utilize them, most PC applications require increasingly communication. Due to that human computer association (HCI) has been a vivacious field of research these most recent couple of years. To accomplish normal and immersive human-PC collaboration, the human hand could be utilized as an interface gadget.

Human utilize hand to hand signal to speak with each other. Finger signs are a simple to utilize and characteristic method for interaction. Finger motions are an effective human to human correspondence channel, which moves information in our regular day to day existence.

Utilizing hands as a gadget can help individuals to interact with computers in a more proficient manner. When we communicate with other individuals, our hand developments assume a vital part and the data they pass on is exceptionally rich in correspondence.

II.EXISTING SYSTEM

In past days, because of computer software and hardware advances are nonstop development and achievement, the social life and data innovation have a cozy relationship in the twenty-first century. Later on, particularly the interfaces of buyer gadgets items (e.g. advanced mobile phones, recreations) will have turned out to be unpredictable. The customary electronic info gadgets, for example, mouse, console, and joystick are as yet the most well-known communication way, however it doesn't imply that these gadgets are the most advantageous and normal information gadgets for generally users.

In recent years, the gesture management technique has become a brand new biological process trend for several

human-based physics product, like computers, televisions, and games. This system let individuals will management these product a lot of naturally. The target of this paper is to develop a hand gesture recognition system supported adaptational color HSV model and motion history image (MHI). By adaptational coloring model, the consequences from lighting, surroundings, and camera may be greatly reduced, and therefore the lustiness of hand gesture recognition might be greatly improved.

III. PROBLEM STATEMENT

“Accessing software mistreatment finger gesture” is predicated on idea of Image process. In recent year there's ton of analysis on gesture recognition mistreatment Kinect detector and HD camera however this is often terribly pricey. This paper is specialize in cut back price and improve system lustiness of the projected system mistreatment easy net camera and by removing pricey sensors.

To control the system while not mistreatment ancient input devices and mistreatment solely 3D air house. User ought to be ready to perform all the fundamental day to day task by simply waving their hand within the air

IV. PROPOSED SYSTEM

Most gesture recognition strategies typically contain 3 major steps. The primary stage is that the object detection. The target of this stage is to discover hand objects within the digital pictures or videos. Common image issues contain poor brightness, noise, unhealthy resolution and distinction. The higher setting and camera devices will effectively improve these issues [2].

It is onerous to manage once the gesture recognition system is functioning within the real atmosphere. Hence, the image process methodology could be a higher answer to resolve these image issues to construct Associate in Nursing accommodative and strong gesture recognition system. The second stage embrace beholding. The detected hand gestures square measure recognized to spot the gestures. The third stage is to research sequent gestures to spot users' instructs or behaviours.

The system order is shown in Fig 1: System Architecture. Initially, when we get an image from the camera, we convert the color space RGB to YCbCr. Then, we define a range of colors as ‘skin color’ and convert these pixels to white, all other pixels are converted to black. Then, we compute the centroid of the dorsal region of the hand. After we identify the hand, we find the circle that best fit this region and multiply the radius of this circle by some value to obtain the maximum extent of a ‘non-finger region’. From the binary image of the hand, we get vertices of the convex hull of each finger. From the vertex and centerdisance, we obtain the positions of the

active fingers. Then by extending any one vertex, we control the mouse movement.

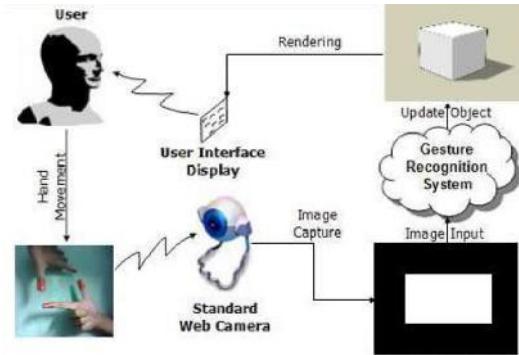


Fig 1: System Architecture

Human hand and finger detection is a vital role in human-computer interface. Fig 2 they typically perform through keyboard, mouse, digitizer, bit panel, and so on. There are several connected works of fingertrip interaction with massive screen show that utilize free hand/finger tracking: interface in VR surroundings, TV management and finger pointer. We tend to describe barehanded interaction between human and huge screen show for Hand and Finger Detection. barehanded means no device and no wires are connected to the user, World Health Organization controls the pc directly with the movements of his/her hand. We tend to establish 3 essential services for tip interaction with massive screen

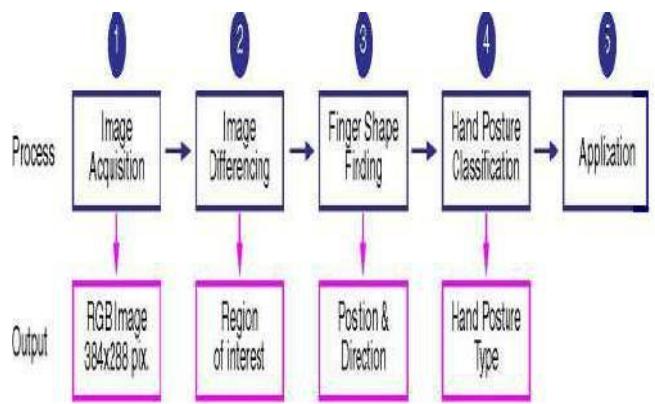


Fig 2: The Hand Posture Recognition and Finger Finding Process

V. ADVANTAGES

High Portability: The projected System cut back the operating of external interface like keyboard, mouse or joystick therefore it makes it high transportable

Reduce Cost: As we have a tendency to square measure removing expensive sensors, it'll facilitate to cut back price of system. Cut back external Interface: The Advantage of System is to cut back External Interface like Mouse and Keyboard.

VI. CONCLUSION

The projected system can solely needs specialised camera to capture input image. This can result in the new generation of human computer interaction (HCI) during which no physical contact with device is required and it'll improve user expertise. Anyone will use this method to control the devices simply, by passing gesture command.

REFERENCES

- [1] Robust Part-Based Hand Gesture Recognition Using Kinect Sensor Zhou Ren, Junsong Yuan, Member, IEEE, Jingjing Meng, Member, IEEE, and Zhengyou Zhang, Fellow, IEEE, 15, AUGUST 2013.
- [2] Accelerometer-Based Hand Gesture Recognition System For Interaction in Digital TV J.Ducloux,P. Colla,P. Petraschine ,Facultad de Ingenieria Universidad Catolica de Cordoba,Argentina,IEEE 2014.
- [3] Real-Time Hand Gesture Recognition System for Daily Information Retrieval from InternetIEEE Fourth International Conference on Ubi-Media Computing
- [4] Mouse Simulation Using Two Coloured Tapes Kamran Niyazi1, Vikram Kumar2, Swapnil Mahe3 and Swapnil Vyawahare4 Department of Computer Engineering, AISSMS COE, University of Pune, India
- [5] Hand gesture recognition for automatic tap system, Sadiq Ullah Department of Computer Science,GIKInstitute of Engineering Topi,IEEE 2015

Health Monitoring System

Mary Margarat Valentine Neela

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai*

Abstract—Numerous systems have been developed to perform health monitoring within the indoor location. Each approach has its own advantages and drawbacks, with the application. In this paper, a system is presented that serves health management system in outdoor location within the specified distance which provides innovative services for mentally disabled people, ranging from alarm, alerts to the caretakers or concerned person. Considering the present scenario at any organization it has been observed that the ratio of mentally challenged patients to the caretaker is not appropriate, and hence there are chances that the caretaker may not be able to assist them efficiently. It may happen that there may be risk to the life of patient in the absence of the caretaker. So we are proposing a solution to address this issue, “A device to monitor mentally challenged person in an organizational environment”

Keywords- *Health monitoring; HealthyPi; wireless infrastructure*

I. INTRODUCTION

Remote healthcare has become a dynamic service with the rising rate of senior citizens and mentally challenged patients. Health monitoring, rehabilitation, and assisted living for the physically/mentally challenged human is an evolving challenge because they need continuous interaction or care between people, and medical and care takers. Mental health care management system face some of the greatest security challenges during physical absence of care takers. Monitoring patient's location is an essential task. An application that can track the location may be beneficial in non-intrusive security, measuring physical activity and in identifying subtle signs of depression.

With an easily accessible, available and worldwide usage of mobile environment deployed with wireless networks can support many current and developing healthcare applications. This technology can achieve the vision of "Ubiquitous Healthcare" or healthcare to anyone, anytime, and anywhere by removing geographical, time and other limitations while increasing the wireless infrastructure and handling the feature. In this paper, health monitoring system is proposed for mentally challenged patients and the patient will be monitored from every location in the organization to avoid potential risk or loss of patient.

It is required to keep a track of the patient's location over a period of time by the care takers. The system should be capable to track several patients. Essentially patient would be wearing a device that relays the identity and placed receivers would record the presence of patients.

A system is proposed with a solution to identify and overcome:

- Risk due to staying in a room that presents a risk for that person without his/her caretaker.
- Inappropriate presence or walking patterns.
- Duration of stay greater than the specified maximum.

- Absence or contact loss.
- Lack of movement for too long in an unusual place.

A. scope:

- a) To monitor the various health parameters of the patients such as: ECG and Respiration front-end, Pulse oximetry (oxygen saturation), Body temperature.
- b) To transfer the above health parameters of the patients to the caretaker and notify them about the same over internet.
- c) Appropriate action can be taken immediately by the caretaker when he is notified with the risk.

II. METHODOLOGY

The patient will be monitored at every location in the organization. The patient will be tagged using passive RFID cards and there will be a long range UHF passive RFID Reader installed at every 25 meter distance. If the patient reaches places like kitchen or medical treatment room in the absence of the caretaker then that potential risk is notified to the caretaker.

Inappropriate presence or walking patterns (nocturnal presence in the dining room and not in the bed) is also notified to the caretaker. Duration of stay greater than the specified maximum (staying in the bath longer than half an hour may indicate that the person has experienced an accident, is sleeping, or simply has some degree of disorientation) is also notified to the caretaker.

Absence or loss of the patient's contact (the patient is moving beyond the limit of the building) is also immediately notified to the caretaker to avoid potential loss or risk. Lack of movement or staying idle at once place for too long in an unusual movement (no movement for 5 min in the middle of a corridor) is also alerted and notified to the caretaker.

All these above risks are notified to the caretaker over internet in his absence with the patient. Now the caretaker of the patient can take the appropriate action against the notification.

Various health parameters including the following are monitored using HealthyPi v3 and the collected data is send over the internet to the caretaker:

- ECG and Respiration front-end is used for monitoring heart pulse rate.
- Pulse oximetry (oxygen saturation) for determining the level of oxygen in the patient's body.
- Checking Body temperature to determine fever etc.

III. PROPOSED SYSTEM

In this system, patient will be monitored from every location in the associated organization. The patient will be tagged using passive RFID cards and there will be a long range UHF passive RFID Reader installed at every 25 meter distance. Within the 25 meter distance of the organization, patient is permitted to move and is monitored by the care takers. In the absence of care takers, potential risk is monitored by the system. This proposed system is to be integrated the existing medical resources and offer smart, reliable, and effective healthcare monitoring service to the mentally disabled people.

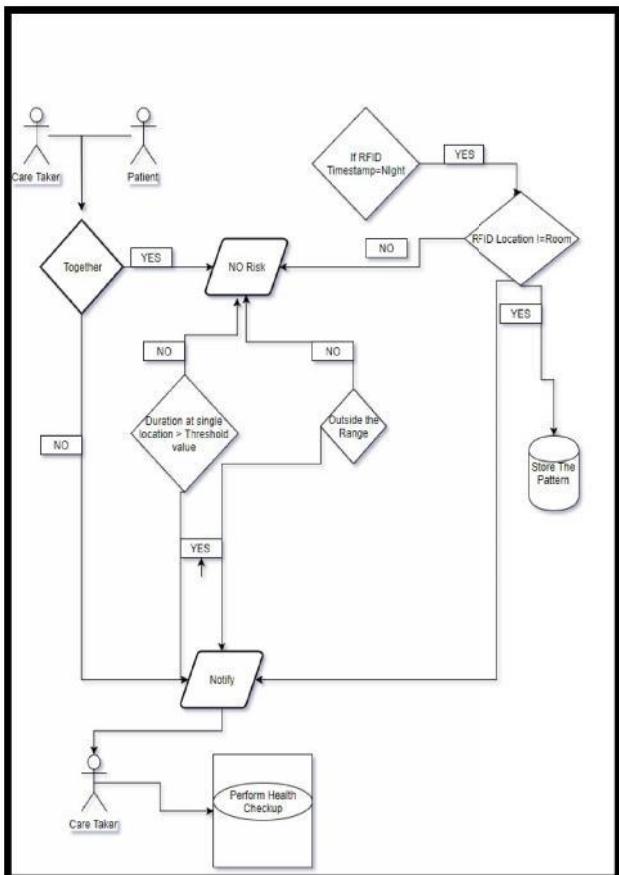


Fig.1: Block Diagram for proposed system

Following are the components which are used for implementation and is described as follows

- RaspberryPi: The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. The Raspberry Pi Model B+ has dual core ARM11 processor with 512MB SDRAM and powers through Micro USB socket of 5V. Sensors are connected to the Raspberry Pi Model B+.

The Raspberry Pi, by contrast, is designed to run an operating system called GNU/Linux hereafter referred to simply as Linux. Unlike Windows or OS X, Linux is open source: it's possible to download the source code for the entire operating system and make whatever changes you desire.

Raspberry Pi directs the data to servers through the GSM module.



Fig.2: Raspberry Pi

- RFID Scanner or RFID reader: RFID Scanner is a plug-play USB device that has only one task – Read the RFID tag.
- RFID Card :RFID Card from which number plate will be read by RFID scanner
- Database :Database of the application will be used to store the data of patients
- Remote Server: The data send by Raspberry pi is stored on a server. Remote server is the place where all patients' data will be stored which are read by RFID card reader. The complete information of patients, caretakers and doctor is registered through website and stored on server. The website can be easily accessible from anywhere.

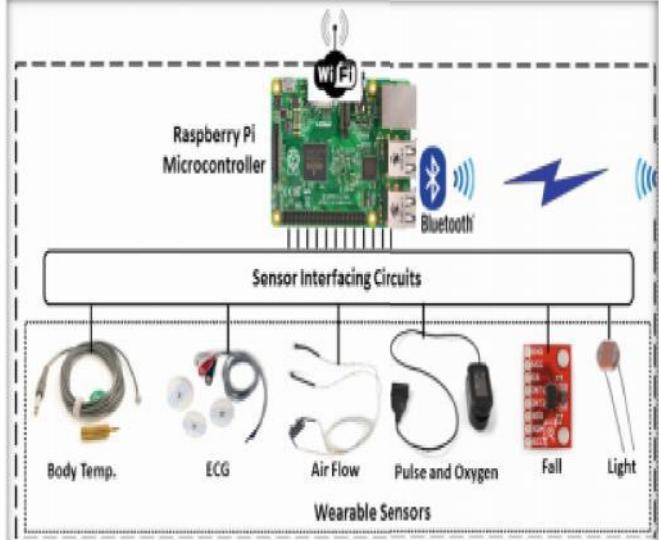


Fig.3: Wearable devices for patients

This system is composed of few components that are described as follows

- capturing / recording a physical view and physical objects or some object feature such as:*

- moving or placing a reader3 in range of a tag;
- moving tags in range of readers;
- capturing a physical view of an object and its surroundings.

This may also involve preprocessing to clean the view, to abstract out an object's main features of interest in a view, e.g., to isolate a voice in a noisy recording.

b) Identifying physical objects:

- detecting a pre assigned object IDs, e.g., RFID, and looking up which object has that ID;
- Assigning an ID to a physical view of the object, e.g., image file ID, and then identifying the object within that view perhaps by relative position, e.g., the object is the top left rectangle in this image

c) Anchoring or relating objects:

- Defining the attributes and relationship of objects with respect to a physical view, e.g., objectmarked on photograph of part of world;
- Defining the attributes and relationship of objects with respect to a virtual view, e.g., objectmarked in some abstract view such a spatial route.

d) Organising or structuring:

Objects to inter relate with different objects within the same view, e.g., a map, and between different views.

e) Presentation:

Superimposing graphic annotation on physical world views such as maps or bydifferent degrees of detaching annotations in different forms from the physical views.

f) Management:

Managing the annotation processes and data including, creating editing, removing, recycling, storing, querying and access control to the annotation data.

IV. RESULTS AND DISCUSSION

Researchers across the world have started to explore various technological solutions to enhance healthcare provision in a manner that complements existing services by mobilizing the potential of the technology. We presented an application system that can automatically create person tracking systems. The system can choose the required modules automatically and optimize their parameters for the given scenario. Patient physical behavior and health improvements can be facilitated by accepting wearable medical devices as landmarks.

As health care services are essential part of our society, automating these services lessen the burden on humans and eases the measuring process. Also the transparency of this system helps patients to trust it. When threshold value is reached, the alarm system that consists of buzzer and LED alerts the caretakers, doctors and the concerned persons who can act more quickly to avoid any potential loss.

In this paper, we have presented a low-power wearable IoT system for active and assisted living healthcare applications. We have outlined the main components of the proposed system and explained their

implementation details. We have built a prototype to illustrate the different performance aspects of the proposed system. The preliminary performance evaluation results have demonstrated the efficiency of the proposed system – despite being a low-cost one. This makes the proposed system a good candidate for implementing a wide set of wearable healthcare systems.

V. CONCLUSION

A Healthcare System using Raspberry-Pi will help Mentally Challenged Patients to create a stimulated environment. The System collects patient's data through RFID Sensors and stored in server. It will be easier for caretakers to track their patients. If any abnormal conditions with patients health will be notified to caretakers which will result safe environment for the patients.

The proposed system will provide the following social benefits:

- An effective way of living provided to the needful.
- A cost-effective solution by using passive RFID tags and open source HealthPi v3.

Our future work will include how to secure the access of the data and will develop a mobile application that allows access of the data on handheld devices

This has three major advantages: associated information, target-oriented healthcare communities, and gamification. In addition, the paper provides detailed research activities concerning how the IoT can address pediatric and elderly care, chronic disease supervision, private health, and fitness management.

REFERENCES

- [1] Xu, B., Xu, L.D., Cai, H., Xie, C., Hu, J., Bu, F.: Ubiquitous data accessing method in IoT-based information system for emergency medical services. *IEEE Trans. Ind. Inf.* 10(2), 1578–1586 (2014).
- [2] Yin, Y., Zeng, Y., Chen, X., Fan, Y.: The Internet of Things in healthcare: an overview. *J. Ind. Inf. Integr.* 1, 3–13 (2016).
- [3] Krishnan, B., Sai, S.S., Mohanty, S.B.: Real time internet application with distributed flow environment for medical IoT. In: International Conference on Green Computing and Internet of Things, Noida, pp. 832–837 (2015).
- [4] R.Sumitha1,C.Vijaya2,S.Parimala: Locations Based Tracking System for Mentally Challenged Person, in International Conference On Innovative Trends In Engineering And Technology2017
- [5] Chetan T. Kasundra1 ,Prof. A. S. Shirsat : Raspberry-Pi Based Health Monitoring System, International Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering
- [6] NuChange - Tracking patients in mental health facilities using Raspberry Pi. Bell Eapen August 11, 2017.
- [7] C. He, X. Fan, and Y. Li, "Toward Ubiquitous Healthcare Services With a Novel Efficient Cloud Platform," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 230–234, Jan. 2013.
- [8] Raspberry Pi 2 Model B (2016). : <https://raspberrypi.org/products/raspberry-pi-2model-b/>

A Proposal for Smart Health Care Center Using IoT for Indian Scenario

Rina Yadav

Thakur College of Engineering & Technology, Mumbai
rina.yadav08@gmail.com

Rajesh Bansode

Thakur College of Engineering & Technology, Mumbai
rajesh.bansode@thakureducation.org

Abstract – The start of this research thought is to create shrewd well-being focuses in India. The proposition fundamentally misuses the possibility of Internet of things and intends to utilize the current innovation and foundation. The work has been spurred by the way that India is deficient as far as number of well-being focuses comparing to the number of inhabitants in nation and in towns particularly. All the more unmistakably if a well-being focus is accessible, Doctors for the most part either are not willing to serve the country territories or they are not accessible 24x7 prompting provincial populace swinging to urban areas notwithstanding for easygoing illnesses. This adds to long lines in clinics and parcel of irritation to patients. Keeping in mind the end goal to maintain a strategic distance from previously mentioned constraints, the proposed demonstrate plans to address the patient without even physically going by the well-being focus and furthermore solution can be produced without Doctor. It is normal that such a shrewd framework would encourage debilitated as well as will add flourishing in a roundabout way to towns. The proposed structure would get a huge change the social insurance area of India.

Index Terms – IOT, RFID TagLoWPAN, CoAP Protocol, Arduino, Mobile Technology.

I. INTRODUCTION

Internet of Things (IOT) is gaining widespread popularity among research community because of its potential to digitize real world physical objects around us. IOT has emerged due to present wireless telecommunication services and ubiquitous presence of Internet. Wireless sensor networks, RFID tags, actuators and various handheld intelligent devices such as mobile phones, PDAs, Tabs etc. are leading to the emergence of IOT

IOT seems appealing and its emergence is being accepted by research and industries, due to its impact on our day to day life both in production and consumption processes. It has opened up wide spectrum of innovative scenarios to improve quality of human life. Sensor networks have already proved their excellence in making human life easy in routine tasks such as controlling water tanks, to save electricity at public places like museums, libraries as well as critical tasks such as habitat monitoring, controlling/ assisting industrial processes etc.

Intelligent devices such as mobile phones have been transformed from embedded keypad based traditional

phones to lightweight touch screen-based devices. Applications of sensors in our day to day life are numerous which indicates their importance, however they are constrained due to limited battery life which limits their use. Radio Frequency Identification Tags (RFID) technology provided substitute to sensors as these tags can be used to identify, track and locate any object using unique Electronic product code (EPC) which is encoded in these tags.

RFID tag comprises of a small chip, an antenna and a cover for encapsulating chip and antenna. Antenna receives signal from RFID reader device and transmits the tag ID to it. These tags can be either active or passive based on power source. Active tags are associated with a battery life and thus depend on it for their lifetime, just like sensors. However, passive tags acquire energy from reader device either through magnetic induction or electromagnetic wave capture techniques [8].

Signal received by RFID antenna, produces a current in it through induction, which is further utilized by the antenna to revert back the tag ID to the reader. This technique can help transmit tag Id to a radio range of few kilometers. Thus RFID tags eliminate battery limitation of sensors, further being small in size they can be embedded in any real life object for its monitoring.

Thus RFID tags are helping great deal to convert every real life physical object into digital entity. Such RFID sensing objects will form their RFID sensor networks with reader devices as sinks of data generated. Emergence of these RFID sensor networks, in our day to day life will fill the gap in omnipresence of Internet and will help IOT spread its roots in our society. ‘Anytime, Anywhere, Any media’ computing has turned into reality with every object embedded with either RFID tag or sensors, these when combined with already existing wireless communication technologies make everything digitized and on Internet.

This gives avenue for large range of innovative applications such as smart homes, E-healthcare, traffic monitoring and route management, resource management at retail stores, automated checkouts at shopping centers, condition based maintain of vehicles are some possibilities. Applications of IOT have been divided into four categories i.e.

Transportation and logistics domain

- Healthcare domain
- Smart environment (home, office, plant) domain

- Personal and social domain

The current proposal focuses on developing smart healthcare centers and it outlines a framework for mobile based portable health care services for rural population in India. Next section provides an overview of relevant literature in this field. Section 3 provides proposed framework and section 4 finally concludes with requirements to develop the proposed smart healthcare center and its future scope.

II. RELATED WORK

This section explores work already done in the field of IOT. Atzori et al. [1] presented a survey on Internet of Things highlighting the most appealing point of IOT which is the integration of several technologies and communication solutions. Their work emphasized that any contribution towards advancement of IOT must be a result of synergistic activities in various fields such as telecommunications, informatics, electronics and social science.

Coetzee and Eksteen [2] have elaborated IOT domain and emphasized that various application domains such as Green IT, energy efficiency and logistics have already started gaining benefits from it. Because of large potential of this domain, IOT has grabbed higher priority on the research agenda of academia, industry and governments such as IBM's Smarter Planet, Microsoft's Eye-on-Earth platform and HP's Earth initiative, just to list a few. European commission and Chinese Government is also making efforts in this direction.

However, advancements in IOT is also raising trust and security issues simultaneously. Standardized protocols and governance strategies are required for IOT to work at global level. Survey on IOT presented by McKinsey Global Institute [6] highlighted that most IOT data being captured today is not used currently. Presently the captured data is used only for anomaly detection and control; however it may be used for optimization and prediction which is of more importance. Further, they pointed that there is large scope for IOT in developing economies such as India.

The critical investigation of available literature clearly reflects that IOT is the demand and strong requirement of developing countries and there is a huge gap prevailing between theory and practice. The proposal submitted aims to fulfill this gap in one of the domains i.e. health care.

Next section presents a proposal for Mobile E-care Health Services using IOT for Indian Scenario.

III. PROPOSED SMART HEALTH MODEL

➤ USING RFID-LoWPAN :

This work aims at designing and implementing an IOT – aware Smart Health Care having as main peculiarity, the capability to readily combine different, yet complementary,

technologies enabling novel functionalities. Basically, the system we envision should be able to collect, in real time, both environmental conditions and patients' physiological parameters and deliver them to a control center.

At this point, an advanced MA(Medical Assistant) should analyze the received data and send alert messages in case of emergency. The conceived SHS has been put into effect according to the architecture illustrated in Figure 2.

- As shown, it is composed of three main parts: The RFID-WSN (Wireless Sensing Network).
- The IOT smart gateway;
- The user interfaces for data visualization and management.

The same interface allows doctors with specific privileges to access both real time and historical patient data. Such information can also be managed remotely by the medical staff through a mobile software application. The HSN consists of an integrated RFID-WSN 6LoWPAN network composed of four typologies:

- 6LowPAN border routers (6LBR);
- 6LowPAN routers (6LR);
- 6LowPAN router readers (6LRR);
- 6LowPAN host tag (HT).

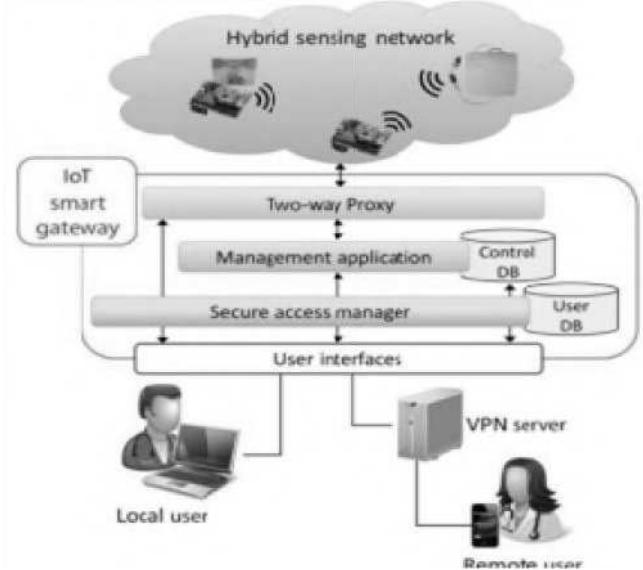


Fig. 1: Architecture for Proposed RFID-LoWPAN

According to the 6LoWPAN standard, the 6LBR is in charge of connecting the network to the Internet by translating 6LowPAN packets into IPv6 packets and vice versa, whereas the 6LR provides forwarding and routing capabilities. Referring to the proposed RFID-WSN integrated system, the 6LRR is defined as a 6LR node interfaced with an RFID reader while HT identifies a typical 6LowPAN Host (i.e., a node without routing and forwarding capabilities) interfaced with an RFID Gen2 tag. The proposed SHS assumes that several 6LR are deployed in the hospital to collect data from the environment, such as temperature, pressure, and ambient light conditions. In addition to the sensing capabilities, the main function of

6LRR nodes, instead, is to track patients, nursing staff, and biomedical devices labeled with RFID tags.

In particular, we envision patients wearing an HT node, which is capable to detect important physiological parameters, such as heartbeat and motion. Sensed data are periodically logged on the user memory of the RFID, thus allowing 6LRR nodes deployed in the environment to retrieve and deliver them to the IOT Smart Gateway. This last one is connected, on one hand, directly with the HSN and, on the other hand, with the Internet through a local area network (LAN). From literature review it is clear that IOT has large possibilities for innovative applications to help improve human life.

Among four main categories of IOT applications listed above, healthcare domain [6] is the one, most beneficial for common people, especially in India. In India large population still lives in villages and is deprived of good healthcare facilities. However, in villages also, the Internet facility is being made available (Owing the credit to Scheme Digital India) and rural population is already making use of mobile phones to its maximum extent. This factor motivates the present proposal that using basic internet or telecommunication services, RFID tags and existing dispensaries in villages, we can facilitate promising basic healthcare services to everyone. This proposal can also contribute towards

Government of India Digital India initiative. Presently, all most all villages in India have at least one health center to provide basic healthcare services, however there is scarcity of doctors in those dispensaries, due to which people have to visit nearest urban cities to avail medical facilities. Because of limited number of government hospitals in urban cities and still limited doctors in those hospitals, creates bottleneck in providing satisfactory medical facilities to all citizens. Long queues at all hospitals in India, clearly indicates demand for a better alternative. Using IOT and its enabling technologies, existing government health centers can provide services for common diseases, such as common flu, cold, cough, typhoid, malaria etc. This meagerly requires establishing one computer system with internet connectivity to healthcare server established at nearest urban government hospital or the server may be established in cloud.

Villagers will be required to visit local health center once where with the help of a medical assistant they will undergo registration on healthcare server. On registration, present status of vital organs will be recorded along with any medical history and a unique RFID tag will be issued to the person, containing registration identity of that person. While RFID tag being very small can be embedded within a wrist band, the RFID reader will remain available online at the health center itself. Now whenever a person is sick, instead of visiting the health center physically, the patient will only be required to press a button in RFID band which in turn transmits the identity information to RFID reader at health center.

On this call, RFID reader identifies the patient with its registration number and will access patients health card

from healthcare server. The health card along with present symptoms of the patient will be submitted to an expert medical system such as Mycin (a new software may also be developed), which will be installed at healthcare server. This expert system is capable of generating a prescription to the patients, even in the absence of a doctor based on present symptoms and past medical history of the patient. Based on generated prescription, medical assistant may provide medicines to the patient (in the absence of Doctor too).

This Expert system and health care database may be kept on cloud, so as to make it accessible everywhere in the country. Thus, the patient can move anywhere in India and can avail medical facilities using that RFID tag, which will help any doctor to become familiar with patient's history and medicines already prescribed and taken. Further, common problem of villagers will get timely identified and resolved avoiding long queues which in turn would save lot of time, money and energy. Figure 2 provides high level view of proposed Smart E- Health Services System.

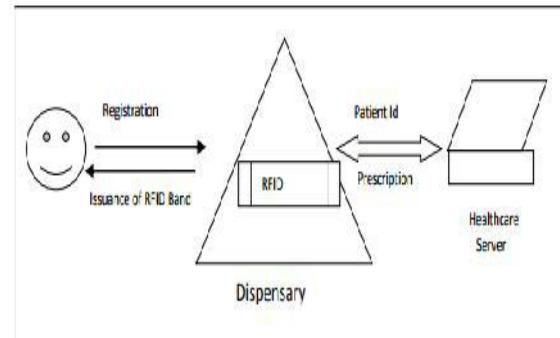


Fig. 2: High Level View of E-Health Care Smart System

➤ USING CoAP

Overview of CoAP(Constrained Application Protocol)

It is a software protocol intended to be used in very simple electronics devices that allows them to communicate interactively over the Internet. It is particularly targeted for small low power sensors, switches, valves and similar components that need to be controlled or supervised remotely, through standard Internet networks [14]. Therefore, efficiency is very important. CoAP can run on most devices that support UDP or a UDP analogue. Message types involved in CoAP are Confirmable, Non confirmable, Acknowledgement, Reset messages. Confirmable requires Acknowledgement whereas Non-Confirmable doesn't require Acknowledgment.

Reset message indicates missing of few contexts.

- GET: The GET method helps in retrieving information from that of server.
- POST: The POST method requests the server to provide data for the user.
- PUT: The PUT method helps in updating or creating the information in the corresponding URI

- **DELETE:** The DELETE method helps in erasing the information stored.

The usage of CoAP in patient monitoring application is to form a client-server relationship, whereas Doctors use the browser as client and the server program runs in the GUI for updating the browser dynamically. Thus the Method definitions are used efficiently for remote monitoring application.

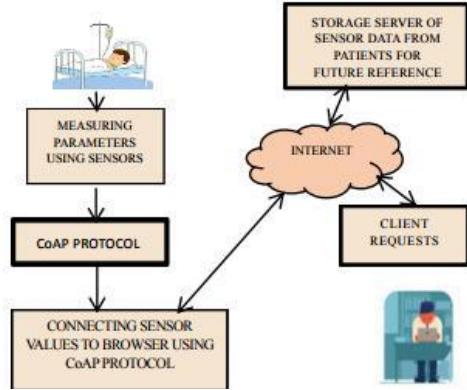


Fig. 3: Architecture of proposed system using CoAP

As per Figure 3, Patient is monitored by the sensors mounted on them. Servers will host the information on the network. The server collects the data from the sensors. On the other side, Doctors are acting as clients. If doctors request server for data, the CoAP protocol helps to connect server with client to display the current status and check for variation of sensor values [11]. If there is any variation, Doctors can give medication indicating emergency situation for the patient. This medication is done in browser using CoAP Protocol. The data of patient is stored in some data storage device for future purposes.

In multi-hop flat wireless topology as shown in figure 4, there are one or more intermediate nodes along the path that receive and forward packets via wireless links. During transmission and reception in multi-hop networks, nodes communicate with each other using wireless channels and do not have the need for common infrastructure or centralized control. The above Figure 4 explains the Multi-hop topology with Border router, Client and server.

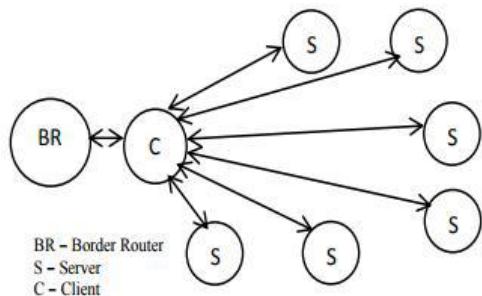


Fig.4: Communication between border router, Client & Server

➤ USING ARDUINO

Here proposed system is designed to reduce the difficulties of rural people. It merges the idea of an active E-Health Care System. The system mainly reduces the human effort by connecting to people from distant places. The sensors play a major role in this human less health care facility. Temperature sensors are used to measure the body temperature of a patient. The sensors are interfaced with Arduino using a coding. Likewise, a heartbeat sensor is used to measure the pulse of the human body. Also, aECG sensor module is interfaced, it is used to measure the rhythm of heart and blood flow through it. These sensors are together interfaced with Arduino. Then, a server and web page part are introduced with respect to IOT which acts as a link between the patient and doctor.

A power supply which is given either by means of USB cable or battery. In order to control the power supply, a charge controller might be used to minimize the effect. These supplies are connected to an Arduino controller. Then, the interfacing of sensors takes place from the temperature sensor. The connection pin of 5V, A0 and GND are connected to Arduino and interfaced. Heartbeat sensor are also connected in same procedure to measure the heartbeat pulse.

Then, an ECG sensor module is used. This sensor has a five-pin connection. They are connected with respect to Arduino. It measures the difference in PQRST waveforms of heart. These things are interfaced together with the Arduino. Next, the Wi-Fi module is connected to the Arduino. It acts as a source of creating a specific hotspot connection through a connection between the hardware and the system is made together. The system gets the specific set of URLs through which the interfacing results are displayed through it.

IV. CONCLUSION

The proposed model can be effectively executed utilizing existing RFID-LoWPAN, CoAP and Arduino innovation and a specialist social insurance framework, for example, Mycin. Another smart e-wellbeing framework can likewise be planned considering the Indian medical problems and conditions. The proposition would go about as a guide to wiped out and would likewise contribute towards Digital India. It is noted from the above studies that:

The RFID peruser is a system associated gadget that can be for all time appended or versatile. It utilizes radio recurrence waves to transmit signals that actuate the tag. Once enacted, the tag sends a wave back to the radio wire, where it is converted into information.

LoWPAN is a class of advances that are intended for low-control, long-go remote correspondence, thus they are perfect for use inside vast scale arrangements of low-control IoT gadgets like remote sensors.

Constraint application protocol is especially focused for little low power sensors, switches, valves and comparative segments that should be controlled.

Arduino can detect the earth by accepting contribution from an assortment of sensors and can influence its surroundings by controlling lights, engines, and different actuators.

REFERENCES

- [1] Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- [2] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things—promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE.
- [3] Babar, S., Stango, A., Prasad, N., Sen, J., & Prasad, R. (2011, February). Proposed embedded Security framework for internet of things (IOT). In Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on (pp. 1-5). IEEE.
- [4] Ukil, A., Sen, J., & Koilakonda, S. (2011, March). Embedded security for Internet of Things. In Emerging Trends and Applications in Computer Science (NCETACS), 2011 2nd National Conference on (pp. 1-6). IEEE.
- [5] Weber, R. H. (2010). Internet of Things—New security and privacy challenges. *Computer Law & Security Review*, 26(1), 23-30.
- [6] Jara, A. J., Zamora, M. A., & Skarmeta, A. F. (2011). An internet of things---based personal device for diabetes therapy management in ambient assisted living (AAL). *Personal and Ubiquitous Computing*, 15(4), 431-440.
- [7] The Internet of Things: Mapping the Value beyond the Hype, Executive Summary by McKinsey Global Institute, and June 2015.
- [8] Ant, R. (2006). An introduction to RFID technology. *Pervasive Computing*, IEEE, 5(1), 25-33, January-March, 2006.
- [9] Istepanian, R. S., Zitouni, K., Harry, D., Moutosammy, N., Sungoor, A., Tang, B., & Earle, K. A. (2009). Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. *Journal of Telemedicine and Telecare*, 15(3), 125-128.
- [10] Istepanian RSH, Jara A., Sungoor A., Philips N. (2010). Internet of Things for M-health applications (IoMT). AMA-IEEE medical technology conference on individual healthcare, Washington.
- [11] Chander RPV, Elias S, Shivashankar S, Manoj P (2012) A REST based design for Web of Things in smart environments. International Conference on Parallel Distributed and Grid Computing pp: 337-342.
- [12] Shelby Z, Hartke K, Bormann C, Frank B (2013) Constrained Application Protocol (CoAP). Corecoapp: 1-118.
- [13] Sunkari V (2015) Framework for Providing Security and Energy Saving Using Elliptic Curve Cryptography in Wireless Sensor Networks. *IJCSIT* 6: 3604-3608.
- [14] Dimicic T, Krco S, Gligoric N (2012) CoAP (Constrained Application Protocol) implementation in M2M Environmental Monitoring System. *E-society Journal* 21.

Raspberry Pi Based Cost Effective Interactive Device for Blind, Visually Impaired, Deaf and Dumb People

Vijay Karotra

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai,
vj.gerrard@gmail.com*

Tanishq Mehra

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, tanishqmehra97@gmail.com*

Shubham Maheshwari

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai,
shubh.maheshwari01@gmail.com*

Shridhar Kamble

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, shridhar.kamble@thakureducation.org*

Abstract - Helping differently able people (visually impaired, physically challenged, deaf and dumb) is a difficult task as there is no solution offering a common platform or a solution that is cost effective. A solution is proposed where four modules viz - text to speech (dumb), gesture recognition (physically challenged), image to voice (visually impaired) and voice to text (deaf) are integrated onto a Raspberry Pi to provide a solution which is both, based on common platform and cost effective. The image to voice module is based on cutting edge neural nets (CNN and LSTM) to read an image and output the description. The gesture recognition module captures the image, identifies the gesture and relays the output by cross checking the gesture with predefined meanings. Both voice to text and text to speech are simple modules which work on natural language processing APIs. The Raspberry Pi is used as its small size and cheap cost suit the project requirements very well.

Keywords - Image-to-voice, gesture-recognition, speech-to-text, text-to-speech, CNN, LSTM

I. INTRODUCTION

According to the World Health Organization, there are more than 39 million legally blind and 285 million visually impaired people living across the world, and this number is still growing at an alarming rate [1]. As currently most of the camera-based devices are capable enough to capture photographs with high resolutions, we can make use of these images to generate captions based on graphics present on images which can be speak out for visually impaired people so that they also have look and feel where they are actually, what is their surrounding environment and finally get a better visualization of what is happening around them. The work focuses on finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. There are many existing solutions for assisting blind or visual impairments people but actually they are not affordable and not

readily available to the community of visual impairments or blind peoples. It's a very tough job to address the issues of People with Visual, Hearing and Vocal Impairment through a single aiding system. The work focuses on gesture recognition and finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. Proposed system will describe the content of an image using well-formed English sentences for better understanding of their surroundings which will have great impact on them.

Comment [V3]: Keep the font size 9

II. OVERVIEW

There are many existing solutions for assisting blind or visual impairments people but actually they are not affordable and not readily available to the community of visual impairments or blind peoples. It's a very tough job to address the issues of People with Visual, Hearing and Vocal Impairment through a single aiding system. The work focuses on gesture recognition and finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. Proposed system will describe the content of an image using well-formed English sentences for better understanding of their surroundings which will have great impact on them.

Comment [V1]: Keep it bold, italic and font size 9

Comment [V2]: Give the number in Roman and Capital and all caps and central align font size 10

III. LITERATURE SURVEY

Text to speech module uses natural language processing to convert letters into phonetics which are combined to create speech [2]. Gesture recognition uses convex hull and adaptive boosting in OpenCV [3]. Image to voice is done in two steps first a CNN is used for feature extraction and the identified features are expressed in a sentence using a LSTM [4]. Speech to text uses an isolated word recognition with limited vocabulary [5].

IV. PROBLEM DEFINITION

A. Phase 1(Information Gathering):

Gather appropriate data set to train the neural networks. All available APIs for speech to text and text to speech need to

find. Most important gestures and their meaning need to be gathered as well.

B. Phase 2(Analysis & Design):

The various algorithms to train the neural net are identified and compared and the best one is selected. The numerous APIs available for speech to text and text to speech need to be compared. Various possible methods of gesture recognition are compared to find the one best fitted for our use.

C. Phase 3(Implementation & Testing):

Separate implementation of each of the modules i.e. Caption generation, Text-to-speech, Speech-to-text and Gesture recognition and checking the efficiency of each of the modules. Finally, integration of the modules with the required hardware equipment and testing of the system as a whole.

D. Phase 4(Deployment):

Deployment of the integrated modules into Raspberry Pi after connection with the equipment's including WebCam, Zigbee, microphone, speaker, and screen. A full-fledged system will be provided which allows access to all the modules as per requirement.

V. METHODOLOGY USED

The figure below shows the connections to the Raspberry Pi. The webcam is to take the input for the Image Caption Generation and Gesture Recognition. The speaker acts as a output for the Image Caption Generation and the Text-to-Voice modules. Microphone takes the input for the Voice-to-Text which is finally displayed through the Monitor.

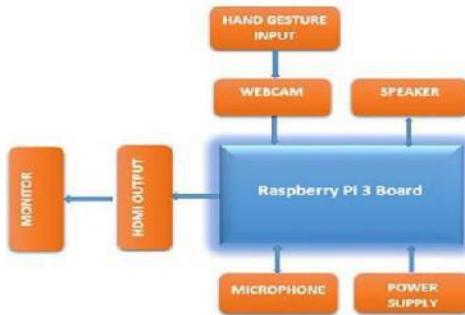


Fig. 1: Architecture of Raspberry-Pi based Automated Device

The speech to text module will be implemented using available open source APIs. A mic will be used for input and the result will be displayed on a screen once the input has been processed.



Fig. 2: Flowchart for Speech to Text

The text to speech module will also use available open source APIs. The input in this module will be a keyboard and the output will be on speakers.



Fig. 3: Flowchart for Text to Speech

The gesture recognition module captures an image from a video then the boundaries between the fingers is detected. Then we identify the region of interest and the gesture in that region is recognized and the corresponding predefined message is outputted on the screen.



Fig. 4: Flowchart for Gesture Recognition

The image to voice module works in two phases and each of these phases use a neural net. The first phase uses a Convolutional Neural Net (CNN) which is used for feature extraction from the input image. The output of this phase is the objects which have been identified and they serve as the input of the next phase. The second phase uses a Long Short-Term Memory (LSTM) which converts the given words into a meaningful sentence. The dataset used in Flickr30k image set.



Fig. 5: Flowchart for Image Caption Generation

All of these modules are finally integrated on a Raspberry Pi via which the users will be able to interact with the modules.

VI. TECHNOLOGY USED

- I. Python is used for scripting where the neural nets are programmed and trained and the APIs required for other modules are called.
- II. All the modules are integrated onto a Raspberry Pi with Raspbian OS installed.

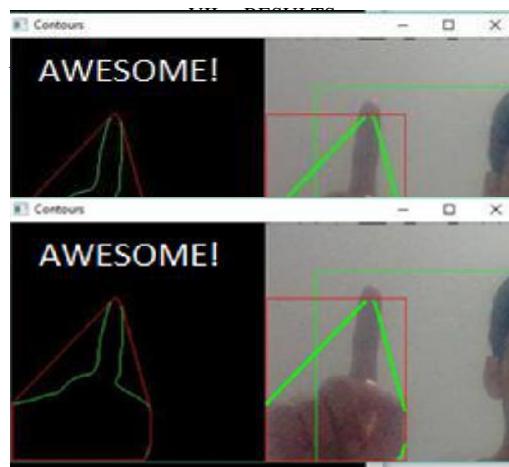


Fig. 6 Output for Finger 1



Fig. 7 Output for Finger 2



Fig. 8 Output for Finger 3



Fig. 9 Output for Finger 4

Table 1: Result for Gesture Recognition

	One Finger	Two Finger	Three Finger	Four Finger
Total No. of Trials	45	43	45	40
No. of correct finger gesture	42	40	44	38
% accuracy	93.33	93.02	97.78	95

B. Speech to Text:

Google API was used for effective speech to text conversion. The input was given via a microphone which was then processed by the Google's API and the output is then printed on to the screen.

The given input was as follows: Hey! This is a test for Google API Speech to text. In this module we are trying to convert the outspoken speech into readable text.

The output for the same input is shown in the figure below.

```
F:\Code\Personal\Major_Project\Ccode>python google_speech.py
Say something!
Google Speech Recognition thinks you said hey this is
a test for Google API speech to text in this module we
are trying to convert the outspoken speed into a read
able text
F:\Code\Personal\Major Project\Ccode>
```

Fig. 10 Output for Speech to Text

C. Text to Speech:

The gTTS Package is used to convert a given input of text into a speech. The program accepts text as an input which is then converted into speech and saved in a mp3 format.

The following text was given as the input: "Look, I didn't want to be a half-blood. If you're reading this because you think you might be one, my advice is: close this book right now. Believe what-ever lie your mom or dad told you about your birth, and try to lead a normal life. Being a half-blood is dangerous. It's scary.

Most of the time, it gets you killed in painful, nasty ways. If you're a normal kid, reading this because you think its fiction, great. Read on. I envy you for being able to believe that none of this ever happened. But if you recognize yourself in these pages-if you feel something stirring inside-stop reading immediately. You might be one of us. And once you know that, it's only a matter of time before they sense it too, and they'll come for you."

The resultant output was a 58 seconds long mp3 file which correctly recites the paragraph verbatim.

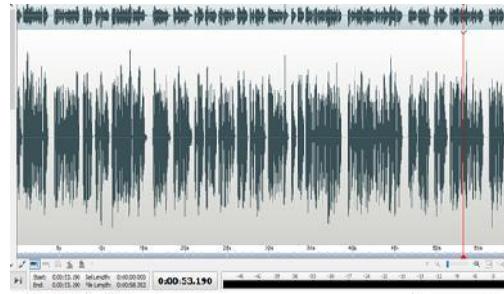


Fig. 11 Output for Text to Speech

VIII. CONCLUSION

It's a challenging task for physically impaired people (blind, deaf or dumb) to communicate with rest of the world. Blind or dumb people generally use sign language for communication. But in reality very few people understand sign language. Proposed system aims to lower this barrier in communication by developing a portable electronic device which will act as a smart assistant for physically impaired people. Primary goal of proposed system is to provide a standard lifestyle for physically impaired people. In a proposed system prototype model for blind, visually impaired, deaf and dumb people has been designed by combining four modules in a single compact device. The important key factor of this project to facilitate physically impaired people and provide them opportunity to interact with rest of the world with the help of technology. Device would be portable and less in weight.

IX. IMPROVEMENTS IN RESEARCH GAPS

The gesture recognition module allows for the accuracy of 70% which is at par with that achieved in the literature survey. Better algorithm is being worked upon which helps enhance the contour detection by the process of adaptive boosting which will make the camera emphasize more on the hand gestures as compared to the surrounding in the field of vision of the camera. The image to voice module is currently working only on images. If we have better computing power and real time data we can apply the same concepts to video and thereby increasing the usability of the module.

The speech to text module which works on Google Speech Recognition does not take into account the usage of punctuations which can make the sentence difficult to understand in certain situations. Larger set of parameters and proper usage of time gaps between words can help improve structuring the sentences and make them grammatically more correct.

REFERENCES

- [1]"Visual Impairment and Blindness." World Health Organization. (2014). Web. 10 Apr. 2016
- [2] Dragos Burileanu, "Basic Research and Implementation Decisions for a Text-to-Speech Synthesis System in Romanian", International Journal of Speech Technology 5, pg. 211–225, 2002
- [3] Premanand K. Kadbe, Ruchi Manish Gurav, "Real time finger tracking and contour detection for gesture recognition using OpenCV", International Conference on Industrial Instrumentation and Control (ICIC), Vol.3 Issue 5, March 2015
- [4] Christopher Elamri, Teun de Planque, "Automated Neural Image Caption Generator for Visually Impaired People", Stanford University, 2017
- [5] Nuzhat Atiqua Nafis and Md. Safaet Hossain, "Speech to Text Conversion in Real-time", International Journal of Innovation and Scientific Research, Vol. 17 No. 2 Aug. 2015, pp. 271-277
- [6]pypi.python<https://pypi.python.org/pypi/SpeechRecognition/> accessed on 19/7/2017
- [7]pythonprogramminglanguage
<https://pythonprogramminglanguage.com/text-to-speech/>
accessed on 19/7/2017
- [8] Google Drive, Flickr 30k dataset
https://drive.google.com/file/d/0B_PL6p-5reUAZEM4MmRQQ2VVSlk/view accessed on 18/7/2017
- [9] Yin, X., Zuo, Z., Tian, S., Liu, C.: Text detection, tracking and recognition in video: a comprehensive survey. IEEE Transaction, Image Processing, 2016
- [10]Jaderberg, M., Vedaldi, A., Zisserman, A.: Deep features for text spotting. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 512–528. Springer, Heidelberg, 2014
- [11]Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons" 2013 IEEE.
- [12]Bien, Z.; Park, K.-H.; Jung, J.-W.; and Do, J.-H. 2005. Intention reading is essential in human-friendly interfaces for the elderly and the handicapped. IEEE Transactions on Industrial Electronics 52(6):1500–1505.

A Comparative Study of Load Balancing Using SDN and Traditional System

VikasKukreja

*Department of Computer
Engineering,*

*Vivekanand Education Society
Institute of Technology,
Chembur, India
vikas.kukreja@ves.ac.in*

Dinesh Panchi

*Department of Computer
Engineering,*

*Vivekanand Education Society
Institute of Technology,
Chembur, India
dinesh.panchi@ves.ac.in*

JatinSajnani

*Department of Computer
Engineering,*

*Vivekanand Education Society
Institute of Technology,
Chembur, India
jatin.sajnani@ves.ac.in*

Hitesh Seedani

*Department of Computer
Engineering,*

*Vivekanand Education Society
Institute of Technology,
Chembur, India
hitesh.seedani@ves.ac.in*

Mrs.Nupur Giri

*Department of Computer
Engineering,*

*Vivekanand Education Society
Institute of Technology,
Chembur, India
nupur.giri@ves.ac.in*

Abstract--With the growth and advancement in Industrialisation there is increase in the service requirement, due to this the load on the server increases and this may result in latency/delay or sometimes the link failure or server crash. To avoid this latency or link failure or server crash there must be a Load Balancer between the client and server to handle this load. By using SDN, a topology can be created and a Load Balancer Application can be deployed on the Controller that will monitor the load on server, links between the server controller and clients. ODL Controller and Mininet are used to display the topology and simulate the topology created. The Load Balancer application will calculate the threshold value dynamically and will compare the results obtained from both traditional system and SDN using certain parameters. And Plot the graphs for same.

Keywords: SDN, Load Balancing, ODL, Mininet, Link failure, Graphs.

I. INTRODUCTION

A. Introduction

Software-defined networking (SDN) is a networking technology that brings programmability in a network by separating control plane from data plane and deploying a single controller (commonly known as SDN controller) to control flow of data packets among data plans.

SDN Architecture has 3 Layers:

- i. Infrastructure Layer.
- ii. Control Layer.
- iii. Application layer.

Control Layer is the brain of the SDN, controlling of different Data Planes is done by this layer. Applications are installed in

the Application Layer, over the Control Layer and the communication between Controller and Application is done using the NorthBound APIs (Eg. Java APIs or REST APIs) and that of between Controller and Switches using SouthBound APIs (Eg. OpenFlow Protocol). If there are more than one Controller than the Controllers can communicate with each other using EastBound and WestBound APIs.

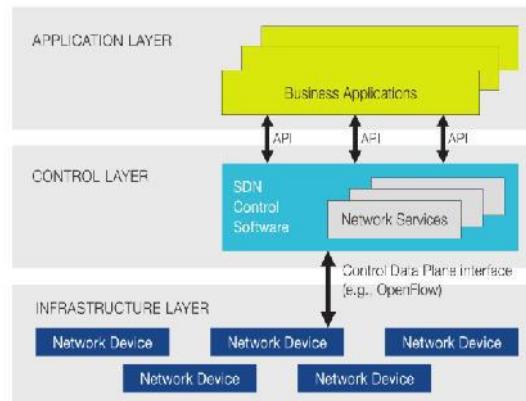


Fig 1. Architecture of SDN

B. Motivation for the project

Software defined networking (SDN) aims to simplify network management by removing the control plane from switches and running custom control applications at a logically central controller. It is a challenging task to code an application which can detect loops or black holes in network. Reliability, Consistency and High-quality products are provided by SDN to all users.

Today's IT leaders are focused on getting more benefits and improving the business agility while lowering the cost of running IT technology. Achieving this is challenging task as IT environment has become complex. SDN technology can help the IT leaders to control cost and boost worker productivity.

With traditional networks, each location requires dedicated devices to provide the services required. For this each location should have their own set of devices for eg router, Firewall etc. SDN can run multiple services on a single hardware platform. This will reduce the consumption of time and hardware as well as operational costs.

C. Problem Definition

With the growth and improvement in Industrialisation there is increase in the service requirement due to this the load on the server increases and this may result in server crashes or link failure.

To avoid this latency or link failure or server crash this paper will install a load balancer between the client and server to handle this load.

By using SDN a topology is created and Load Balancer application is installed on the Controller that will monitor the load on server as well as the links between the server and controller and controller and clients. Load Balancer Application calculates Threshold value dynamically using 2-Threshold policy to transfer processes between the nodes, and Altruistic Priority assignment policy is used in Under loaded Conditions to prioritize the clash between Local and Remote Process.

II. CONCEPT AND METHODOLOGY

A. Concept

Software-defined networking (SDN) is a networking technology that brings programmability in a network by separating control plane from data plane and deploying a single controller (commonly known as SDN controller) to control flow of data packets among data planes. By using SDN, this paper creates a custom topology and installed a Load Balancer application on the Controller that will monitor the load on server and also on the links between the server, controller and clients. Load Balancer Application uses 2-Threshold policy to transfer processes between the nodes where Threshold is calculated Dynamically and Altruistic Priority assignment policy for migration of processes from one node to another or transferring via some other links. This paper provides a solution for controlling the load in this era of increasing industrialization.

Mininet is a tool which is used as network simulator for openflow and SDN applications. It is a software program which allows an entire network consisting of virtual hosts, controllers, switches, and links to be created and emulated on single PC. Mininet uses lightweight virtualization. Minnet connects hosts and switches using virtual ethernet pairs. New network applications such as firewall, load balancer. Can be developed and tested on Mininet. The same application code can be moved to another infrastructure.

It emulates the network using :

- i. Command Line Interface.
- ii. Interactive User Interface.
- iii. Python Application.

Mininet has few inbuilt topologies:

- i. Simple.
- ii. Linear.
- iii. Tree.
- iv. Minimal.
- v. Custom.

Methodology

1. Run OpenDaylight SDN Controller using command:
`.karaf`
2. Run the custom topology using Mininet:
`sudomn --custom topology.py --topomytopo --controller=remote,ip=127.0.0.1,port=6653`
3. Run the pingall command to ensure all hosts are up and running.
4. Check the GUI of OpenDaylight using link:
`localhost:8181/index.html#/topology`
5. Ping any two hosts to get real simulation of data transmission between nodes.
6. Run the load balancing application using command:
`python odl.py`
7. Python file will run the application for load balancing between client and server, application uses Altruistic Priority Assignment Policy and 2-threshold policy for deciding the workload of node.
8. Node can be in any of the following state
 - a. Overloaded
 - b. Normal
 - c. Underloaded
9. Dynamically threshold value is calculated and the load balancing application will handle the load based on the node states described in the previous step.

III. IMPLEMENTATION

1. **Creation of Topology :** This paper had considered 12 hosts named as Host 1 to Host 2 and 15 switches named as switch 1 to switch 15 which are connected as shown in Fig 2 below.

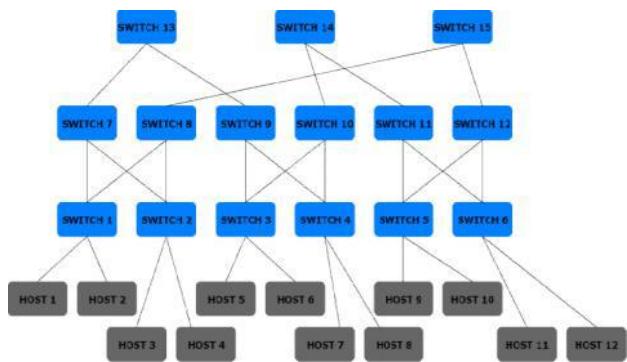


Fig 2 . Topology.

- 2. Finding the Paths:** find all the possible paths from destination to source along with load on each path. There can be more than one path to reach the destination from source.
- 3. Selecting the Shortest path:** Out of all the paths select the one with the least load. Packets will be transferred through this path from source to destination.
- 4. Dynamically Change Of Path :** As the load on paths depends upon the topology created and the packets present in the topology. As the time flies and number of packets increases it might be possible that path selected as shortest may have more load than the alternative path which was rejected first, It will automatically switch to the path with least load.

IV. PARAMETERS

- 1. Overload Rejection:** In situations where node is Overloaded, certain measures need to be taken in order to balance the links. Overload rejection measures are stopped when node which is overload situation terminates and after a short interval Load Balancing is also closed down.
- 2. Process Migration :** This factor decides when does a system should transfer a process based on threshold value used in algorithms.
- 3. Reliability:** In case of failure, this factor determines the reliability of Static and Dynamic Algorithms. Static load balancing algorithms distribute the workload in the start based on their performance evaluation and therefore it can not transfer the workload to other machine in case of failure occurs at run time. While dynamic algorithms distribute the workload at run time and therefore it can balance the load in case of system failure.
- 4. 2-Threshold Policy:** The two threshold values (upper and lower) of various parameters (CPU Queue Length, CPU Utilization and Memory Utilization) are computed using multiplication of average load of every parameter and a constant value.

$$\begin{aligned} tH &= H * Lavg \\ tL &= L * Lavg \end{aligned}$$

Where, tH is high threshold, tL is low threshold, H and L are constants(calculated based on average load).

5. There are different parameters on which this paper will compare the Load Balancing using Traditional and SDN:

- a. Throughput
- b. Threshold
- c. Number of Users
- d. Process Migration
- e. Response Time
- f. Waiting Time
- g. Robustness
- h. Scalability

The graph of above parameters can be plot for different loads on a single node.

V. PROPOSED GRAPH

1. Load vs Throughput for Threshold parameter.
2. Number of Users vs Throughput for Number of Users.
3. Load vs Number of Users for Scalability and Robustness.
4. Load vs Threshold for Process Migration.

VI. CONCLUSION

SDN has brought about a revolution in the field of Networking because now the brain of the devices i.e., the Control Plane is separated from Data Plane. Flexibility is achieved in the Networks as Developers can now develop and deploy their own Applications on the Controller. The reason SDN is called Future of Networking is because SDN provides lots of functionalities to the Developers and Network Administrators. Even intelligent applications can be programmed and deployed on the Controller to maintain the network intelligent which was not possible in traditional networks.

Due to very small topology the results here may not seem promising but as the network grows, SDN proves to be the best solution.

Load Balancing using SDN is the most efficient way to monitor and balance the load between Clients and Servers. The overhead now remains with the Developers is to develop an efficient Application as per the requirements and deploy that application on the Controller and the Application will perform its function.

REFERENCES

- [1] Mr. Marlon Esteban Olaya, Mr. Ivan Bernal, Mr. David Mejia, "Application of Load Balancing in SDN", [10.1109/EATIS.2016.7520102](https://doi.org/10.1109/EATIS.2016.7520102).
- [2] Mr. Mohammad Mousa, Mr. Ayman .Bahaa-Eldin, Mr. Mohamed Sobh, "Software Defined Networking concepts and challenges", [10.1109/ICCES.2016.7821979](https://doi.org/10.1109/ICCES.2016.7821979).
- [3] Mr. Alexander Gelberger, Mr. Niv Yemini, Mr. Ran Giladi, "Performance Analysis of Software Defined Networking(SDN)", [10.1109/MASCOTS.2013.58](https://doi.org/10.1109/MASCOTS.2013.58).
- [4] Mr. G. Araniti, Mr. J. Cosmas, Mr. A. Lera, "OpenFlow over wireless networks: Performance analysis", [10.1109/BMSB.2014.6873559](https://doi.org/10.1109/BMSB.2014.6873559).

Web Application on Result Analysis

Kirtan Patel

Dept. of Information Technology

*Thakur College of Engineering and Technology
Mumbai, India*

Rishab Tiwari

Dept. of Information Technology

*Thakur College of Engineering and Technology
Mumbai, India*

Maulik Shah

Dept. of Information Technology

*Thakur College of Engineering and Technology Mumbai,
India*

Ashish Soni

Dept. of Information Technology

*Thakur College of Engineering and Technology Mumbai,
India*

Abstract-The goal of this project is to create web application for computer science and engineering department; purpose of this is to manage the “result analysis and time table generation, progress report, placement”. It is helpful for students as well as the college authorities. In the current system all the activities are done manually. It is very time consuming and costly. This web application gives authorization to the staff to view their result details, academic details, allotted subject and Time Table. The project is more user-friendly & easily operated so that anyone have minimum computer knowledge can approach and handle the software without having any convoluted. The commonly used tools in the system are Data Flow Diagram, ER Diagram, Block Diagram, Flow chart. Training, experience and common sense are required for collection of relevant information needed to develop the system.

Keywords- *Data Mining, Natural language Processing, Cobweb Algorithm.*

I.INTRODUCTION

This is a tracking application where at a time we can see all the year’s result in a single sheet and a choice of viewing the candidates result separately is also possible. The proposed work is defined by the work to be done in the project and also the ways to complete the project successfully. It will help give clarity on what the project is going to achieve. The structure of the project is defined in this section. The flow of the processes that will be followed in order to produce successful results or outcomes is also listed in this section for better understanding of the project. Also states the probable solutions that could be achieved by the project. The main objective of the project is to provide the examination result to the student in a simple way.

This project is useful for students and institutions for getting the results in simple manner. By a result analyzer with subject status and marks is an application tool for displaying the results in secure way.

The system is intended for the student. And the privileges that are provided to student are to read and execute his/her result by providing user name and password for secure login and in case of new student the registration is available. And the guest user has the privilege only to read. The whole result analyzer will be under the control of the administrator and the admin as the full privileges to read, write and execute the

Result. And admin gives the privileges to the Teacher and student and the guest user to access the result.

II. IMPORTANCE

In this project students’ performance is evaluated according to how they perform academically. The main importance is that it becomes easy for the faculties to list out the students list according to their performance. Using this web application, the students with the lower grade can be sorted easily so that more attention can be put on that particular list of the students.

III. OVERVIEW

Student report generator is a web based application designed and engineered for colleges that need to manage results across multiple branches students that need to track, manage and report results. This application can run on any kind of operating system. At a time we can see all the years result in a single sheet and we can see the individual candidate’s results separately.

IV. PROBLEM DEFINITION

The traditional approach of maintaining the data of student is done manually and this work is too tedious and time consuming. As a human the possibility of entering wrong information by a single person is also high. Hence the Analysis of student is not accurate as per his criteria such as marks, attendance, Term work etc. This website bridges the gaps of all the previous proposed software like Outlier analysis of data according to Student record [1] , Human Friendly results is obtained with the help of NLP [Ref 2] and K-means algorithm help to analysis the data in an efficient way [Ref 4].

Phase 1(Planning, Analysis, Design, Coding, Implementation):
-

1. Perform literature survey and identify all the research gaps, learn about the marks obtained by individual in each semester.
2. Design the database and different modules of the Project.
3. Integration of data from all the tables of the database into a data warehouse.
4. Create code for Analysis of data and generate a pie chart and bar Graph according to marks obtained by students
5. Code to create a software that fills all the research gaps such as Outlier analysis of data according to Student record, Human Friendly results is obtained With the of NLP and K-means algorithm help to analysis the data in an efficient way.

Phase 2(Testing, Deployment): -

1. Create various cases of test case data to check if the integrated data warehouse functions as desired by the client.
2. Conduct beta testing for identifying any further errors and the identify and improvements and that can be performed.
3. After testing and approval, deploy the proposed system to the college.

V.FEATURES OF THE PROJECT

“Progress report of student and, result analysis” this project is mainly designed to reduce manual work.

In the Progress report of student generates the report according to students roll number, semester, year and other details.

The result analysis is generating the graph according to semester university result of students. User friendly interface. Fast access to database. It consists of modules like student login, faculty login. Providing the result in a pattern like internal marks, external marks, total marks. Providing results in CGPI as well as Percentage.

VI. METHODOLOGY

I. Progress report analysis

Importing data from excel sheet-

The data is give as an input in the form of excel sheet which contains different number of fields. The fields like roll number, percentage of attendance, number of practical's, lectures attended by all student.

We take 3 excel sheet as an input those are:

- 1) Student Details
- 2) Month wise Attendance Details
- 3) Month wise unit test marks Details

Student details are provide information like serial number, roll number, name of student, batch. Month wise attendance details contains information like number of taken lectures, practical's, total attendance, total percentage, status about defaulter, etc. Month wise unit test marks details contain subject wise marks of each roll number.

II. Store data into database in required format-

After getting all data it is stored into database by cropping only required field of the imported data. A structure of imported sheet is totally different then database structure so it select only required field from imported data.

III. Generation of crystal report-

Progress report is created as per required format by using the crystal report tool. When proctor provides the inputs like class, year and proctor name, etc. the data will be fetched from the database and crystal report of student will be generated.

IV. Search data of particular student according to roll number and class-

When there is need to search details of particular student then this facility is useful. We must require to give input as roll

number, class name, and year of studding to finding data within database.

V. Print the progress report on crystal report-Printing the report on print paper is one of the main features of our application. Printing only required data on paper like printing of bank passbook. It means that we are able to print the details of same student, on same paper having result of different semester.

VI. Result analysis:

I. Importing data from excel sheet-

II. We take excel sheet as an input those are:

Subject wise Student marks Details are provide information like serial number, roll number, marks of each subject, and name of student.

III. Storing data into database in required format-

After the getting all data is stored into the database by cropping only required field of the imported data. Whatever fields present into imported data that all as it is not stored into database. A structure of imported sheet is totally different then database structure so it's required to select only required field from imported data.

IV. Generation of semester report for overall performance-

Here we are generating the report of overall performance of student for all subjects by considering there mark. We are maintaining all details according to semester. When user needed to find any record then there is need to get input as semester, year, and class.

Time Table Generation:

I. Taking data from user-

- a. Subject Details
- b. Subject teacher
- c. Batch
- d. Class rooms/Lab
- e. Time Details

All input is providing as per semester and year for all classes.

II. Storing data into database in required format-After getting all data it is stored into database by insert only required field of the inputted data. Whatever fields enters into input data that all as it is stored into database.

III. Generate time table-

After getting all the details an algorithm automatically generate time table. It does not contain any overlapping of lectures or practical.

IV. Make the changes in time table as per need-

If any faculty wants to changes in timetable then make changes into editable slots of timetable. It should not any overlapping of lectures or practical.

VII.FEASIBILITY STUDY

Result analysis is a web based application designed for colleges that need to manage results across multiple branches students

that need to track, manage and report results. The project is analyzed first by the project team and after the analysis phase, planning phase is designed. In this phase, the layout of the project is designed and also the working flow. Project planning includes Costing, budgeting and scheduling of the project.

- **Technical Feasibility** – To determine whether the proposed system is technically feasible, we should take into consideration the technical issues involved behind the system. If the system is not comprehensive which supports many

Platforms to work with that goes to proposed system is technically feasible.

- **Economic Feasibility** – To decide whether a project is economically feasible, we have to consider various factors as:
 - I. Cost benefit analysis
 - II. Long-term returns
 - III. Maintenance costs.

- **Operational Feasibility** - To determine the operational feasibility of the system we should take into consideration the awareness level of the users. This system is operational feasible since the users are familiar with the technologies and hence there is no need to make changes to the system. Also the system is very friendly and to use.

Table 1: Project Planning Table

Organizational Impact	Value	Metric	Time Frame
Strategic	Increases the amount of analyzing large amount of student data	Increase the amount of data compiled by 40%.	12 months
Customer	Reduce the result analyzing time	Reduce average time between evaluation of papers and display of results	6 months
Financial	Reduce personnel & communication cost of scheduling.	Reduce personnel & communication cost of analyzing by 75%	
Operational	Reduce analyzing cycle time.	Reduce analyzing time.	12 months
Social	Increase the number of institutes accessing this service.	Increase the Number of institutes Accessing this service from 10 to 50	24 months

The flowchart contains the following modules:

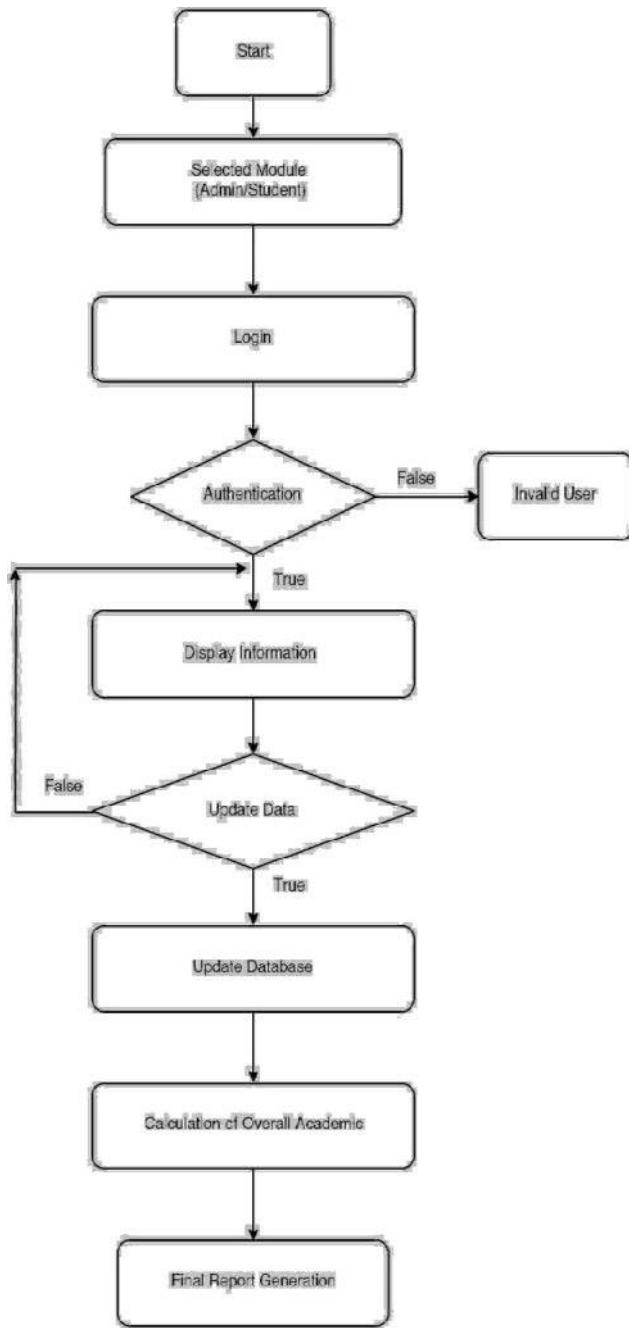


Fig. 1 Flow Chart

ADMIN:

The admin is responsible for uploading the sheets' generated. He can add the subject marks and make any kind of modifications like storing, updating, deleting. Apart from the regular results, the supplementary results can also be uploaded by the admin.

STUDENT:

Student must be an authenticated user of the college to access this application. The student can view the semester marks, individual subject marks and aggregate till the current semester. The student can also make a comparative analysis with the results generated.

FACULTY:

Faculty must also be authorized users of the college. Faculty can view the class result, subject wise result. The percentages of students pass/fail in a subject can be viewed. The total marks (internal, external and both) can be calculated.

VIII. RESULTS

The computer software application is required to be independent of any platform; It is designed to have three main sections, namely: the login window, the main menu and sub menu. The login window requests a valid user name and password from the Administrator to be able to gain access into the software.

The Administrator is any staff that is authorize by the management of the school to be in charge of exams and records unit (e,g the Director of studies), hence he should have a valid user name and password created by him to be able to login to the software.

(a)The Login:

The login form/window enables the user to enter his \ her password. If the password entered is valid, the software will then display the main menu. But if the password is not valid the user will be denied access to the program.

(b)The Main menu:

The main menu contains dashboard and panel, on the dashboard sub menu which serves as a link to all other modules in the program, the sub menu are; Home, Results, Register, profile, supports and contact us. While the panel holds the forms for capturing data and display query result.

(c) Register Sub Menu:

This sub-menu is use for capturing fresh students" data or record into the data base such as student's surname, first Name, other name, class, e-mail, date of birth, sex and admission year/number. It handles students Registration at the beginning of the first term especially because it's the term in which fresh students are registered into the school.

The student"s personal information are all captured in this menu.

VIII. RESULTS

The computer software application is required to be independent of any platform; It is designed to have three main sections, namely: the login window, the main menu and sub menu. The login window requests a valid user name and password from the Administrator to be able to gain access into the software.

The Administrator is any staff that is authorize by the management of the school to be in charge of exams and records unit (e,g the Director of studies), hence he should have a valid user name and password created by him to be able to login to the software.

(a)The Login:

The login form/window enables the user to enter his \ her password. If the password entered is valid, the software will then display the main menu. But if the password is not valid the user will be denied access to the program.

(b)The Main menu:

The main menu contains dashboard and panel, on the dashboard sub menu which serves as a link to all other modules in the program, the sub menu are; Home, Results, Register, profile, supports and contact us. While the panel holds the forms for capturing data and display query result.

(c) Register Sub Menu:

This sub-menu is use for capturing fresh students" data or record into the data base such as student's surname, first Name, other name, class, e-mail, date of birth, sex and admission year/number. It handles students Registration at the beginning of the first term especially because it's the term in which fresh students are registered into the school.

The student"s personal information are all captured in this menu.

(d) The Results Sub Menu:

This sub menu contains modules such as register subject, view subject, view students, compute results, and view results. Clicking on any of these modules, operation takes place.

(e) Register subject module:

All subjects offered by students in the school are captured and recorded into the data base in this module.

(f) View subject module:

All subjects registered into the data base can be viewed in this module.

(g) View student's module:

All registered students in the database are viewed in this module

(h) Compute results module:

This module allows the students" Scores to be captured for the purpose of results computation. In fact this is the Modules that handle the result processing.

(i) View results module:

This module displays each student result or report sheet as the case may be, the result sheet can be printed from this module as well.

(j) Edit module:

This module allows the already entered data about students to be modified in case of errors or updates.

IX. CONCLUSION

Our main aim through this project is to reduce manual work, to provide efficient way of handling data. The progress report „module our project reduces the work of staff of making entries for each student manually. Result analysis generate graph of students marks and timetable generation includes by taking only the input it will automatically generate timetable. In future our project can be done as mobile application. So that student can see their progress report on their mobile.

X. SCOPE OF THE PROJECT

The aim of this project is to create software for computer science and engineering department; purpose of this is to maintain the “PROGRESS REPORT, RESULT ANALYSIS.

The system has an inbuilt feature for adding the details of the students by themselves. So the overhead of the college authorities and the teachers become less.

Acknowledgment

We sincerely thank to our guide Mr. Aaditya Desai, our HOD Dr. Rajesh S. Bansode, our Dean Dr. Kamal Shah and our principle Dr. B. K. Mishra for his/her guidance and support for carrying out our project work.

REFERENCES

- [1] C. Romero S. Ventura P. G. Espejo and C. Hervás "Data mining algorithms to classify students ". In EDM 2008 pp. 8-17.
- [2] Y. Zhang S. Ousseina T. Clark and H. Kim "Use data mining to improve student retention in higher education-A case study=" In ICEIS (1) 2010 pp. 190-197.
- [3]B.Namratha, Niteesha Sharma "Educational Data Mining – Applications and Techniques". In IJLTET Vol 7 issue 2 July 2016.
- [4]Z. Erdogan and M. Timor "A Data Mining Application in a Student Database " in Journal of Aeronautics and Space Technologies 2005 vol. 2 No. 2 2005 pp.

Digital India and Its Impact

Sandeep Yadav
Thakur College of Engineering
Mumbai – 400101

Sangeeta Vhatkar
Thakur College of Engineering
Mumbai - 400101

Abstract- It is a widely known indisputable fact that digital Bharat is that the outcome of the many innovations and technological advancements. These remodel the lives of individuals in many ways and can empower the society in a very better manner. The Digital Bharat Programme, associate degree initiative of honorable Prime Minister man. Narendra Modi can transpire new development in each sector. The motive behind the thought is to build participative, clear and responsive system. The Digital Bharat drive may be a dream project of the Bharat Government to rework India into a knowledgeable economy and digitally authorized society, with sensible governance for voters by transfer synchronization and co-ordination publicly accountability, digitally connecting and delivering the govt. Programs and services to mobilize the capability of knowledge technology across government departments. Today, each nation desires to be absolutely digitalized and this programmed strives to supply equal profit to the user and repair provider. Hence, a trial has been created during this paper to know Digital Bharat – as a campaign where technologies and property can close to create a bearing on all aspects of governance and improve the standard of lifetime of voters.

Keywords-Development, Digital, E-governance, Government, Internet access

I. INTRODUCTION

Digital technologies, which consist of Cloud Computing and cell programs, have emerged as catalysts for brief financial boom and citizen empowerment throughout the globe. Virtual technologies are being an increasing number of utilized by us in regular lives, from retail shops to authorities offices. They assist us to hook up with each other and also percentage statistics on troubles and worries faced with the aid of us. Honorable high Minister envisions transforming our nation and growing possibilities for all citizens by harnessing virtual technologies. His imaginative and prescient is to empower each citizen with access to

virtual services, information, and information. Digital India is the subsequent large issue that India is witnessing. It aims at profoundly touching the lives of anybody with the transformation traveling the paths of each rural and concrete India. Nowadays, the sector has transformed from a know-how savvy to techno expertise savvy. Consider something and it's far available in one click on. So, virtual India is a step with the aid of the government to inspire and join Indian financial system to one of these expertise savvy worlds. This system targets to make government services available to humans digitally and experience the benefit of the latest facts and technological improvements. It brings out numerous schemes

like E-fitness, digital Locker, E-signal, education etc. And national scholarship portal. The program strives to offer an identical advantage to the user and provider. The clients can be benefited by way of saving time, cash, physical & cognitive energy spent in prolonged government procedures. The digital India Programme has been released with an intention of remodeling us into a digitally empowered society and know-how economy. The digital India would make sure that authorities services are to be had to residents electronically. It might additionally convey in public duty through the mandated delivery of presidency's offerings electronically; a completely unique identity and e-Pramaan based on authentic and fashionable primarily based interoperable and incorporated government applications and information foundation [1].

Digital India is a dream venture of the government for the citizens and Industries of India which could help in connecting the diverse past and present projects to bring India to a global platform. Through this mission, government offerings are available for city and rural citizens digitally or electronically. It will help to gain virtual innovation and create positive impact on the human beings residing in rural and urban areas. It will attract investment in all product production industries. Andhra Pradesh is the first state to have opted this implementation. The digital India assignment pursuits to transform our country into a digital financial system with participation from rural or urban citizens and companies. This could make certain that each one government services and data are to be had everywhere, anytime, on any tool that is clean-to-use, noticeably available and secured. Virtual India undertaking eliminates virtual hole among the rural and concrete India.

II. VISION OF DIGITAL INDIA INITIATIVE



Fig 1: Vision of Digital India

Here is what the government of India aims to achieve through Digital India initiative.

A. Infrastructure:

The digital India initiative has an imaginative and prescient to provide excessive speed net offerings to its citizens in all Gram Panchayats. Financial institution money owed could be given priority at man or woman stage. Human beings might be provided with safe and comfortable cyberspace within the united states [2].

B. Governance and services:

Government offerings will be available online where citizens might be ensured smooth get entry to it. Transactions may be made easy thru digital medium.

C. Digital empowerment of citizens:

One of the maximum crucial factors of the virtual India initiative to offer typical digital literacy and make virtual sources without problems available. The services also are supplied in Indian languages for lively participation.

III. NINE PILLARS OF DIGITAL INDIA

The idea of digital India is to deliver abroad commitments to invest closer to this initiative so that the Investments might be applied towards making internet devices and smart telephones at a low-cost [2]. The fee in India which would help in generating jobs, reduction inside the time and cost of uploading them From abroad. Digital India comprises of various projects below the single programme each targeted To put together India as a knowledge economic system and for bringing properly governance to residents via Synchronized and coordinated engagement of the whole authorities. Nine initiatives were Undertaken which might be proven inside the diagram given beneath:

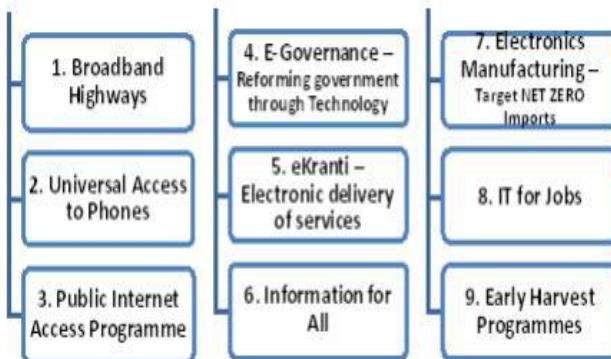


Fig 2: Nine Pillars of Digital India

A. Highways to have broadband services:

The first step is to provide excessive speed broadband highways thru fiber optics that join all of the remote regions, authorities departments, universities, studies and development etc. Internet based portals and cellular apps could be advanced to get entry to online records while on the pass.

B. Easy access to mobile connectivity:

The authorities are taking steps to ensure that via 2018 all villages are covered via cell connectivity. The purpose is to boom

community penetration and cowl gaps in all 44,000 villages. The widespread public will get entry to the online authorities offerings with the help of handheld devices. The state is prepared to be well-connected, efficient, and greater efficient in each component.

C. IT Training for Jobs:

The government is preparing to provide education and coaching abilities to the young people for employment possibilities in the IT quarter [3]. BPO industries will be mounted for the quickest developing section of the records technology Enabled services enterprise. It offers eservices 24/7 in every subject and gives extra jobs potentials. This initiative seeks to teach 10 million humans in cities and villages for IT area jobs in five years. It also aims to train zero.3 million agents to run feasible agencies delivering IT services.

D. Manufacturing of electronics:

This milestone will create a big base for electronics manufacturing in India with the useful resource of digital technologies and abilities. goal net 0 Imports is a placing demonstration of reason. This ambitious intention requires coordinated motion on many fronts.

E. Provide public access to internet:

Virtuous technology that supports fee containment, collaboration, protection, offerings-on-the-pass, social-connect, and in-built intelligence that supply remote get admission to any information or carrier available across the domain [4]. this transformation will open new doorways of e-offerings to every citizen. The authorities' goals to provide network services to two & half lakh villages which comprise of 1 in every Panchayat with the aid of March 2017 and 1.5 lakh post places of work within the subsequent two years. these post workplaces will become Multi-provider centers for the human beings.

F. E-Governance:

This governance will transform every manual work into fully automation system.

It will revolutionize the system in the following ways:

- Online access to applications i.e. availability of all databases and information in electronic format.
- Effortlessly tracking of assignments.
- Interface between departments for superior production of work.
- Quickly respond, analyze and resolve persistent problems and many more.

G. E-Kraanti:

This Kraanti will absolutely consciousness on digital know-how software wherein training, fitness, farming, rights, monetary and many greater offerings can be brought on a very high bandwidth [5]. Physical barriers now not are a quandary whilst nearly every person and the entirety are a digital handshake away.

H. Global Information:

Hosting data online and engaging social media platforms for governance is the aim of the government. Information is also easily available for the citizens. Mygov.in is a website launched by the government for a 2-way communication between citizens and the government. People can send in their suggestions and comment on various issues raised by the government, like net neutrality.

I. Early harvest programs:

This programme will generate quick timeline tasks in which every guide service is altered via e-service. E-services like:

- Implementation of wireless in all the universities.
- Public wi-fi sports to get right of entry to online records.
- educational books to e-books.
- human beings will use the e-offerings for leisure, climate data, brand new updates etc.
- substitute of manual attendance to Biometric process

IV. INITIATIVES UNDER DIGITAL INDIA

Some of the facilities provided under the initiative of Digital India are as follows:

A. Digi Locker

Digital Locker facility allows citizens to digitally keep their great files like PAN card, passport, mark sheets and diploma certificate. Digital Locker offers secured get right of entry to authorities issued files. It uses authenticity offerings supplied through Aadhaar. It's miles focused on disposing of the use of physical documents and allows sharing of verified electronic files throughout government businesses. Three key stakeholders of digilocker are Citizen, provider, and the requester.

B. Attendance.gov.in

Attendance.gov.in is a internet site, released with the aid of PM Narendra Modi to maintain a document of the attendance of government employees on a real-time foundation. This initiative started with the implementation of a not unusual Biometric Attendance gadget (BAS) within the central government offices located in Delhi.

C. MyGov.in

Mygov.in is a platform to percentage inputs and ideas on matters of coverage and governance. It's miles a platform for citizen engagement in governance, through a "speak", "Do" and "Disseminate" approach.

D. SBM Mobile app

Swachh Bharat Mission (SBM) Mobile app is being used by people and Government organizations for achieving the goals of Swachh Bharat Mission.

E. e-Hospital

The e-Hospital application provides important services such as online registration, payment of fees and appointment, online diagnostic reports, enquiring availability of blood online etc.

F. National Scholarship Portal

Country wide Scholarship Portal is a one-step solution for quit to cease scholarship technique right from submission of student software, verification, sanction and disbursal to end beneficiary for all of the scholarships furnished by way of the government of India [6].

G. E-Sign framework

e-Sign framework allows citizens to digitally sign a document online using Aadhaar authentication.

V. PROPOSED IMPACT OF DIGITAL INDIA

A. Economic impact

In keeping with analysts, the virtual India plan ought to raise GDP as much as \$1 trillion by using 2025. It may play A key role in macroeconomic elements along with GDP increase, employment technology, labor Productivity, growth in a wide variety of agencies and revenue leakages for the government. As in step with the world financial institution file, a ten% increase in mobile and broadband penetration increases the According to capita GDP via 0.eighty one% and 1.38% respectively inside the growing international locations. India is the second Biggest telecom market within the international with 915 million wireless subscribers and global's 3rd biggest Internet marketplace with nearly 259 million broadband customers. There's still a massive financial Opportunity in India as the teledensity in rural India is simplest 45% in which more than 65 % of the Populace lives [7]. Destiny increase of telecommunication industry in terms of number of subscribers is Expected to come back from rural regions as city regions are saturated with a teledensity of extra than a160%.

B. Social impact

Social sectors consisting of training, healthcare, and banking are unable to reach out to the citizens due to obstructions and barriers such as intermediary, illiteracy, lack of knowledge, poverty,loss of inances,information and investments. Those demanding situations have led to an imbalanced increase in the rural and concrete areas with marked differences within the economic and social repute of the humans in those regions. Modern information and Communications generation (ICT) makes it easier for humans to reap get admission to offerings and resources. The penetration of cellular devices can be quite beneficial as a complementary channel to public provider delivery apart from the advent of absolutely new services which can also have a big impact on the excellent of life of the customers and lead to social modernization. The poor literacy rate in India is due to unavailability of bodily infrastructure in rural and remote areas. That is wherein m-schooling services can play an important function by using achieving faraway hundreds. In line with estimates, the virtual literacy in India is just 6.5% and the net penetration is 20.83 out of one hundred populations. The digital India challenge can be useful in presenting real-time training and partly cope with the project of loss of instructors in training device through clever and digital school rooms. Education to farmers, fisher guys can be supplied via cell devices. The excessive speed community can offer the adequate infrastructure for online training systems like huge Open online courses. Cell and net banking can enhance the economic inclusion in the country and may create a win-win scenario for all events inside the value-chain by way of growing an interoperable environment and revenue sharing commercial enterprise models. Telecom operators get extra sales streams at the same time as the banks can attain new consumer corporations incurring lowest feasible expenses [8].Factors inclusive of a burgeoning populace, negative doctor-affected person ratio (1:870), high toddler mortality charge, increasing existence expectancy, fewer excellent physicians and a majority of the populace residing in far-off villages, aid and justify the need for tele medication within the country. M-fitness can sell innovation and beautify the attain of healthcare offerings. Digital platforms can help farmers in expertise (crop desire, seed range),

context (climate, plant safety, cultivation pleasant practices) and marketplace statistics (marketplace expenses, marketplace call for, logistics).

C. Environmental impact

The essential modifications in the technology area will now not simplest brought modifications to the financial gadget, however, will also make contributions to the environmental adjustments. The following technology technologies will assist in Reducing the carbon footprint through lowering fuel intake, waste management, greener workplaces and for that reason leading to a greener environment. The ICT (information and Communications era) sector allows inefficient management and utilization of scarce and non-renewable sources. Cloud computing era minimizes carbon emissions through improving mobility and flexibility. The Strength intake can be decreased from 201. 8 terawatt hour (twh) in 2010 to 139.8 twh in 2020 by higher adoption of cloud data facilities inflicting a 28% discount on carbon footprint from 2010 ranges.

D. Governance impact

The final vicinity wherein we analyzed the effect of digitization turned into government effectiveness. As for the evaluation reviewed above, we relied on three metrics: the transparency of governmental activities, the transport of e-government services, and the provisioning of public schooling a key authorities carrier. Our correlational analysis demonstrates that extra digitization enables a society to be extra transparent, growing public participation and the government's potential to disseminate information in a handy way. Virtual era gives the population more perception into government guidelines and features a perception that would, in flip, lead to more energetic political participation and aid the development of human rights. Additionally, as anticipated, e-government offerings are greater powers in a digitized surrounding. Contemporary research suggests that causality in this situation acts both methods. Higher digitization contributes to extra green transport of e-authorities services, while higher e-government services stimulate a boom in digitization. In the end, digitization helps the better transport of primary authorities services, along with public education. Usual, our analysis indicates that digitization simply has a nice effect on monetary development, societal nicely-being, and authorities effectiveness, despite the fact that this effect varies in line with a rustic's level of digitization. Digitization has an increasing impact at the economy and first-rate of lifestyles as countries enhance via the degrees of digitization, and more impact on access to fundamental services and schooling in international locations that are just starting their adventure.

VI. CHALLENGES FOR DIGITAL INDIA

The primary challenge is to establish popular performance indicators to the degree the extent to which ICT is being assimilated in societies. All through most of the sector's improvement, ICT stakeholders targeted mostly on access, constructing the networks that today join plenty of the planet; they devised metrics, therefore. In a world of close to ubiquity in terms of getting right of entry to, policymakers want a brand new way to observe the ICT area. The second assignment issues the lack of gear to decide the effect that the mass adoption of connected virtual technologies and applications is having on societies and economies. With sensible, dependable equipment to the degree the

advantages of digitization, governments could probably be extra formidable in developing and investing in the ICT region. The 1/3 mission is for policymakers to undertake new coverage gear to accelerate digitization and acquire its accompanying benefits. In this paper we laid emphasis on the need to gain a similar know-how of the ways wherein they can encourage adoption and boost the use of virtual packages by means of purchasers, businesses, and public institutions.

VII. OBSTACLES IN THE PROCESS OF DIGITALIZATION

1. The resistance to change people show during emergence of new technology.
2. Building trust among the people for change is difficult.
3. Lack of knowledge about its use and benefits.
4. Infrastructure requirements and their unavailability also hamper the reaching of these basic facilities to the people.
5. The basic hindrance is the lack of literacy and literacy is not just to be able to write ones name and do the signature but to have an understanding of the changes going in the technology, society and the country for their own betterment.

A. Jio helps to digitize rural India

The digital India initiative is one of the most ambitious plan of the Indian government and each and every sector should make some contribution in the process of achieving this goal. One such initiative or the opportunity seen by the Reliance Industries in this perspective is their Reliance Jio Infocom Limited (RJIL) subsidiary which has set a tough benchmark for its competitors. Tarun Pathak, senior analyst, Mobile Devices and Ecosystems, Counterpoint Research, told IANS "To make the digital revolution a success in our country, especially for millions of rural Indians, data mining is the future. I feel Jio is targeting nearly 450 million feature phone users who will make a shift to smartphones in the near future". Experts say that Jio's move will increase data usage across the board. Reliance Jio has partnered with several smartphone brands where Jio Preview offer is available for a period of 90 days. Jio's 4G-LTE services include unlimited HD voice calls and video calls, unlimited SMS, unlimited high-speed data and a host of Jio Premium apps. According to Prashant Singhal, Global Telecommunication leader, Ernst and Young "For a price sensitive market like India, launch of affordable data services and free voice calls is indeed a welcome step [9]. This is expected to drive greater data adoption across segments". He also added that "Telcos form the backbone of 'Digital India' and need to make the higher investments in data network. Any market erosion, at this stage, may impact the outcome of the upcoming Rs.800 billion spectrum auctions critical for the 'Digital India' vision".

Such contributions from every sector can act as catalyst to the digital transformation in India. But along such steps it should be seen that the quality of the service is maintained and does not affects the market negatively.

VIII. CONCLUSION

The digitalization brings innovation, ease of operating, new process possibilities and increase in the financial system. It enables to bring transparency in the gadget and greater obvious is the float of price range within the financial system less is the trouble of tax evasion, parallel economic system and so forth. But with these

types of advantages to being had it additionally makes it important for the human beings to have simple monetary information and a push towards the importance of the economic literacy. With the assist of which they are able to defend their money in conditions like inflation, depression, and understand about distinct monetary services and products to store it for his or her better destiny. Digitalization can also play a crucial function in success this purpose as it can have a greater reach to the humans. By this, we are able to attain on a conclusion that the new technology needs to harness properly and for this, it is not only the provision however additionally the expertise to apply it and get blessings from it.

REFERENCES

- [1] Sobhesh Kumar Agarwalla, Samir K. Barua, Joshy Jacob and Jayanth R. Verma, "Financial literacy among working young in Urban India", W.P. No. 2013-10-02 IIMA.Research and Publication, Oct 2013.hahnazNurulMahdzan, and Tabiani, S. "The Impact of Financial Literacy on Individual Saving: an Exploratory Study in Malaysian Context", Transformation in Business & Economics, Vol. 12, No 1(28), pp.41-55. ISSN 1648-4460,2013.
- [2] Articles in Financial Express, Economic Times, The Hindu.
- [3] economictimes.indiatimes.com/markets/expertview/demonetization-effect-more-and-morepeople-are-goingfor-digital-payment-chandakochhar/articleshow/55495397.cms
- [4] Network Readiness Index 2016 www.weforum.org
- [5] www.business-standard.com/article/news-ians/disruptive-jiooffers-will-help-digitise-rural-indian-experts1160901005534_1.html
- [6] www.digitalindia.gov.in
- [7] www.deity.gov.in
- [8] www.dnaindia.com/
- [9] www.nber.org/papers/w15352.pdf NBER working paper series
- [10] www.scientificamerican.com
- [11] www.strategyand.pwc.com/media/file/Strategyand_Maximising-the-Impact-of-Digitization.pdf

Efficient and Secure Data Dynamic Operations in Cloud Computing

Ms. Shital H. More
*IT Department
Thakur College of
Engineering &
Technology, Mumbai*

Ms. Swati Abhang
*IT Department
Thakur College of
Engineering &
Technology, Mumbai*

Mr. Sudhir Dekhane
*IT Department
Thakur College of
Engineering &
Technology, Mumbai*

Mr. Shridhar Kamble
*IT Department
Thakur College of
Engineering &
Technology, Mumbai*

Abstract- In Cloud Computing our applications and data is moved in large data centers of Cloud Service Provider (CSP). In this case management of data and services may not be trustworthy. This poses a security challenge to the data stored inside cloud. As we know cloud data storage security is an important aspect of quality of service, data storage security is having paramount importance. Here we want to focus on correctness of users data inside a cloud. We are using homomorphic token for distribution of data and for verifying its correctness. It also helps us to find out data localization errors. This approach is efficient for detecting and recovering from malicious data modification attacks.

Keywords-Cloud Computing, Cloud Service Provider, data storage security, Homomorphic token, localization errors

I. INTRODUCTION

Innovations are necessary to ride the unavoidable tide of change. Most of enterprises are striving to reduce their computing cost through the means of virtualization. This demand of reducing the computing cost has led to the innovation of Cloud Computing [1]. Cloud computing is a term used to describe a set of IT services that are provided to a customer over a network on a leased basis and with the ability to scale up or down as per their service requirements. Usually cloud computing services are delivered by a third party provider who owns the infrastructure.

Cloud Computing has become one of the most talked about technologies in recent times and has got lots of attention from media as well as analysts because of the opportunities it is offering Desktop PC Laptops Mobile / PDA Internet Cloud Provider Clients subscribes to the service via internet, & they pays as per usage of service to which they have subscribed [4]. Clients need not to be stationary it can be PDAs, Mobiles Laptops. Clients doesn't need to have in house infrastructure, they can purchase to the service they want on hourly, weekly or monthly reasonable basis. Cloud computing users can

avoid capital expenditure on hardware, software, and services when they pay a provider only for what they use. Consumption is usually billed on a utility or subscription basis with little or no upfront cost.

II. METHODOLOGY

2.1 Cloud Storage Architecture

The general architecture of cloud storage system is illustrated in Figure1 Generally two different network entities can be identified. We have assumed that users have direct peer to peer connection between them & cloud. Users will upload their data to cloud and only they can access it & not any other cloud users. Different network entities are mentioned below:

- User: users, who have data to be stored in the cloud and rely on the cloud for data computation, consist of both individual consumers and organizations.
- Cloud Service Provider (CSP): CSP is who has the capabilities to host data & applications of users. They have huge resources that they can provide dynamically for satisfying various user needs. CSP having expertise in building & managing cloud servers, having their own data centers for hosting user's data [5].

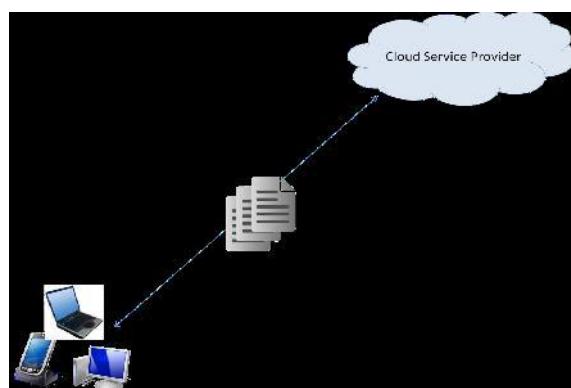


Fig 1. Architecture of cloud data storage

Figure.1 shows how the data is outsourced in cloud and users have no control over it [3]. This also gives perception of the problem with the storage and to ensure the integrity of the data in the cloud. In cloud data storage, a user stores his data through a CSP into a set of cloud servers, which are running in a simultaneous, cooperated and distributed manner. Data redundancy can be employed with technique of erasure-correcting code to further tolerate faults or server crash as users data grows in size and importance. Thereafter, for application purposes, the user interacts with the cloud servers via CSP to access or retrieve his data [8].

In some cases, the user may need to perform operations on his data. The most general forms of these operations we are considering are update, delete, insert and append. As users no longer possess their data locally, it is of critical importance to assure users, that their data are being correctly stored and maintained. That is, users should be equipped with security means so that they can make continuous correctness assurance of their stored data even without the existence of local copies[10]. In case those users do not necessarily have the time, feasibility or resources to monitor their data, users can delegate the tasks to an optional trusted TPA of their respective choices. But users need to pay to the Third Party Auditors for that. This is not our aim, what we want is to give freedom to users to ensure intactness of their data in cloud.

In our scheme, we assume that the point-to-point communication channels between each cloud server and the user is authenticated and reliable. Security threats faced by cloud data storage can come from two different sources. On the one hand, a CSP can be self-interested, un-trusted and possibly malicious. It may also attempt to hide a data loss incident due to management errors, Byzantine failures and so on. On the other hand, there may also exist an economically motivated adversary, who has the capability to compromise a number of cloud data storage servers in different time intervals and subsequently is able to modify or delete user's data while remaining undetected by CSPs for a certain period. So we have attackers with different purposes in different context & we need to classify them as per the severity of damage they can do to storage.

Depending upon various motivations of attackers we have classified them into categories. Specifically, we consider two types of adversary with different levels of capability [17]:

- a) *Weak Adversary*: The adversary is interested in corrupting the users data files stored on individual servers. Once a server is comprised, an adversary can pollute the original data files by modifying or introducing its own fraudulent

data to prevent the original data from being retrieved by the user.

Strong Adversary: This is the worst case scenario, in which we assume that the adversary can compromise all the storage servers so that he can intentionally modify the data files as long as they are internally consistent.

In fact, this is equivalent to the case where all servers are colluding together to hide a data loss or corruption incident. To ensure the security for cloud data storage under the aforementioned adversary model, we aim to design efficient mechanisms for dynamic data verification and operation.

2.2. Securing the Data Storage

In cloud data storage system, users store their data in the cloud and no longer possess the data locally. Thus, the correctness and availability of the data files being stored on the distributed cloud servers must be guaranteed. One of the key issues is to effectively detect any unauthorized data modification & corruption, possibly due to server compromise [16]. Besides, in the distributed case when such inconsistencies are successfully detected, to find which server the data error lies in is also of great significance, since it can be the first step to fast recover the storage errors. Our system operation is divided into four modules,

- File Distribution Preparation
- Token Pre-computation
- Correctness Verification & Error Localization
- File Retrieval & Error Recovery.

1.2.1 Notation & Preliminaries

- F – The data file to be stored. We assume that F can be denoted as a matrix of m equal-sized data sectors, each consisting of l blocks. Data blocks are all well represented as elements in Galois Field $GF(2^w)$ for $w=4,8,16$.
- R – The dispersal matrix used for Reed-Solomon coding.
- D – Data matrix constructed over data vectors.
- C – The encoded file matrix, which includes a set of $n=m+k$ vectors, each consisting of l blocks.
- PRF – Pseudorandom function.
- PRP – Pseudorandom permutation.

1.2.2 File Distribution Preparation

It is well known that erasure-correcting code may be used to tolerate multiple failures in distributed storage systems. In cloud data storage, we rely on this technique to disperse the data file F redundantly across a set of $n=m+k$ distributed servers. R $m+k,k$ Reed-Solomon erasure-correcting code is used to create k redundancy parity

vectors from m data vectors in such a way that the original m data vectors can be reconstructed from any m out of the $m+k$ data and parity vectors. By placing each of the $m+k$ vectors on a different server, the original data file can survive the failure of any k of the $m+k$ servers without any data loss, with a space overhead of k/m . For support of efficient sequential I/O to the original file, our file layout is systematic, i.e., the unmodified m data file vectors are distributed across $m+k$ different servers. We are using Reed Solomon Algorithm to disperse the file redundantly over m storage devices [18].

2.3 Rs-Raid Algorithm

Let there be n storage devices ($D_1, D_2, D_3, \dots, D_n$) each of which holds k bytes. These are called the Data Devices. Let there be m more storage devices ($C_1, C_2, C_3, \dots, C_m$) each of which also holds k bytes. These are called the Checksum Devices. The contents of each checksum device will be calculated from the contents of the data devices. The goal is to define the calculation of each c_i such that if any m of ($D_1, D_2, D_3, \dots, D_n, C_1, C_2, C_3, \dots, C_m$) fail, then the contents of the failed devices can be reconstructed from the non-failed devices. In other words we have n data words $d_1, d_2, d_3, \dots, d_n$ and m checksum words $c_1, c_2, c_3, \dots, c_m$ which are computed from the data words in such a way that the loss of any m words can be tolerated[3].

To compute a checksum word c_i for the checksum device C_i , we apply function P_i to the data words: $c_i = P_i d_1, d_2, \dots, d_n$. If a data word on device D_j is updated from d_j to d'_j , then each checksum word c_i is recomputed by applying a function $G_{i,j}$ such that $c'_i = G_{i,j} d_j, d'_j, c_i$. When up to m devices fail, we reconstruct the system as follows. First, for each failed data device D_j , we construct a function to restore the words in D_j from the words in the non-failed devices. When that is completed, we recomputed any failed checksum devices C_i with P_i . For example, suppose $m=1$. We can describe $n+1$ parity in the above terms. There is one checksum device C_1 . To compute each checksum word c_1 , we take the parity (XOR) of the data words: $c_1 = P_1 d_1, d_2, \dots, d_n = d_1 \oplus d_2 \oplus \dots \oplus d_n$. If a word on data device D_j changes from d_j to d'_j , then c_1 is recalculated from the parity of its old value and the two data words: $c'_1 = G_{1,j} d_j, d'_j, c_1 = c_1 \oplus d_j \oplus d'_j$. If a device D_j fails, then each word may be restored as the parity of the corresponding words on the remaining devices: $d_j = d_1 \oplus d_2 \oplus \dots \oplus d_n \oplus c_1$. In such a way, the system is resilient to any single device failure.

We are given n data words d_1, d_2, \dots, d_n . We define functions P and G which we use to calculate and maintain the checksum words c_1, c_2, \dots, c_m . We then describe how to reconstruct the words of any lost data device when up to m devices fail. Once the data words are reconstructed, the checksum words can be recomputed from the data words and P . Thus, the entire system is reconstructed.

2.3.1 Calculating & Maintaining Checksums

We define each function P_i to be a linear combination of the data words: $c_i = P_i d_1, d_2, \dots, d_n = d_j P_{i,j} \quad n, j=1$. In other words, if we represent the data and checksum words as the vectors D and C , the functions P_i as rows of the matrix P , then the state of the system adheres to the following equation: $PD=C$. We define P to be the $m \times n$ matrix $p_{i,j} = j-1$, and thus the above equation becomes[18]:

$$\begin{matrix} \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,n} \\ p_{2,1} & p_{2,2} & p_{2,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & p_{m,2} & p_{m,n} \end{pmatrix} & \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} & = & \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \end{matrix}$$

When one of the data words d_j changes to d'_j , then each of the checksum words must be changed as well. This can be affected by subtracting out the portion of the checksum word that corresponds to d_j , and adding the required amount for d'_j . Thus, $G_{i,j}$ is defined as follows: $c'_i = G_{i,j} d_j, d'_j, c_i = c_i + p_{i,j}(d'_j - d_j)$. Therefore, the calculation and maintenance of checksum words can be done by simple arithmetic.

2.3.2 Recovering From Failures

To explain recovery from errors, we define the matrix A and the vector E as follows: $A = IP$ and $E = DC$. Then we have the following equation $AD=E$. We can view each device in the system as having a corresponding row of the matrix A and the vector E . When a device fails, we reflect the failure by deleting the devices row from A and from E . What results a new matrix A' and a new vector E' that adhere to the equation: $A'D=E'$. Suppose exactly m devices fail. A' is $n \times n$ matrix. Because matrix P is defined to be a Vandermonde matrix, every subset of n rows of matrix A is guaranteed to be linearly independent. Thus, the matrix

A' is non-singular, and the values of D may be calculated from $A'D=E'$ using Gaussian Elimination. Hence all data devices can be recovered. Once the values of D are obtained, the values of any failed $C1$ may be recomputed from D . It should be obvious that if fewer than m devices fail; the system may be recovered in the same manner, choosing any n rows of A' . Thus system can tolerate any number of device failures up to m [15].

2.3.3. Rs Raid Algorithm Summarized

Given n data devices and m checksum devices, the RS-Raid algorithm for making them fault-tolerant to up to n failures is as follows[12]:

- i) Choose a value of w such that $2w > n+m$. It is easiest to choose $w=4$ or $w=8$ or $w=16$, as words then fall directly on byte boundaries. w =word size .
- ii) Set up the tables $gfllog$ and $gfilog$. These tables are used to perform multiplication over Galois Fields.
- iii) Set up the matrix P to be the $m \times n$ matrix: $pi,j = ji-1$ where multiplication is performed over $GF(2w)$.
- iv) Use the matrix P to calculate and maintain each word of the checksum devices from the words of the data devices. Again, all addition and multiplication is performed over $GF(2w)$. Create the matrix D as actual data matrix & Calculate C by equation $PD=C$.
- v) If any number of devices up to m fails, then they can be restored in the following manner. Choose any n of the remaining devices, and construct A' and a vector E' . Then solve for D in $A'D=E'$. This enables the data devices to be restored. Once the data devices are restored, the failed checksum devices may be recalculated using the matrix F .

So, as per RS Raid algorithm, we divide the input file to the n data vectors, where n is number of storage devices present in the system. The data vectors that are generated are of equal size, so the load will be distributed equally to all the storage devices. We create $m \times n$ matrix D & store all the data vectors in matrix D . In next step we create a Reed Solomon matrix R which is generated over Galois field, i. e. $GF(2w)$. In our case we have assumed word size $w=4$. After this stage, we perform matrix multiplication to generate checksum matrix C . We multiply data matrix D with Reed Solomon matrix R . The resultant matrix is the redundant matrix which contains original data from data matrix D & parity vectors added by Reed Solomon matrix. It means matrix D will be stored redundantly across the different storage devices & it will be used for token computation as well as data recovery purpose.

2.4. Challenge Token Pre-computation

To verify the correctness of users data & to locate the errors, we entirely rely on the pre-computed verification

tokens. These tokens are calculated before file distribution & they are very short. We are computing the tokens by pseudorandom function PRF & pseudorandom permutation function PRP . We pre-computes short verification tokens on individual vector, each token covering a random subset of data blocks. We have assumed

Algorithm: TOKEN PRE-COMPUTATION

- i) Begin
- ii) Choose file F to upload & encrypt the file using AES .
- iii) Generate $n \times m$ Vector Matrix D on file F .
- iv) Create Reed Solomon Matrix P over Galois Field $GF(2w)$ where $w=4$.
- v) Generate Matrix $C=D \times P$. It is Checksum Matrix created for fault tolerance.
- vi) Compute Token over Matrix C i.e. $ComputeToken(C,l,t,r)$, where l – block size, t – no. of tokens, r – indices per verification. Compute the tokens by pseudorandom function PRF & pseudorandom permutation function PRP .

2.5. Correctness Verification & Error Localization

To eliminate the errors in storage systems key prerequisite is to locate the errors. However, many previous schemes do not explicitly consider the problem of data error localization, thus only provide binary results for the storage verification. In our scheme we integrate the correctness verification and error localization in our challenge-response protocol. The newly computed tokens from servers for each challenge are compared with pre-computed tokens to determine the correctness of the distributed storage. This also gives information to locate potential data errors.

Algorithm: Correctness Verification:

- i) Begin Challenge i , for $i=1$ to n , where n – total number of cloud servers.
- ii) Get $TokenArr()$ // Getting pre-computed tokens from main cloud server.
- iii) $HandleChallenge()$ // Reading file blocks from all cloud servers for calculating new tokens.
- iv) Generate Vector Matrix D on all file blocks that are read in step 3.
- v) Create Reed Solomon Matrix P
- vi) Generate Matrix $C=D \times P$. On this matrix, new tokens will be computed.
- vii) Compute token on Matrix C . $ComputeToken(C,l,t,r)$
- viii) If $((Precomputed\ token == newly\ computed\ token))$ then,

- ix) Data is intact Else Data is Corrupt. For that i , initiate the recovery.
- x) End

2.5.1. Error Recovery & File Retrieval

Once the data corruption is detected, next important step is to recover the corrupted data and bring data storage back to consistent state. The comparison of pre-computed tokens and received response values can guarantee the identification of misbehaving server. Therefore user can recover the corrupted data. Our system recovers data from backup server & distributes all data vectors to corresponding servers. This will results in successful recovery of corrupted data. But due to file splitting we made at the time of file distribution, user's need to recover file from all the servers. Error localization is limited to misbehaving servers only, i.e. servers giving false assurance of posing users data.

Algorithm: Error Recovery

- i) Begin (Assume that the data corruptions have been detected & $s \leq k$ servers have been identified misbehaving.)
- ii) Download consistent data blocks from backup server.
- iii) Create the data vectors as per number of cloud storage servers.
- iv) Distribute the consistent data blocks to corresponding servers & recover the data. 5. End.

2.5.2 Dynamic Operations

In cloud data storage, there are many potential scenarios where data stored in the cloud is dynamic, like electronic documents, photos, or log files etc. Therefore, it is crucial to consider the dynamic case, where a user may wish to perform various operations of update, delete and append to modify the data file while maintaining the storage correctness assurance. The straightforward and trivial way to support these operations is for user to download all the data from the cloud servers and re-compute the whole parity blocks as well as verification tokens. This would clearly be highly inefficient. In cloud data storage, sometimes the user may need to modify some data stored in the cloud, from its current value to a new one. We refer this operation as data update.

To perform update operation on particular data block client need to recalculate the verification token on updated data. Also client need to update this value of newly calculated token to all the replicas of file in storage cloud. When user want to perform update operation, the split file from all storage servers is merged and given to the user to perform data updates. Once user has finished with the

updating the data, new tokens are calculated on whole file and they are stored on main cloud server. After this, updated file is divided back and dispersed onto corresponding cloud storage servers. Update operations include modifying file, inserting data, as well as deleting data from file. Sometimes, after being stored in the cloud, certain data may need to be deleted. The delete operation we are considering is a general one. When user wants to delete some file, he can simply delete it.

In delete operation, file blocks that are distributed among cloud storage servers are all deleted. Once file is deleted, we cannot perform any recovery of deleted files as there won't be any backup available in main cloud server. In some cases, the user may want to increase the size of his stored data in file by adding data at the end of the data file, which we refer as data append operation. So in case of append operation whenever user append data to his file, new verification tokens are calculated & stored on main cloud server & file is divided as before and dispersed among the cloud storage servers. Previous schemes [5, 6, 7] don't support insert operation on data. In our scheme user can insert data at any location wherever he desires. In case of insert operation, we are treating as a part of update operation and we are relying on update operation for insert operation.

III. IMPLEMENTATION & RESULTS

3.1 System Implementation

We have implemented our system with the help of web services in .Net. The functionalities of cloud servers are provided by web services. We have used C#.Net to implement the system with 4.0 frameworks. At the back end side we used MySQL server. Visual studio 2010 editor is used to develop the application. As we know, to use any application we need to register as well as to authenticate ourselves as a user. In our system, a user has to register them to use the system.



Fig 3. User registration snapshot

As compared to its predecessors, our system is proactive one i.e. as soon as user is registered with the system, he will be notified with an alert message on his mobile and with an email. After that user each and every action is monitored by the system and information is conveyed to the respective users as an alert message on their mobiles. Figure 3 shows the new user registration screen. Users can use the system after authenticating them at login.

Users can use the system after authenticating them at login. After logging into system, user can update his information filled at the time of registration. The mobile number with which users have been registered, are used of providing alerts on every action of user. File upload facility is provided to users. Any type of file can be uploaded into cloud but, dynamic data operations are performed on only .txt, .doc and .rtf file formats. File upload screen is shown Figure 4. In file uploading, file is read into byte array and divided into 3 equal parts as we have assumed three cloud servers. We call these parts as data vectors, which we multiply by Reed Solomon erasure codes to create redundancy. After this, tokens are calculated & saved into cloud server & file is dispersed among cloud servers.



Fig 4. File upload snapshot

After uploading files, we have provided download tab which displays all the files uploaded by the user. Here user can download the file to his local machine. Also, user can open the file for appending or updating without downloading it. As we have mentioned, dynamic operations like append, update, delete can only be performed on .doc, .rtf, and .txt file formats only.

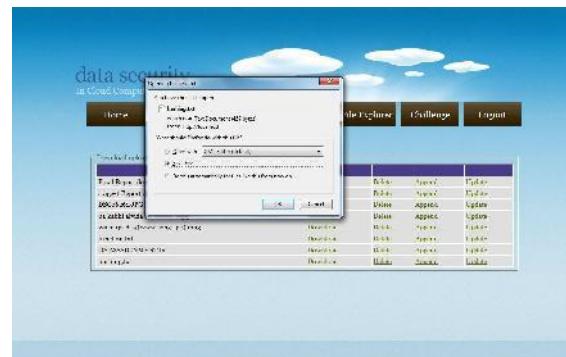


Fig 5. File download snapshot

When user performs any dynamic operation, the new token values are calculated and stored back into cloud. Modified file is divided again & dispersed to all cloud servers to maintain consistency.

To demonstrate data modification attacks, we have provided a hacker account with access to all the uploaded files by all users. Hacker performs the dynamic operations on files like an authorized user. Our system detects every single data modification done by an unauthorized user. Figure 4. shows file download screen with file download box. File append operation is shown in Figure 6.



Fig 6. File append snapshot

We have provided file explorer tab to user, where user can enter a string & system will search & display matching files inside a grid view. From this page, user can also verify the correctness of his data.



Fig 7. File explorer snapshot Most

Most important part is correctness verification & recovery of data. This facility is provided under challenge tab, where user can select the file to verify & server on which the correctness to be verified.

Any modification to the file done by an authorized user will leave file in correct or intact state but, if file is modified by an unauthorized user then system will give alert as "File has modified by unknown user...Please recover your file on selected server." On same tab we have provided an option to correct the data blocks. In this case user needs to select the particular file & server and to click on correct data blocks.

As we have divided the file according to number of servers, at the time of data recovery we manually needs to select all the servers and recover file on each server to ensure correctness of our data. The main advantage of our system is that every unauthorized modification to data will get detected. Again, user's can challenge the cloud servers as many times they want to verify correctness of their data. There is no limit on number of challenges. No burden on user's to store pre-computed tokens locally. Figure 8. shows challenge screen.



Fig 8. Challenge snapshot with alert message

3.2. Results

Our model is not probabilistic rather it is systematic & it guarantees detection of every single data modification attacks. To demonstrate an unauthorized access to the system, we have created hacker account which has given access to all registered users data. Hacker can modify any file from any user. Token pre-computation technique helps us to find out any modification to users data. It guarantees that no single unauthorized data modification is left undetected. Figure 9 shows a screenshot where data modification has been detected.

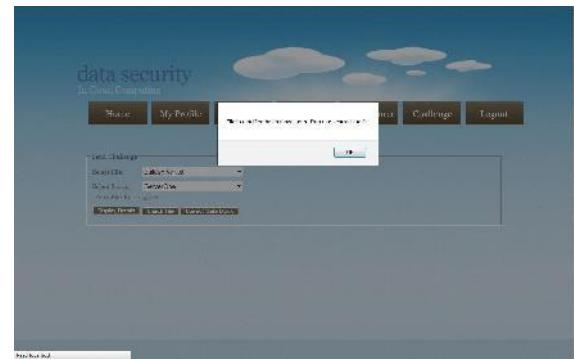


Fig 9. Data modification detected snapshot.

IV. PERFORMANCE ANALYSIS

Our model is systematic & it guarantees detection of every single data modification attacks. So there is no probability of detection as previous work rather there is guarantee of detection of each modification attack. We have evaluated the performance of our system in following cases:

1. Token pre-computation time.
2. File distribution time.
3. Server Token computation time.
4. Server response time.

Our experiment is conducted using C#.Net on a system with an Intel core 2 duo processor running at 2.10 GHz, 4 GB of RAM, and a 7200 RPM Western Digital 320 GB Serial ATA drive. We have tested our system under upload speed of 1Mbps & varied file size up to 10MB. We restricted to this much file size because generally we don't have .txt, .doc, .rtf files with greater size. We can upload up to 25 MB file.

The token pre-computation is long process. It includes encrypting users file then converting users file into data vector matrix D. Next step is to create Reed Solomon matrix P. Later we generate checksum matrix C on which we computes our short tokens. After that the file is distributed to all cloud storage servers.

V. CONCLUSION

We have analyzed the data security concerns in cloud data storage, which is a distributed storage system. We proposed a distributed scheme to ensure users that their data are indeed stored appropriately and kept intact all the time in the cloud. To provide redundancy we used erasure correcting code in the file distribution preparation. As we all know cloud is not just a third party data warehouse. So providing support for dynamic operations is very important. Our scheme maintains the same level of storage correctness assurance even if users modify, delete or append their data files in the cloud. Challenge response protocol along with pre-computed token is used to verify the storage correctness of user's data & to effectively locate the malfunctioning server when data corruption has been detected. Through detailed performance analysis, we show that our scheme is having very low communication overhead & guarantees to detect every single unauthorized data modification. Our scheme has no limitation on number of pre-computed tokens used for challenging the cloud servers. Unlimited number of challenges can be made. We removed burden of calculating pre-computed tokens & storing the locally from the users. By splitting the file according to the number of servers we are added extra security to system. But we still believe that data storage security in Cloud computing is an area full of challenges and of paramount importance.

REFERENCES

- [1]. Juels, J. Burton, and S. Kaliski, "PORs: Proofs of Retrievability for Large Files," Proc. ACM CCS '07, Oct. 2007, pp. 584–97.
- [2]. G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc. ACM CCS '07, Oct. 2007, pp. 598–609.
- [3]. G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," Proc. SecureComm '08, Sept. 2008.
- [4]. H. Shacham and B. Waters, "Compact Proofs of Retrievability," Proc. Asia-Crypt '08, LNCS, vol. 5350, Dec. 2008, pp. 90–107.
- [5]. K. D. Bowers, A. Juels, and A. Oprea, "Hail: A High-Availability and Integrity Layer for Cloud Storage," Proc. ACM CCS '09, Nov. 2009, pp. 187–98.
- [6]. C. Wang, Qian Wang, Kui Ren, Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing," Proc. IWQoS '09, July 2009, pp. 1–9.
- [7]. Q. Wang, C. Wang, Wenjing Lou, Jin Li, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing," Proc. ESORICS '09, Sept. 2009, pp. 355–70.
- [8]. Erway, Alptekin, Charalampos Papamanthou, Roberto Tamassia, "Dynamic Provable Data Possession," Proc. ACM CCS '09, Nov. 2009, pp. 213–22.
- [9]. R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "MR-PDP: Multiple-replica provable data possession," in Proc. of ICDCS'08. IEEE Computer Society, 2008, pp. 411–420.
- [10]. T. Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in Proc. of ICDCS'06, 2006.
- [11]. N. Gohring, "Amazon's S3 down for several hours," Online http://www.pcworld.com/businesscenter/article/142549/amazons_s3_down_for_several_hours.html, 2008.
- [12]. M. Arrington, "Gmail Disaster: Reports of Mass Email Deletions," Dec. 2006; <http://www.techcrunch.com/2006/12/28/gmail-disaster-reports-of-massemail-deletions/>
- [13]. Peter Mell, Tim Grance, "The NIST Definition of Cloud Computing", Online at <http://www.nist.gov/itl/cloud/upload-def-v15.pdf>.
- [14]. K. D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Cryptology ePrint Archive, Report 2008/175, 2008, <http://eprint.iacr.org/>.
- [15]. M. Lillibrige, S. Elnikety, A. Birrell, M. Burrows and M. Isard, "A Cooperative Internet Backup Scheme", *Proc. of the 2003, USENIX Annual Technical Conference (General Track)*, pp. 29–41, 2003.
- [16]. Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE Transactions on Parallel & Distributed Systems, Volume: 22, Issue: 5, pages: 847-859.
- [17]. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for storage security in cloud computing", in *Proc. of IEEE INFOCOM'10*, San Diego, CA, USA, March 2010.
- [18]. J. S. Plank and Y. Ding, "Note: Correction to the 1997 Tutorial on Reed-Solomon Coding," University of Tennessee, Tech. Rep. CS-03- 504, 2003.[10] T. Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in Proc. of ICDCS'06, 2006.

Android App for Renting Daily Use Items

Harsh Parekh

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Kartik Iyer

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Hetal Amrutia

*Thakur College of Engineering and Technology
Department of Information and Technology
Mumbai, India*

Harsh Doshi

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Abstract —This paper discusses about a Rent Application provides a way of solving the problems faced by the people by saving cost of buying new products. The objective of our application is to make money sitting at home by just renting the products which are not used for time being, and to provide the customers with an effective application to find and give on rent, monitor purchases, handling performance and make improvements in the manual process. Rent Application is a management technique for renting products. This software is used to rent the products and respond to changes taking place in the new products, as well as provide an outlook of business opportunity to customers.

Keywords – analysis; payments; e-commerce; database; dashboard; security.

I. INTRODUCTION

Android is popular with technology companies that require a ready-made, low-cost and customizable operating system for high-tech devices. Its open nature has encouraged a large community of developers and enthusiasts to use the open-source code as a foundation for community-driven projects, which deliver updates to older devices, add new features for advanced users or bring Android to devices originally shipped with other operating systems. The success of Android has made it a target for patent (and copyright) litigation as part of the so-called "smart phone wars" between technology companies.

Android's default user interface is mainly based on direct manipulation, using touch inputs that loosely correspond to real-world actions, like swiping, tapping, pinching, and reverse pinching to manipulate on-screen objects, along with a virtual keyboard.

This is an application which would first ask the users to sign up. The users can then look for things to rent by searching or visiting the categories which include but not limited to electronics, cars, clothes. The user can also look up the requirements for the rented item. The user can know about the deposit upfront, the rent/day and the documents requested for verification. If the user agrees to those conditions, then he can upload the documents and send the request to seller. The seller will check and accept/reject the request. If the deal goes forward, then the buyer will send the deposit through the app. The money will be kept within the company for any complaints later. It would be visible to both parties but be in a locked state, only accessible one the item is returned, or damage complaint verified.

The user can also sell items by uploading relevant data, and requirements, he can choose to ask for additional user data like driving license in case for cars. This will help create a sense of security for the user. The system is automated maintaining detail about products and the payment can be done easily which will prove to be cost effective to customers.

II. BACKGROUND

Android is a mobile operating system developed by Google^[6], based on the Linux kernel and designed primarily for touch screen mobile devices such as smart phones and tablets. Android's user interface is mainly based on direct manipulation, using touch gestures that loosely correspond to real-world actions, such as swiping, tapping and pinching, to manipulate on-screen objects, along with a virtual keyboard for text input. In addition to touch screen devices, Google has further developed Android TV for televisions, Android Auto for cars and Android Wear for wrist watches, each with specialized user interface. Variants of Android are also used on notebooks, game consoles, digital cameras, and other electronics.

This project for renting products is Android app based renting products for customers, and is alternative to existing buying and selling process which need more money and effort. A Rent Application is any defined platform in the market that is used to sell or buy products. There are existing online portals like Amazon, Flipkart, OLX and Quickr which allows user to buy and sell products but none that allows user to rent products for limited time. So, this renting application fills that void. This system is automated and will provide features of better management, monitoring, and maintenance of the products and renting process. It will ease the entire purchasing process and lower the cost needed to use a product by renting it. Our system is concerned with customers, which will take buy or sell products on rent and will be managed by an administrator or concerned person.

OLX^[6] is an online buying and selling platform, which does not include the ability to rent. The unique features of the platform is the ability to chat with other people before buying.

Quickr is another competitor to OLX, in terms that it only allows the user to buy and sell, they are almost equivalent in terms of features available.

Amazon and Flipkart are online portals for buying brand new products directly from retailers themselves, providing easy purchases and discounts to users.

As our project is based on an app, the user can check it out through their mobile and just have to login to the app and search for the relative application. This isn't time consuming. There is no need to manually go and check the products. Normally customers have some forms filled up probably there are some transaction too if the problem must be resolved within some deadline. By using this app customers can rent their products using their smart phones get their products rented faster. We came across various rent paying methods and the survey for different types of products to get acquainted with how the industry works today.^[4] We have also studied the market share of E-Commerce sites for various categories in India. In this we studied detailed Comparison of Amazon and Flipkart.^[3] Additionally, we learned how trust plays an important part in online portals thereby providing trust to the users of our application the trust and transparency to give them the appropriate services.^[2] On studying a paper we learnt about various rent paying methods for different types of products. In this we studied how to use payment gateway for our application.^[1] Customer needs are important information

reflecting customers sound and is a primary measure of customer behavior. An Effective and Efficient response to these complaints is an essential index of organization's performance. This system 'Android App for Rent Application' will discourage users to buy new things for short period of time and on the other hand it can encourage participating in renting their household items. This system is for everyone and will be user to resolve their problem regarding not being able to effectively use items that remain in the house. It will be an alternative to current offline process which will reduce time and effort needed for current process.

III. MOTIVATION

Problems regarding purchasing the products for which are to be used for very short span of time is major part for development of this service. Nowadays customers face problem relating to time. In the sense, they must go to the dealer's office to rent any product such as vehicles, electronics or clothes for renting or buying on rent and wait for their response. This tends to waste of time. So, taking this issue in consideration we thought of developing an application which will be useful to the customers.

If someone is going to move into a new city for just a few days then, they shouldn't need to buy everything that they need for the stay, this is where this system would be exceptionally helpful. Even for sellers, who are going out for a few weeks or days, they can rent out their products which would otherwise just sit at home gathering dust. All products depreciate over time and this is a way of earning back some of the value of the product.

The main purpose of our project is to save time and money. Through the survey we found that there is need to develop such an application which will be of use to customers. And this motivated us to develop such an application which will fulfil their needs. On a larger front this project also helps us to discourage the using and throwing of daily household items, thereby putting a step forward towards a greener tomorrow. The main objective is to reach to the customers and help them by letting customers buy products which they want to use for short span of time or sell products on rent which are not of use to the customer for some time thereby generating revenue.

- a. To reduce the cost by renting instead of buying and selling.
- b. To provide faster access to the products because customers itself are selling and buying products on rent.

IV. SCOPE

The main purpose of this project is to generate revenue for both buyers and sellers from the goods or items which are unused. It will help the buyers buy items on rent which are needed on temporary basis instead of purchasing thereby saving the amount of purchasing it. And for buyers it would help to sell those things which are not needed in near future thereby generating revenue sitting at home easily. To gain customers trust by providing the appropriate services and payment facility.

V. PROBLEM DEFINITION(PHASE WISE)

There are no online portals for renting, this helps to bridge the gap between online buying and selling products and renting them so that the products which are not used can generate an alternate source of income. Since there are renting based services available in the market this is a new platform where the users can then look for things to rent by searching or visiting the categories which include but not limited to electronics, cars, and clothes.

A. Phase 1

Perform literature survey and identify all the research gaps. Learn about the functioning of the E-Commerce websites by observing the websites and applications by them and analyzing the current working situation of the organization. Prepare a report or synopsis to set a path that is to be followed throughout the project. After understanding the requirements and making a synopsis the development is to be started. In development first step is to plan on how the project is to be done. The second step is to design a prototype or a basic structure of the complete database. A star schema is to be created to join the data objects together. So first all the dimensions and the facts are to be written and a design of star schema is to be created. After both the steps are done completely actual development is to be started taking help of the plan and the designs created. The basic application design and code is to be created by using Visual Studio and Android Studio. The server-side code is to be deployed on Microsoft Azure Cloud. After all the steps are completed successfully the application is to be tested in the next phase.

B. Phase 2

Create various test cases to check if the integrated data table functions as desired by the client. Create various test cases to check the proper functioning of the server and application. After testing and approval, deploy the proposed system to the IT department.

VI. METHODOLOGY

The application would first ask the users to sign up. The users can then look for things to rent by searching or visiting the categories which include but not limited to electronics, cars, and clothes. The user can also look up the requirements for the rented item. The user can know about the deposit upfront, the rent/day and the documents requested for verification. If the user agrees to those conditions, then he can upload the documents and send the request to seller. The seller will check and accept/reject the request. If the deal goes forward, then the buyer will send the deposit through the app. The money will be kept within the company for any complaints later. It would be visible to both parties but be in a locked state, only accessible one the item is returned, or damage complaint verified.

The user can also sell items by uploading relevant data, and requirements, he can choose to ask for additional user data like driving license in case for cars. This will help create a sense of security for the user.

VII. FEASIBILITY STUDY

Technical Feasibility: The system works on the app model and is the system implemented using Java, C#, etc. as programming and scripting languages. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms.

Economic Feasibility: The cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud. Cloud computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud. Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

Legal Feasibility- The project is developed under legal license of Windows and other software's and databases. The data will not be used for some illegal purpose and will be encrypted using algorithms.

Operational Feasibility- An overview of the process portrayed that the proposed Rent Application would be beneficial by providing a unique concept to users hence being the only player in the market.

Social Feasibility- The Application allows users to chat with each other giving a sense of social connect before actually doing through the rent agreement. This offers another aspect of the feasibility.

VIII. FLOW DIAGRAM

Figure 1 is a flowchart diagram. Flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem.

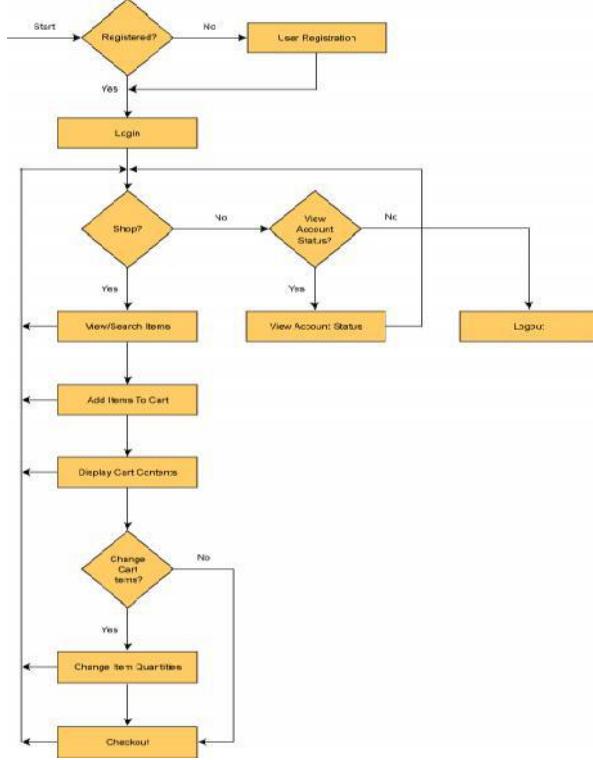


Fig1: Flow Chart Diagram for registration and sign in.

IX. RESULTS

To create a product that is commercially viable with a good business model ensuring good service, profits and discounts to both buyers and sellers so that they can generate a source of income sitting at home. There are screenshots of some of the pages of the application below.

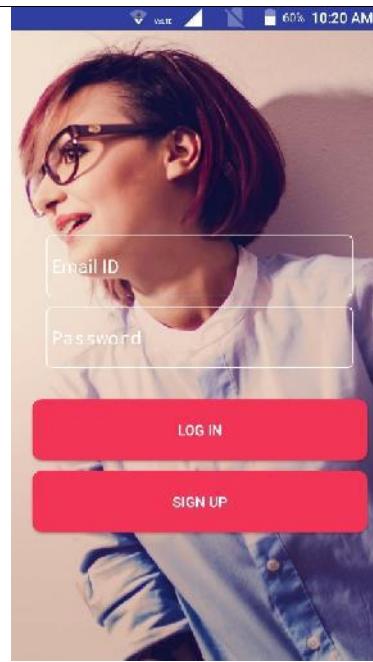


Fig 2: Login page of the application.

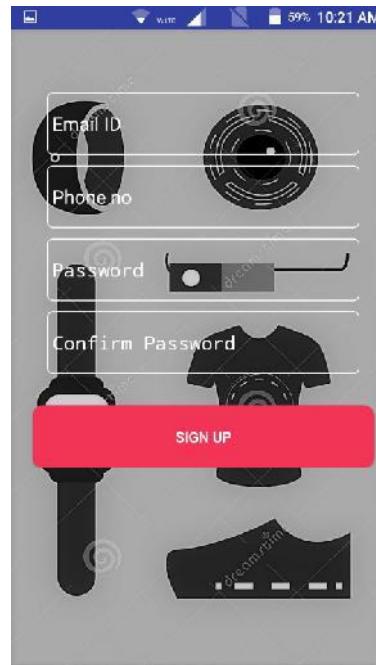


Fig 3: Sign Up page of the application.

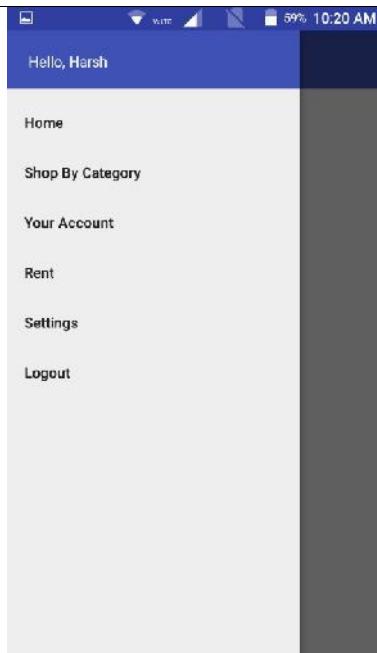


Fig 4: First page with hamburger menu.

X. CONCLUSION

The rent application has been a work in progress in terms of the backend connectivity and the user interface. The final version of the app is expected to support the entire feature set, with future version having payment option built in as well. The application is expected to strive based on the lack of any competition in the market currently.

REFERENCES

Paper References

- [1] Michael H. Anderson "Rent-to-own agreements: Customer characteristics and contract outcomes" in ISSN: Volume 61, Issue 1 EDM 2009
- [2] Nidhi Vishnoi Sharma, Varsha Khattri "Study of online shopping behavior and its impact on online deal websites" in ISSN 2229 – 3795 EDM 2016
- [3] Vidyashree DV, Neelam Bhandari, Richa Sharma, Pooja M Verma, Priya Chauhan "A Comparative study on consumer Preferences towards online retail marketer" in ISSN: 2394-5869 in EDM 2015
- [4] Swapnil V. Mishra & Dr. Shamkant N. Kotkar "A Study on Current Status of E-Commerce in India" in ISSN:232 7782 Volume 3,Issue 2 EDM 2015

Web References

- [5] <https://stackoverflow.com/search?q=android+development>
- [6] <https://www.olx.in/posting/>
- [7] <https://azure.microsoft.com/en-in/>

Smart Mirror Using IoT

Charvika Banjan

*Department of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India*
charvikabanjan@gmail.com

Vamil Gandhi

*Department of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India*
yamilg@gmail.com

Saurabh Khandelwal

*Department of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India*
saurabhkhandelwal100@gmail.com

Anil Vasoya

*Department of Information Technology
Thakur College of Engineering
&Technology Mumbai, India*
vasoyaanil32@mail.com

Abstract-Effective time management is an essential factor in increasing production of day-to-day life. Integration of technology into people's daily lives has made that time management possible. The use of products such as tablets, PCs, and smart phones have given people access to the tools needed to be productive. However, though successful technological products have been used to increase productivity, the use of technology has become another task on everyone's daily to-do list. Technology should mold to our schedule, not the other way around. That is where the "Smart" mirror idea originated. "Smart" mirrors have been envisioned for years, part of the broader trend in imbuing everyday objects with various "smarts" to improve our lives. The smart mirror idea aimed to integrate technology seamlessly into people's lives by putting it where everyone's routine eventually collides. The goal of the smart mirror is to increase a user's productivity by saving themtime.

Parameters that user wants is the effective solution for checking updates on the go.In turn saving their time to check mobile notifications, weather, checking traffic and booking the cab. In this project we have tried to implement all this services and provide effective solution behind the mirror which will help the user to use this functions while they do their

According to the literature survey, we have found that the smart mirror takes high power consumption which is an overhead on smart mirror. We found that the users consented that mirror interaction was understandable but the users were unhappy with the visual quality of the system. On a scale of 5 the parameters ratings were observed as 3.7 in visual

quality and 3.9 in ease to use. This project will try to increase the visual quality and it will try to be more user friendly so it becomes easier to use. In consumer applications the corefunction is basically twofold: Be a general informational hub, derived from viewing one's reflection as well as augmented with other useful information and provide some style and comfort benefits, in addition to mirror reflectivity and other functions. By connecting IoT to smart mirror, it is possible to implement a variety of application services. Smart mirror that has been linked with IoT platform is friendly and provides varieties of information to user. Generally the humble mirror gains such sentence in one of two ways: adding layers of smart coatings, or embedding a variety of electronics: sensors, cameras and display, touch interface, lighting, and the software and processing to stitch everything together a prototype of a smart mirror is proposed which has been linked with IoT platform and the artificial intelligence features where an user can interact with the mirror and get different types of information like weather report, time, calendar, maps, compliments, browsing and much more.

I. INTRODUCTION

Smart mirror has been developed within the context of a time where every day we see more and more connected devices. The Internet transformed our lives by connecting us more easily to information and other people in the virtual world. Mobile phones then became smartphones and since then this concept has erupted and morphed into the Internet of Things, things which connect us to everyday objects. Our research is exploring a new way to use technologies

to be more informed. The state of innovation currently is to provide more information, but less interaction to get it. The device that has been researched and designed is called SmartMirror. It is a wall mounted mirror which displays relevant items to the user such as weather, scheduling, messages and other fields of interest. The mirror will solve the problems that many people experience every day, getting information without distraction.

Before going to bed, the user may want to know whether it will show the next morning so that they can plan their commute. The aim of this research project is to explore the impeding shift in how people receive information. We see wall mounted clocks, which provided people with access to the time at a glance. We have displays in airports to show the timetable so that passengers can see the information. We also see massive tickers and advertisements. We have tried to personalize this concept and bring it to the home. The Smart Mirror lets people use their device less, while being presented with more. This is a major step in the evolution of being informed. This is the moment technology is starting to get out of the way. Heads will be up, hands will be set free.

II. OVERVIEW

Smart mirror is useful for the busy individuals that want to multitask and stay informed while on the go. Instead of constantly pulling out a device, one could get informed while finishing daily grooming tasks. Today everyone is busy, but for a while, when you look into the mirror and see something more than yourself; the mirror could tell you that it's cold outside and recommend you to wear a sweater. We wish to develop Smart Mirror for application in a personnel room; its features include voice detection, display notifications, clock, maps, messages, weather, date, news update. We will use Raspberry pi, Display monitor, Two-way mirror, USB mic etc. The programming language we wish to use is Python

III. LITERATURE SURVEY

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they can move forward. [1] As the mirror interaction was understandable but the users were unhappy with the visual quality of the system, the parameters that we'll improve the observed

ratings by 15% [2]. Rather than confined to home we can implement the functionality to a glass material. So that it can have wide range of applications like one can setup this functionality to a glass table. We'll be using Raspberry Pi 3 which contains 1 GB SDRAM which will in turn increase the efficiency by 50% [3]. The X10 protocol is not as reliable as expected and sometimes commands get lost in the transfer. Also the COM part in the .Net framework can at times be locked by the previous command centre. To fix these issues a delay of 1sec should be created after a command is sent to a device. Theoretically the accuracy was stated to be 96.5% but then we found it to be only 94% so we are improving the accuracy to 95%

IV. PROBLEM DEFINITION

Our lifestyle has evolved in such a way that optimizing time is the most important thing. Based on the user studies and prototype implementation, we present the development of an innovating appliance that incorporates interactive services of information, offered through a user interface on the surface of a mirror. Our work is based on the idea that we all look at the mirror when we go out, so why wouldn't the mirror become smart. The framework will offer basic services, like the presentation of personalized weather data, time, date and will incorporate some additional functionality, like reminder service by mobile synchronization and through social media. Our framework is based on detecting presence of human using Passive Infrared sensors and Wi-Fi connectivity. Once a person comes in front of the mirror, it displays the information that is being fed from the web. This data or information includes calendar, time, weather, news feed, notifications and so on. Our framework also discusses about the speech recognition and its application in control mechanism in home appliances and opening and closing of shelf. We use speech recognition to automate many tasks that usually requires hands-on human interaction, such as recognizing spoken commands to perform something like turning on lights or shutting a door. Our framework also introduces speech activated music player, and plays the music when a person gives a command.

PHASE 1: (Planning, Analysis, Design, Coding)

Planning: Conceptualizing and designing the architecture for proposed solution. Whether to use Raspberry Pi 2 or Raspberry Pi 3 systems.

Analysis: Critical study, analysis and review of feasibility for proposed solution. Choosing the right type of components such as Pimicrocontroller.

Coding: Creating various modules of an IT based solution product for analyzing the data of the guests.

Design: RaspberryPi 3, line of sight antennae to set up the network architecture.

PHASE 2: (Integration, Testing, Deployment)

Integration: Integrating of various proposed modules such as Display of Date, Time, Weather and Traffic updates, social media

Testing: Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.

Deployment: Use the completed prototype for evaluation purpose and beta testing.

V. METHODOLOGY USED

The proposed Smart Mirror has various functionalities implemented and is designed and developed using a building block architecture. This therefore prepares the Smart Mirror for future scalability as it can support extended home automation. This means that through the Smart Mirror interface, in the future, other home automation IoT devices can be controlled like ambient lighting systems, centralized air conditioning systems, and even other modular home automation systems like Google Home. Our Smart Mirror is designed to mimic a natural mirror interface. A two-way acrylic mirror is used for the mirror display, thereby mimicking the function of a regular mirror. A flat LCD monitor is used to provide real time display of the content on the Smart Mirror which is powered by a Raspberry Pi, thereby mimicking the function of a SmartMirror.

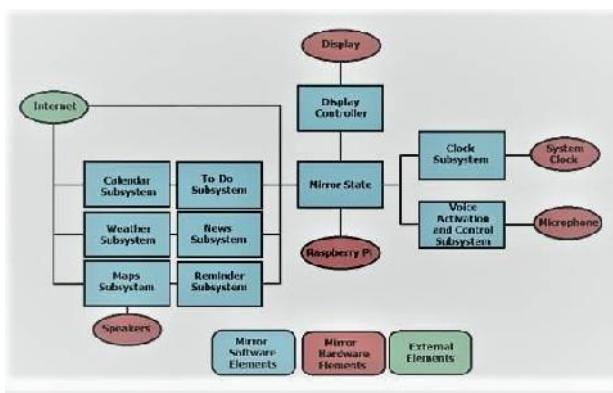


Figure: Data flow in the smart mirror

The Smart Mirror has the functionality of many modern day smartphone applications implemented on it in the form of widgets. Some of the applications incorporated into the working design of the Smart Mirror are those of Bing Maps (Along with real time display of traffic in a particular location), YouTube (Functionality to play

videos along with the audio output), a Date and Time widget (Displays the current system date and time. It can also update this information by obtaining it from the internet), a Weather widget (Displays the weather information based on either automatically detected location, or manual location information fed to it in terms of Latitude and Longitude). The temperature units system (Fahrenheit or Celsius) displayed in the widget is based on the popular units system used in the location. This too can manually be changed by the user based on the user's units system of choice. Users will be able to obtain minute updates of latest news and public headlines, with the Rich Site Summary (RSS) feed widget. Along with the aforementioned widgets and applications, the Smart Mirror also has application functionality. These functionalities are those of a Countdown Timer which counts down from a user specified time measurement, a Reminders application which displays the user's manually set reminders. There is also a To-Do List which has a list of to-do items that the user manually inputs, and a widget that displays the next-air date of a list of the user's favorite televisionshows.

VI. FEASIBILITY STUDY

- A. *Technical Feasibility* - Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.
- B. *Economic Feasibility* - As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.
- C. *Legal Feasibility* - Legal issues can affect a system's acceptance by Bottom of Form users, its performance, or the decisions on whether to use it in the first place so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any log analysing system.
- D. *Operational Feasibility* - It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.
- E. *Social Feasibility* - The acceptability of a log analysis

system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness

VI .OUTCOMES

Phase 1.1 Planning

- Proposing Statement of work, scope definition and scope boundary for planning the prototype fromthe problem definition to decide what needs to be done and what not to be done

Phase 1.2 Analysis

- Creating the synopsis, literature survey andfeasibility study for proposed solution which will act as an input to the designphase.
- Choosing RaspberryPi3 over Arduino for higher computing power and betterperformance.

Phase 1.3 Design

- Proposing network architecture for Home for implementing the surveyed case studies andsynopsis.

Phase 1.4 Coding

- Various modules will be createdlike Analyzing the logfiles
- Creating the dashboard for everyprominent location

Phase 2.1 Integration

- Constructing the various modules of the proposed solution to integrate them into aprototype.

Phase 2.2 Testing

- The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stresstesting.

Phase 2.3 Deployment

- The benefits of Analysis on after deployingthe proposed prototype include

 1. Proper and more efficient way of monitoring ofusers.
 2. Effective way of managing and storing userdata.
 3. Easier to manage the flow of users in different locationseffectively.
 4. Reduce the time consumed for checkingnotifications onphone
 5. Controlling home through mirror

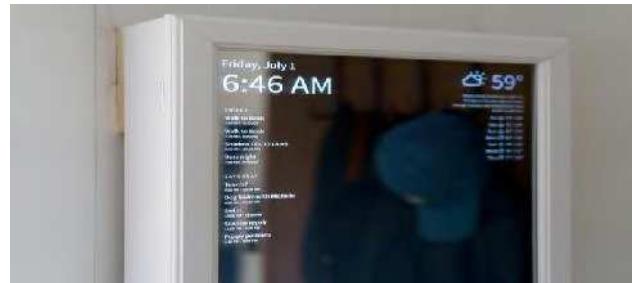


Figure: Home Screen of smart mirror

VII. RESULTS

The Smart Mirror system we have developed combines the concepts and methodologies that have been implemented in other systems like smartphones. It is a novel method of creating a smart interactive system that is reliable and easy to use. We have concentrated on an interactive and user friendly home appliance that is beneficial to the user. The architecture has been adapted for the development and deployment of various services which all use web service communication mechanisms. These services of the Smart Mirror can be further extended to support future enhancements and developments. This means that features such as extended home automation (control ambient lighting systems, centralized air conditioning systems, and other modular home automation systems/devices like Google Home) and additional applications and widgets can be implemented. A user's personalized data such as calendar, news feeds, and other information relevant to their lifestyle have been implemented and displayed. The Smart Mirror uses voice commands to switch between each view, and performs the functions corresponding to the input voice command. Rather than become confined to a home, we can implement the functionality of our Smart Mirror system to commercial work, and public environments. Each widget and applicationhas

been developed using a building block approach, keeping the aspects of future enhancement and scalability in mind. The main advantage of using this building block approach is that any new application or widget can be implemented and put to functional working by developing them as plugins and using them as required to meet the environment and user's needs. The Smart Mirror we have developed takes an average of 15 seconds to begin execution once the command to initiate its process has been input into the Terminal window in the Raspberry Pi. On initiation, when the Raspberry Pi is connected to the internet, a unique IP address with the port

-8080 suffixedtoit, isgenerated. ThisIPaddressisdisplayed in the terminal window and can also be viewed in the Smart Mirror process by saying the command —Show

Remote Link . The IP address is used to connect the user's laptops, smartphones, or any other device capable of connecting to the internet, remotely to the Smart Mirror. That is, the Remote Configuration Tool (RCT) is initiated whenever a device uses the generated IP address to connect to the Smart Mirror. As a prerequisite, the device running the RCT needs to be connected to the same network connection as that to which the Raspberry Pi of the Smart Mirror is connected. When the Smart Mirror is running, voice commands can be masked by background noise that is present in the environment in which the Smart Mirror is placed. This therefore can result in false positives and true negatives of voice command input to the Smart Mirror. In order to overcome this, the user can either manually alter the Configuration File or use the RCT to alter the sensitivity of the microphone. The voice command to activate the Smart Mirror is user-defined and was set to "Smart Mirror" in our implementation. Once this command has been given, the Smart Mirror actively listens for the user's command, and once the subsequent voice command is input, the Smart Mirror performs the

function associated with it.

REFERENCES

- [1] S.Athira;—Smartmirror:Anovelframeworkforinteractive display in 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2016, pp.1-6
- [2] M. Anwar Hossain;—Smart mirror for ambient home environment ,3rd IET International Conference on Intelligent Environments, 2007, pp. 589 – 596
- [3] Derrick Gold, David Sollinger, Indratmo;—SmartReflect: A modular smart mirror application platform , IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016, pp. 1 -7
- [4] Guangyu Gao;—The Future of Smart Dressing Mirror: An Open Innovation Concept Video , IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing, 2015, pp. 369 –369
- [5] A.S.M.MahfujurRahman;—AugmentedRenderingofMakeup Features in a Smart Interactive Mirror System for Decision Support in Cosmetic Products Selection , IEEE/ACM 14th International Symposium on Distributed Simulation and Real Time Applications, 2010, pp. 203 –206
- [6] ERCIM Working Group SESAMI,—Smart Environments and Systems for Ambient Intelligence <http://www.ics.forth.gr/sesami/>.

*International Conference on
Research Challenges in Security*

Video Processing in Visual System for Color Detection for People with Tritanomaly

Nazneen A. Pendhari

Department of Electronics & Telecommunication Engg
Thakur College of Engineering & Technology
Mumbai, India
mailnazneen@gmail.com

Raghavendra R. Sedamkar

Department of Computer Engineering
Thakur College of Engineering & Technology
Mumbai, India
rr.sedamkar@thakureducation.org

Abstract: In this paper, a video re-coloring algorithm have been performed for the anomalous trichromats type tritanomaly is presented. We firstly propose a method to handle the weaknesses faced by the anomalous trichromacy and finally build an interface that will bridge the gap for the colorblind to view visual media without any hindrance. The RGB to LMS theory has been utilized for the required purpose. The detailed study about the RGB to LMS theory with respect to the required equations and pseudo code pertaining to the type tritanomaly i.e. blue-yellow weakness has been proposed to be done here.

Keywords: Color Vision Deficiency (CVD), Trichromacy, Tritanomaly.

I. INTRODUCTION

Color seems to be the most pleasing and attractive characteristic for human beings and also most valuable in today's world of technology and digital media. As it is said, there are two faces to every coin where we enjoy this luxury with great comfort at the same time there are people who are deprived of this necessity. After luminance striking the human eye, the next most important entity that plays a vital role in human visual perception is color which is then categorized. Many individuals all over the globe suffer from distinct forms of color blindness. Some category of colors becomes completely unperceivable by the color blind. Consequently, they may encounter numerous trifles and drastic Tritanomatic vision CVD to experience the color difference between all the three primary colors which are namely Red, Green and Blue.

We all know that to perceive anything for normal humans, vision plays very important role which is solely driven by our ability to understand and perceive 3 primary colors (Red, Green, Blue) distinguishably. Tritanopia is type of CVD (Color Vision Deficiency) in which the patient cannot perceive Blue color at all as the S-cone cells are totally absent. Thus we have done in-depth studies of Tritanopia and have proposed an algorithm in this paper which does multimedia recoloring, so that the CVDs can atleast perceive all 3 primary colors distinguishably. It is distinct to the image mapping schemes, in order to perform video recoloring the consistency of colors in every frame of the video should be

reflected i.e. the color that is changed or recolored should be seen or reflected in all the frames that are bi-furcated. Color vision deficiency (CVD) is the vision problem where person is unable to perceive color such as Red, Green and Blue. There are three types of cone cells namely S-cones, M-cones and L-cones and each type has a different sensitivity to light wavelengths. S-cones perceives short wavelength color, M-cones perceives medium wavelength color and the L-cones perceives long wavelength color.

Protanopia and Deutanopia are two of the most common forms of color blindness and third one being Tritanopia caused by genetic disorder.

There is no scientific method or process to remove color blindness. However, people with color blindness may be able to use a special set of lenses/glasses to help them perceive colors more accurately. These lenses/glare are very costly to be taken for consideration. These lenses can only be used out-doors under bright lighting conditions thus making it difficult to use.

We know that to visualize any object for normal human vision colors plays vital role which helps to perceive the environment. Tritanomaly is type of CVD (Color Vision Deficiency) in which the patient confuses to perceive Blue and other 2 primary colors, as the S-cone cells have certain kind of malfunction or defect in them. We have proposed an algorithm in this paper which performs multimedia recoloring to help tritanomaly suffering CVD to distinguish Blue color with other colors in the frame.

Simple everyday tasks which includes cooking meat to selecting ripe produce can be a challenge for some housewives.

The issue over here is the similar color appearance in distinct frames cannot be assured to be mapped to the similar or twin new color, because the temporal coherence between the different frames is not surfaced.

The purpose of visual correction is to improve the perception of people viewing color information in images or videos with this type of deficiency. Basically, the core idea behind this algorithm is to remap the colors of the original image for accommodating the trichromacy in order to provide a better color representation in the image and videos

Table I : Analysis of color blindness

Type	Name	Cause of defect	Prevalence
Monochromacy	Rod Monochromacy	No functioning cones	Very rare
Dichromacy	Protanopia	L-cone absent	1%
	Deutanopia	M-cone absent	1.1%
	Tritanopia	S-cone absent	0.02%
Anomalous trichromacy	Protanomaly	L-cone defect	1.3%
	Deutanomaly	M-cone defect	4.9%
	Tritanomaly	S-cone defect	0.01%

II. PROPOSED SYSTEM

Front End System: It is built keeping the end user in mind wherein the colorblind can log into the system, load the media he/she wishes to recolor and according get the output of the loaded image or video above. The Steps for the working of the interface are as follows:

1. User logs into the system.
2. User enters the image or video he/she wishes to recolor.
3. The algorithm will be loaded and applied to the image.
4. Processing will be performed on image and video.
5. Finally the recolored image and video will be presented to the end user.
6. Finally apply the re-coloration process using the matrices predefined with threshold value.
7. If video is uploaded then it is first divided into frames and then the same steps of color detection , application of RGB to LMS strategy to every individual frame.
8. Finally, the frames are merged to form a single video as a whole and the recolored video is given as output to the color blind user.

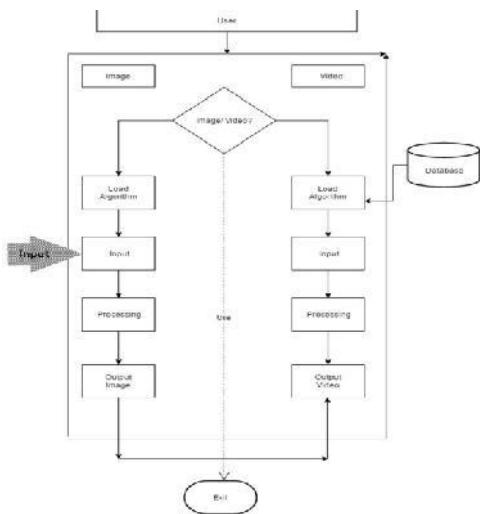


Figure:1 Block diagram of the System

III. Implementation Algorithm

Step 1: Convert the video into individual frames. Since video file has a number of frames. It can vary from a hundred frames to even thousand frames. We can operate on the video all at once. Hence, we need to operate on single frame basis. Therefore, we convert the video into individual frames to process them.

Step 2: Extract the RGB color space of a frame. Each frame has various attributes associated with it. Attributes such as frame size, height, width, and the pixel values. By default, the frame has RGB color space. Hence, we extract that to process it and modify the colors.

Step 3: Convert the RGB into LMS color space.

$\text{rgb2lms} = \dots$

Colorblindness is associated with cones of our eyes. The cones such as Long Median or short are affected. Hence it is easier to work in LMS workspace. We convert from RGB to LMS color space.

Step 4: Transform to colorblind LMS values. We are working on people with Tritanomaly. They have S cone absent. The matrix lms2lmst converts the LMS matrix to the new LMS matrix which shows us how the person with Tritanopia sees the image.

Step 5: Transform the new LMS values back to RGB Values. After the processing we have to reconvert the frame from LMS space to RGB space.

$\text{newRGB} = \text{lmst} * \text{inverse}(\text{rgb2lms})$

Step 6: Calculate the error between original RGB and new RGB. Now we have the new RGB values we calculate the error. Error is calculated by subtracting the new RGB value from old RGB value.

Step 7: Modify the Error for people with Tritanopia. $\text{err2mod} = \dots$

Once the error is calculated we know what parts the affected people cannot see. We modify these errors so that people with tritanomaly can easily distinguish colors.

Step 8: Add the error to original RGB. Now we add the corrected error to the original image. This will keep the original image same and modify the colors which earlier people with tritanomaly couldn't distinguish earlier.

Step 9: Merge the individual frames to finally make the new video. Once we have processed all the frames individually we merge them to get the output. The output is stored in AVI file.

END.

This whole process can be summarized as:

This program takes in media files and first makes the RGB to XYZ conversion. Then, subtracting this from the original image, we find the information lost when the image is seen by a trichromat. We then make a transformation on the error function so as to map it to something that could be perceived by a trichromat. Finally, previously invisible details become visible

IV. RESULTS

Time to run video in seconds

Table II : Performance Table

Length of Video	Frames per second of video				
	12 fps	15 fps	30 fps	60 fps	
4 s	800	1200	1600	2200	
6 s	1200	1600	1800	2400	
8 s	1600	1800	2000	2800	

In terms of video processing, the video is divided into frames. Each frame is then processed on the same algorithm of the image processing for the respected type of color blindness. These recolor frames are then combined to form the video. The resulting video appears vivid to CVD.

1. Balloons:

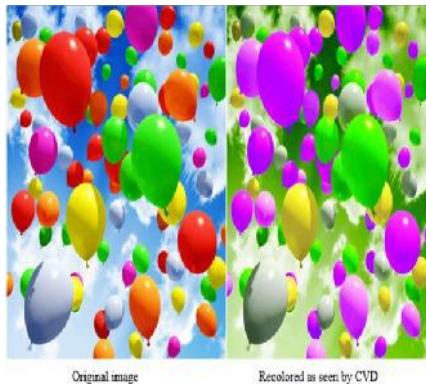


Figure 2 : Comparative view of image processing

2. Color Spectrum

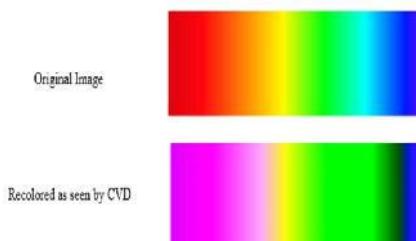


Figure 3: Comparative view of image processing 1



Figure 4: Comparative view of image processing 2

3. Tibet

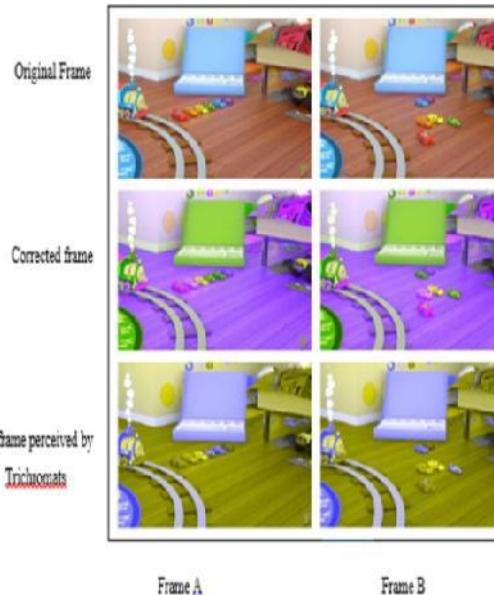


Figure 5: Recolor on frames

Results of Survey

Table III : Survey results

	Worst	No Difference	Better	Much Better
Balloon	0	1	2	4
Cars	1	0	1	5
Tibet	0	0	3	4

Our system was tested by color blind people and we gathered the results. Each respondent was asked to give rating to the new recolored media file. Worst means the condition if image deteriorated further. Some people saw no changes in the new and original media. Where as significant number of respondents found our system to be

better. The people who responded much better were able to distinguish the colors which they were unable to do so before. The survey was conducted online using social media.

V. CONCLUSION

We have studied the details about tritanomaly and designed a visual system for Color Vision Deficient people. Our system takes input in the form of image or video, processes it and give the output. This system can be implemented real time and used by various social medias and online websites such as Youtube and Facebook.

REFERENCES

- [1]Gustavo M. Machado, Manuel M. Oliveira, Member, IEEE, and Leandro A. F. Fernandes, "A Physiologically-based Model for Simulation of Color Vision Deficiency," IEEE Transactions on Visualization and Computer Graphics, November/December 2009.
- [2]Yong Man Ro, Seungji Yang (2004), "Color Adaption for Anomalous Trichromats", International Journal of Imaging Systems and Technology 14(1):16 - 20 · January 2004.
- [3]Seungji Yang, Yong Man Ro, Edward K. Wong and Jin-Hak Lee, "Color compensation for anomalous trichromats based on error score of FM-100 Hue Test," Proceedings of the 2005 IEEE Engineering in Medicine and Biology, 27th Annual Conference, China, September 2005.
- [4]J.P.Srividhya, P.Sivakumar and Dr.M.Rajaram, "The Color Blindness Removal Technique in Image by using Gradient Map Method," Proceedings of 2011 International Conference on Signal Processing,
- [5]Pankaj Kumar Nigam; MahuaBhattachary, "Color vision deficiency correction in image processing," 2013 IEEE International Conference on Bioinformatics and Biomedicine.
- [6]Hyun-Ji Kim; Jae-Yun Jeong; Yeo-Jin Yoon; Young-Hyun Kim; Sung-JeaKo, "Color modification for color-blind viewers using the dynamic color transformation," 2012 IEEE International Conference on Consumer Electronics (ICCE)
- [7]Manuel M. Oliveira, "Towards More Accessible Visualizations for Color-Vision-Deficient Individuals," Computing in Science & Engineering 2013
- [8]Image Processing with Color Compensation using LCD Display for Color Vision Deficiency, Journal of Display Technology, September 2015.
- [9] Chung-Rong Huang, Member, IEEE, Kuo-Chuan Chiu, and Chu-song Chen, Member, IEEE "Temporal Color Consistency-Based Video Reproduction for Dichromats" IEEE, Vol.13, NO.6, OCTOBER 2011.
- [10] S.M.W.Masra, A.A.M.A Shafiee, and M.S.Muhammad "Color Blind Image Correction" University Malaysia Sarawak (UNIMAS).
- [11]Vischeck <http://www.vischeck.com>
- [12]Daltonize. <http://www.daltonize.com>
- [13]<http://www.colorblindness.com>
- [14]https://en.wikipedia.org/wiki/Color_blindness
- [15]<http://www.colorvisiontest.com>
- [16]<http://www.webexhibits.org>
- [17]<http://www.testingcolorvision.com>
- [18]https://nei.nih.gov/health/color_blindness/facts_abot
- [19]https://en.wikipedia.org/wiki/Color_blindness

NOSQL DB with Python PaaS

MirzaZainab
JNU,Manipal University
Jaipur, India
Mirza_zainab@yahoo.com

V.S Dhak
Jaipur,India
vijaypal.dhaka@gmail.com

Abstract-NOSQL technology gives solution to scalability and performance issues associated with SQL Database. All Cloud providers are now a day's providing variety of NOSQL based services and storages. These language-agnostic Cloud platforms support multiple Programming languages, Databases and Framework. This paper compares two NOSQL based services including Big data Analytics services provided by Google Cloud providers and Amazon AWS with Python Programming language.

Keywords: *NOSQL; Big-Data; PaaS; Python*

I. GOOGLE- NOSQL SERVICES

Datastore -Google Cloud Datastore (Cloud Datastore) is a highly scalable, fully managed NoSQL database service offered by Google on the Google Cloud Platform. Cloud Datastore is built upon Google's Bigtable and Megastore technology. Originally released as a feature in Google App Engine in 2008, Cloud Datastore was announced as a standalone product in 2013 during Google I/O [1]. Cloud Datastore can be used for any E-commerce application for example to store product Catalog with various attributes / details of the product. Similarly any social Networking and gaming application, which requires to store user's profile and profile grows based on user's past activities. In such situation Datastore DB can be sued. Moreover, a NoSQL Datastore is able to accept all types of data – structured, semi-structured, and unstructured – much more easily than a relational database. For applications that have a mixture of data types, a NoSQL database is a good option [2]

Features of Google DATASTORE:

- Rich Admin dashboard
- Multiple Access Method
- Fully Managed
- Diverse Data Types
- ACID transactions

Google Datastore is type of NOSQL Document store DB, some document store is simply key-value DB. Key –value DB contain semi structured Data. This DB can be sorted, values can be retrieved. It contains row / entities/items and attribute can vary for different rows in a

same table.

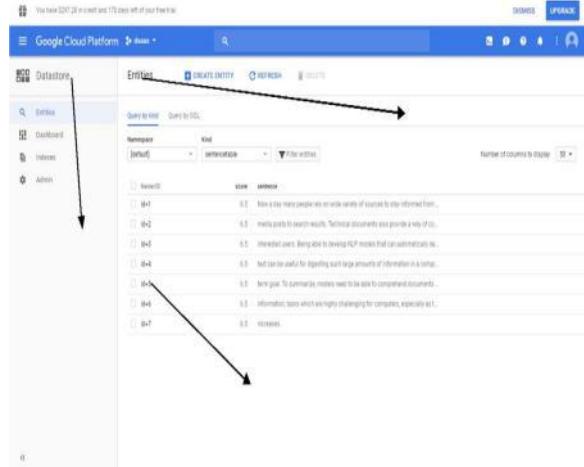


Fig.1:Google Datastore

Table 1: Comparison of Google Cloud DataStore with Relational DB

Name	Google Datastore	Relational DB
Document	Automatically scaling, Versioned Database as a Service (PaaS) on the Google Cloud Platform	A relational database in a unit of tables containing data fitted into predefined categories
Category of one object	Entity or Entity	Table
Individual object	Properties of object	Field column
Value of one attribute	Key	Primary key
Schema	Schema test	Fixed Schema
Document ID	None	NTT
Document	None	Java
Document	None	PHP
Document	None	Python
Document	None	MySQL

Document contains a description of data types and its values. Entities can be grouped together based on attributes/keys so it gives fast query response to search documents with particular attribute. Figure 1 shows screenshot of Datastore with Entities having key –value. Table1 shows the terminology associated with Document data store.

II. GOOGLE NOSQL DATASTORE WITH PYTHON

Using python Platform with Google Datastore is very simple and quick to establish the connectivity. Once Connectivity is established irrespective of whether a development environment is on local machine or cloud based, it is very convenient to read, update and create entities in Data Store. Python is very strong in handling the databases , it supports many compound data types like dictionary ,List,tuple , Set etc. It support many existing SQL based DB as well as Cloud based SQL/ NOSQL DB. Google cloud SDK and Google Cloud Client Libraries need to be downloaded for python.

III. DATASTORE CONNECTION STEPS WITH PYTHON

1. Go to Google Cloud platform -->Products-->and select Datastore-->create a new table in Datastore (tables are referred as 'kind' in Datastore.)
2. Configure API keys for your project to access Datastore APIs
 - a. Go to the Cloud Platform Console.
 - b. From the projects list, select a project or create a new one.(copy project id for future use).
 - c. If the APIs & services page isn't already open, open the left side menu and select APIs & services.
 - d. On the left, choose Credentials.
 - e. Click Create credentials and then select API key
3. Adding restrictions to API keys. Keys are unrestricted by default
4. To add a restriction:

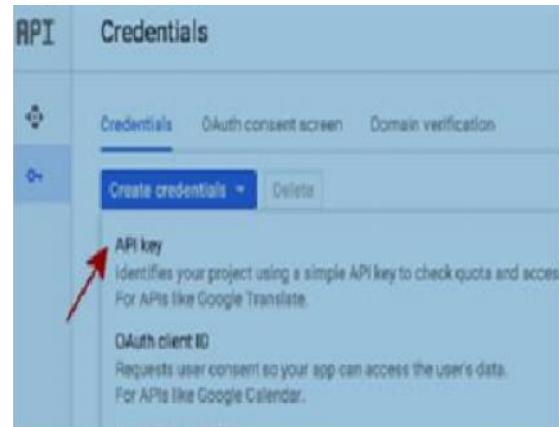


Fig. 2: Datastore Credentials

Click Restrict key within the API key created dialog box. The API key configuration panel appears:

Once Datastore is configured, it is automatically initialized and connected to Google APP Engine. Now we need GoogleAPI's for Datastore in python. Import

DataStore on python shell by following command:

```
>>>import googledatastore as datastore
>>>datastore.set_options(project_id='project-id')
>>>req=datastore.BeginTransactionRequest()
>>>datastore.begin_transaction(req)
```

OR import in editor by using following Command

```
from google.cloud import datastoreclient =
datastore.Client()
key = client.key('id')
entity = datastore.Entity(key=key)
entity['score'] = 'Your score' entity['level'] = 2
client.put(entity)
```

IV. GOOGLE NOSQL BIGTABLE

BIGTABLE - Bigtable is a compressed, high performance, and proprietary data storage system built on Google File System, Chubby Lock Service, SSTable (log-structured storage like Level DB) and a few other Google technologies. On May 6, 2015, a public version of Bigtable was made available as a service [1]. Cloud platforms are language agnostic and support different programming Languages like, Dot NET, JAVA, PHP, Python, Ruby, etc. Cloud is very effective for the Cloud and Big data analytics. E-commerce web site, Social networking and IoT is generating a huge amount of data in every second. This data is unstructured, comes in different format and from different sources. Python can be used effectively in fetching

live stream data.

Extensive list of Libraries' are available for working with Stream data. Google big table is very powerful NOSQL based service, and used by Google for many of its services like search engines, map, analytics and gmail etc. Bigtable can be accessed through Python by installing Native Bigtable API's or Hbase compatible API.

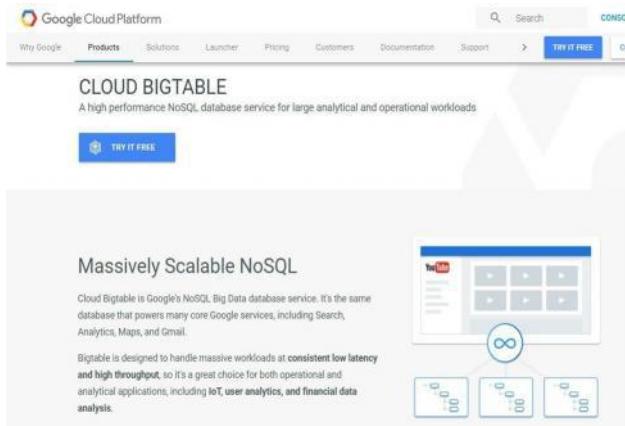


Fig. 3: Google Bigtable
Features of Google BIGTABLE

- Massive scalable NOSQL
- High Performance
- Security and Permission
- Low Latency Storage
- Fully managed
- Redundant Auto scaling Storage
- Seamless cluster resizing
- HBase Compatible
- Global Availability

V. AWS NOSQL DYNAMODB

AWS supports variety of NOSQL databases like Columnar DB, Graph DB and in memory key-value stores. These DB scales horizontally and its Cloud based so availability, scalability and performance and reduced cost are advantages. Performance of thses NOSQL DB are dependent on Cloud Clustersize, Network Latency etc. AWS provides a rich set of API to communicate with NOSQL based infrastructures. These NOSQL DB uses Partition –key for load balancing and Performance.

Amazon DynamoDB is available, fully managed NOSQL document DB. This Database supports both document and key value data model. This is best suited for developer making any E-commerce application, web applications or gaming application, where application setting and users profile grows as application advances. Now a day's many IoT based applications are using NOSQL based cloud storage. Table 2 Compares Amazon DynamoDB with Google DataStore.

Table 2: Datastore V/S DynamoDB

Description	Google DataStore	Amazon DynamoDB
service	Hosted, available , scalable DB by Google	Hosted, Available, scalable DB by AWS
DB Model	Document store Key-value Store	Document store Key-value Store
Data schema	Schema Free	Schema Free
Python	Python API available	Python API available
SDK	Google sdk for Python	AWS SDK for Python

VI. AWS NOSQL DYNAMODB

Boto is the Amazon Web Services (AWS) SDK for Python, which allows Python developers to write software that makes use of Amazon services like S3 and EC2. Boto provides an easy to use, object-oriented API as well as low-level direct service access[3] Its very simple to integrate Python applications with AWS service like s3, DynamoDB .Steps for using DynamoDB with Python

- Create an account on AWS
- Use AWS Access keys
- Download AWS Sdk for Python Use following instructions in python editor

```
import boto3
# Get the service resource.

dynamodb = boto3.resource('dynamodb')    table = dynamodb.Table('users')
```



Fig. 4: AWS DynamoDB

VII. CONCLUSION

Google Datastore and Bigtable both fully managed NOSQL , scheme free Tables provided by Google.DynamoDB provided by AWS as one of the NOSQL DB.Depending upon the requirement of the project it becomes very convenient and cost effective to use this available service with Python Platform.

REFERENCES

- [1] <https://cloud.google.com/datastore/docs/concepts/overview>
- [2] <https://cloud.google.com/bigtable/docs/>
- [3] <https://boto3.readthedocs.io/en/latest/>
- [3] V. S. Dhaka and M. Zainab, "Big Data Management as a Service (BDMaas) using NOSQL model from Cloud," no. 3, pp. 18–22, 2016.
- [4] B. Y. D. Corporation, "Why NoSQL?," no. October, pp. 1–10, 2013.
- [5] EMC Corporation, "Big Data-as-a-Service," no. July 2012, p. 16, 2012
- [6] V. S. Dhaka and M. Zainab " Big Data Analytics as a Service (BDAaaS) system layer"INDIACOM.2017 IEEE Conference

Real-Time Location Detection Using Face Recognition Through CCTV Cameras/Mobile Camera

Huzaifa Saboowala

Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Mufaddal Kapadia

Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Mustafa Poonawala

Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Farhana Siddiqui

Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Abstract-Face recognition allows identification and targeting of individuals of interest from faces captured across a network of video cameras. In such applications, face recognition is challenging because faces are captured under limited spatial and temporal constraints. Facial models for recognition are designed as reference samples in order to compare them with faces captured by the camera(s).In the past few years, many real time systems have been developed. These systems carry out face recognition by performing image processing on human faces and produce throughputs according to the purpose.The development of human location detection and tracking system has been progressing for several years. Detecting locations using face recognition is a very active research area specializing on how to recognize faces within images or videos and detecting the location of the individual. Our goal is to create a portable, user-friendly, low-cost system using various advanced technologies like Principle Component Analysis (PCA).

Keywords- Face Recognition, Location Detection, Image Processing, Low-cost, Principle Component Analysis.

I. INTRODUCTION

Generally, face recognition is a big challenge as there is a chance of high uncertainty due to certain external light conditions, so we are taking the advantage of grayscale images and Principle Component Analysis (PCA), which are less effected to the external environment changes^[1]. Prior step of face recognition involves face detection of the individual which is also a big challenge.

For this we are taking help of pre-designed cascades whose detection of objects is satisfactory. OpenCV (Open Source Computer Vision Library) [1] which is a

library of programming functions mainly aims at real time computer vision. It focuses mainly on real-time image processing. OpenCV can support all the image and signal processing algorithms which can be ported onto the Android platform very easily. The major applications of OpenCV include 2D & 3D feature extractions, ego motion estimation, facial recognition system, gesture recognition, human computer interaction, mobile robotics, motion understanding, object identification, motion tracking etc.

Intelligent human recognition and location detection system is integrated with a lot of hardware modules such as image and video capture, stored facial models for reference, GPS positioning, etc. The design of the system software uses embedded software developing platform based on keil IDE and Android environment. By the hardware/software co-design, the new intelligent human recognition and location detection system will be implemented successfully.

II. EXISTING SYSTEM

The current existing system(s) follow the approach for 3D face recognition which involves size variation that makes use of a purely curvature-based representation. They can handle size change between faces, but run into problems with change in facial expressions when comparing between captured image and the stored reference image. This is one of the major drawbacks of existing system(s). Face recognition is done with PCA-based 2 dimensional (2D) and 3 dimensional (3D) algorithm. The main problem in experimental validation and comparison of 3D face recognition is lack of appropriate datasets.

Certain desirable properties of such dataset include:
1. A large number and demographic variety of people represented.

-
- 2. Images of a given person taken at repeated intervals of time.
 - 3. Images of a given person that represent substantial variation in facial expression.
 - 4. High spatial resolution, for example, depth resolution of 0.1 mm or better.
 - 5. Low frequency of sensor-specific artefacts in the data.

III. LITERATURE SURVEY

Authors in [1] mainly address the building of face recognition system by using Principal Component Analysis (PCA). PCA is a statistical approach used for reducing the number of variables in face recognition. In PCA, every image in the training set is represented as a linear combination of weighted eigenvectors called eigen faces. These eigenvectors are obtained from covariance matrix of a training image set. The weights are found out after selecting a set of most relevant eigen faces. Recognition is performed by projecting a test image onto the subspace spanned by the eigen faces and then classification is done by measuring minimum euclidean distance.

System proposed in [2] provides a new framework for providing robust location detection in emergency response systems, based on the theory of identifying codes. The key idea of this approach is to allow sensor coverage areas to overlap in such a way that each resolvable position is covered by a unique set of sensors. In this setting, determining a sensor-placement with a minimum number of sensors is equivalent to constructing an optimal identifying code, an NP-complete problem in general.

System proposed in [3] presents a comprehensive and critical survey of face detection algorithms. The algorithms presented in this paper are classified as either feature-based or image-based and are discussed in terms of their technical approach and performance. Due to the lack of standardized tests, a comprehensive comparative evaluation is not provided, but in cases where results are reported on common datasets, comparisons are presented. System proposed in [4] tries to implement an AFR system based on eigen face method where noise filtering is also added. It also provides comparative study between systems and tries to explain some pre-processing difficulties (illumination, pose variation, image quality etc.) and its explanation for face recognition system.

System proposed in [5] makes use of Radio Frequency Identification (RFID) tags. These RF-ID tags can be made much smaller and cheaper than ultra-wideband nodes, and they can be attached to just about anything. Each tag contains an antenna and a simple integrated circuit. It sends identification data in response to a query from a reader or scanner at distances of centimetres to meters. The tags can be passive, borrowing energy from the scanner's RF signal to send their reply, or they can be active, with battery power and the ability to transmit their own signals. The tags use security codes and encryption.

System proposed in [6] performs location detection with the help of mobile stations. The mobile station measures signal strength of each of a set of signals associated with the mobile station, then determines whether a number of signals in the set is less than a predetermined threshold. Based on the determination, the mobile station calculates a difference in power between a first and second signal associated with the mobile station. Thereafter, power on a channel corresponding to the mobile station is increased for a predetermined number of frames based on the calculation. The location of the mobile station is determined during the power increase.

The proposed system in [7] is based on performing real-time position detection and motion tracking of mobile communications devices moving about in a defined space comprised of a plurality of locales is provided. A plurality of access points are disposed about the space to provide an interface between mobile devices and a network having functionality and data available or accessible therefrom. Knowledge of adjacency of locales may be used to better determine the location of the mobile device as it transitions between locales and feedback may be provided to monitor the status and configuration of the access points.

The system in [8] proposes a technique to give a high percentage of face detection based on correlation and post-processing. The first module concentrates on the approach used in solving this problem, namely: correlation, false/repeated hits removal techniques, color segmentation and multi resolution approach. The second module includes the actual implementation, where results will be obtained after performing the algorithm on training images.

IV. FACE RECOGNITION

Images of faces, represented as high-dimensional pixel arrays, often belong to a manifold of lower dimension. In statistical approach, each image is represented in terms of d features. So, it's viewed as a point (vector) in a d-dimensional space. The dimensionality -number of coordinates needed to specify a data point- of this data is too high. Therefore, the goal is to choose and apply the right statistical tool for extraction and analysis of the underlying manifold. These tools must define the embedded face space in the image space and extract the basis functions from the face space. This would permit patterns belonging to different classes to occupy disjoint and compacted regions in the feature space. Consequently, we would be able to define a line, curve, plane or hyper plane that separates faces belonging to different classes. Many of these statistical tools are not used alone. They are modified or extended by researchers in order to get better results. Some of them are embedded into bigger systems, or they are just a part of a recognition algorithm. We will be using Principal Component Analysis (PCA) to perform face recognition.^[9]

One of the most used and cited statistical method is the Principal Component Analysis (PCA). It is a mathematical procedure that performs a dimensionality reduction by extracting the principal components of the multi-dimensional data. The first principal component is the linear combination of the original dimensions that has the highest variability. The n^{th} principal component is the linear combination with the maximum variability, being orthogonal to the $n-1$ first principal components. The idea of PCA is illustrated in Fig 1. The greatest variance of any projection of the data lies in the first coordinate. The n^{th} coordinate will be the direction of the n^{th} maximum variance - the n^{th} principal component. Usually the mean \bar{x} is extracted from the data, so that PCA is equivalent to Karhunen-Loeve Transform (KLT). So, let $X_{n \times m}$ be the data matrix where x_1, \dots, x_m are the image vectors (vector columns) and n is the number of pixels per image. The KLT basis is obtained by solving the eigenvalue problem

$$C_x = \Phi \Lambda \Phi^T [9]$$

where C_x is the covariance matrix of the data.

$$C_x = \frac{1}{m} \sum_{i=1}^m x_i x_i^T [9]$$

$\Phi = [\Phi_1, \dots, \Phi_n]$ is the eigenvector matrix of C_x . Λ is a diagonal matrix, the eigenvalues $\lambda_1, \dots, \lambda_n$ of C_x are located on its main diagonal. λ_i is the variance of the data projected on Φ_i . PCA can be computed using Singular Value Decomposition (SVD). The SVD of the data matrix $X_{n \times m}$ is

$$X = UDV^T [9]$$

It is known that $U = \Phi$. This method allows efficient implementation of PCA without having to compute the data covariance matrix C_x -knowing that $C_x = U^T X$. The embedding is done by $y_i = U^T x_i$ thus obtaining the mapped points $y_1, \dots, y_{m \cdot n}$

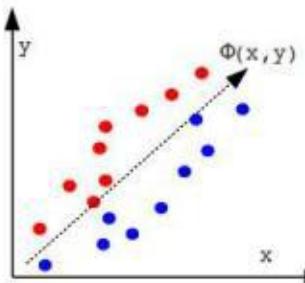


Fig 1. Principal Component Analysis (PCA)

V. METHODOLOGY

The overall objective is subject to the success of three phases within the system:

Phase 1:

Phase 1 is also known as the authorization phase. In this phase, the user logs in to the system (registers first if he/she is a new user) using their credentials. Based on these credentials, we can classify the types of users and grant access to view locations of individuals accordingly.

Phase 2:

Phase 2 consists of portable system with Android Operating System including OpenCV library, USB and serial port to perform the image processing part. Initially using a USB camera interface (Raspberry Pi), continuous images are captured by segmenting the live video into frames and these images are processed with help of OpenCV and compared with existing database. If the current images are matching with any of the existing images, the system generates command to the Phase 3.

Phase 3:

Phase 3 will perform the location identification using GPS and forward the necessary information about the identified individual using GPRS to administrator. The administrator updates the new location in the database. Thus, the current location of the individual is displayed to the user. In addition, it will predict the path of the individual and alert those cameras that are placed on that path.

The block diagram shown in Fig 1 explains the phases efficiently.

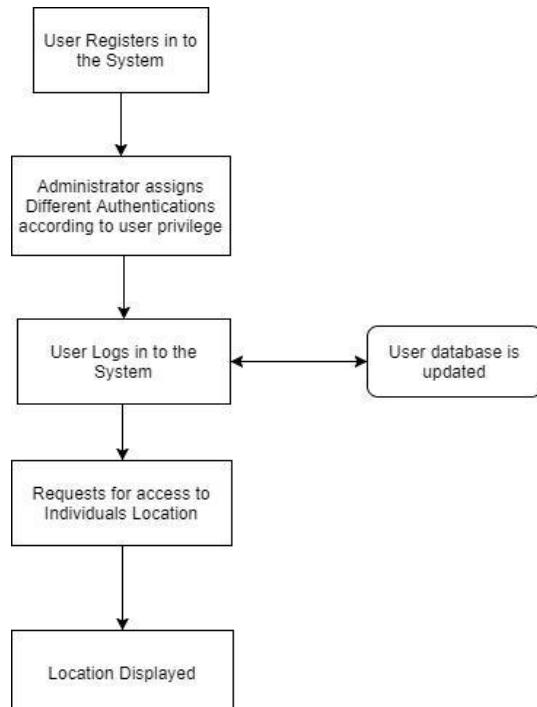


Fig 2: Block Diagram for the proposed system

Technical efficiency: The effectiveness with which the given system produces the output, it calculates the amount of space and time required by the system to produce the desired output.

Economic efficiency: The total cost required for implementation of the system and the extent to which

it minimizes the inefficiency, given the amount of resources.

Management efficiency: The ease with which the system can be handled or maintained. For example, the proposed system reduces a large amount of paper work.

Table I: Efficiency Ttable

Parameter	System based on ontology	System based on affinity model	Traditional system	Proposed system
Technological Efficiency	Medium	Medium	Poor	Very Good
Economic Efficiency	Good	Good	Medium	Very Good
Management efficiency	Medium	Good	Poor	Excellent

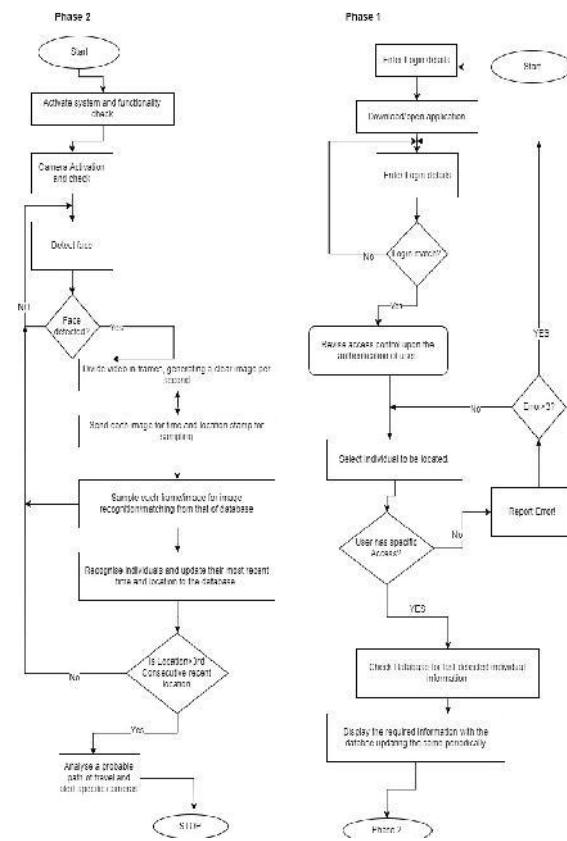


Fig 3: Implementation flowchart of the proposed system

Algorithm:[9]

Let a face image $\Gamma(x, y)$ be a two dimensional $M \times N$ array of intensity values. In this thesis, I used a set of image by 200×149 pixels. An image may also be considered as a vector of dimension $M \times N$.

Step 1: Prepare the reference faces

Obtain face images $I_1, I_2, I_3, I_4, \dots, I_M$ (reference faces). The face images must be of the same size.

Step 2: Prepare the data set

Each face image I_M in the database is transformed into a vector and placed into a training set S .

$$S = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \dots, \Gamma_M\}$$

Each image is transformed into a vector of size $MN \times 1$ and placed into the set. For simplicity, the face images are assumed to be of size $N \times N$ resulting in a point in N^2 dimensional space. An ensemble of images, then, maps to a collection of points in this huge space.

Step 3: Compute the average face vector

The average face vector (Ψ) has to be calculated by using the following formula:

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n$$

Step 4: Subtract the average face vector

The average face vector Ψ is subtracted from the original faces T_i and the result stored in the variable Φ_i ,

$$\Phi_i = \Gamma_i - \Psi$$

Step 5: Calculate the covariance matrix

We obtain the covariance matrix C in the following manner,

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T$$

Step 6: Calculate the eigenvectors and eigenvalues of the covariance matrix

The covariance matrix C in step 5 has a dimensionality of $N^2 \times N^2$, so one would have N^2 eigenfaces and eigenvalues. Computationally, this is not very efficient as most of those eigen faces are not useful for our task. In general, PCA is used to describe a large dimensional space with a relative small set of vectors.

Compute the eigen vectors u_i of AA^T

Step 6.1: Consider the matrix

$$L = A^T A \quad (M \times M \text{ Matrix})$$

Step 6.2: Compute eigen vectors v_i of $L = A^T A$

$$A^T A v_i = \lambda_i v_i$$

Step 7: Keep only K eigen vectors (corresponding to the K largest eigen values)

Eigen faces with low eigenvalues can be omitted, as they explain only a small part of characteristic features of the faces.

Step 8: Parse the location information to the application.

Step 9: Update the location of the recognized individual.

VI. CONCLUSION

Our proposed system can create an impact at a local as well as at a global level with room for modifications. Further modifying the system using the concept of Artificial Intelligence, Data Structures, Machine Learning etc., may also enhance the system by creating a map which demonstrates the activities related to the individuals which may be linked to the database. At a local level, this system can be used in public schools, colleges in order to keep a track of the teacher's/student's whereabouts. At a global level, this system can be used in bank chains, hotel chains, corporate offices, transport sector, police stations and much more. It can bring security to a whole new level.

REFERENCES

- [1] N Papa Rao, G Bhavana, M Lakshmi Ravi Teja "RTOS Based Image Recognition & Location Finder Using GPS, GSM and OpenCV" - International Advanced Research Journal in Science, Engineering and Technology Vol. 2, Issue 12, December 2015.
- [2] S. Ray, R. Ungrangsi and De Pellegrini "Robust location detection in emergency sensor networks"-INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies
- [3] ErikHjelmås¹,Boon Kee Low "Face Detection: A Survey"- Survey" (Department of Informatics, University of Oslo, 23rd October 2000).
- [4] Ferdous, Shariful Islam, Ahmed, Sabir "Literature survey of automatic face recognition system and eigenface based implementation"- Department of Computer Science and Engineering, BRAC University
- [5] "They know where you are" – Jay Warior, McHenry and Kenneth McGee. IEEE Spectrum July 2003.
- [6] Timothy J Neher "Personal location detection system" - Patent application Ser. No. 09/536,009 March 26, 2000.
- [7] Jun Miao, Baocai Yin, Kongqiao Wang, Lansun Shen, Xuecun Chen "A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-centre template"-Department of Computer science, Beijing polytechnic university, 28th October 1998.
- [8] Face detection through template matching and color segmentation. By Scott Tan Yeh Ping (yptan@stanford.edu) Chun Hui Weng (cwengc@stanford.edu) Boonping Lau (bppower@stanford.edu) (EE 368 Final Project)
- [9] Face Recognition Algorithms by Proyecto Fin de Carrera, Ion Marqués Supervisor: Manuel Graña. June 16, 2010.
- [10] Liton Chandra Paul , Abdulla Al Sumam " Face Recognition Using Principal Component Analysis Method" - International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 9, November 2012
- [11] Roberto M. Cesar, Jr., EndikaBengoetxea, Isabelle Bloch, Pedro Larrañaga "Face Recognition Algorithms and techniques"- Journal Pattern Recognition archive Volume 38 Issue 11, November, 2005
- [12] "Face Recognition Algorithms" University of Basque Country, June 16th 2013.
- [13] Face Recognition Algorithms and techniques: <https://facedetection.com/algorithms/>
- [14] Libraries for face detection available at: <https://opencv.org/>
<http://dlib.net/>

Analysis of Real Time Surveillance System on Hadoop Image Processing Interface

Er. Mohammed Ahmed Abdul Mannan
Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Shehbaz Ansari
Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Shamuwel Ahmed Ansari
Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Sayed Adnan Husain
Computer Engineering Department,
M.H. SabooSiddik College of Engineering,
Byculla, Mumbai, India

Abstract - Traditional security systems works to avoid crimes as much as possible. Real time Surveillance gives an opportunity to prevent any criminal activity before they can happen. Implementing security measures are also very complex and takes a lot of time and also requires human interference. An autonomous security system will make security economically feasible and it works quickly. Using facial, object and behavior recognition on the input feed provided by CCTV cameras, various criminal activities can be detected, and authorities will be assisted to take desired action. Covering large number of CCTV's distributed over wide space can generate lots of data and requires tremendous processing power to process this data. Hence, we will use "Hadoop's image processing interface (HIPI)" to distribute the processing task over the cloud network, so communication between authorities of various areas is enhanced.

Keywords - HIPI, OpenCV, Tensorflow, Ffmpeg, CCTV.

I. INTRODUCTION

In the current time, at almost all locations, the security systems works in a rather passive way. CCTV cameras installed in these system record videos and feed them to a human supervisor. Such a security system is vulnerable to human errors. Quick actions are not possible which are necessary in many conditions to prevent adversary. The entire security works locally and provides with limited cloud capabilities. Such a static system is outmoded and itself is under security threat of being misused and hacked. Hence we presented a modern, dynamic system with capabilities to work in cloud with powerful real time surveillance and arguably cheaper than existing system. Footages from numerousCCTV cameras will reach to a local station. These video feed will be provided to a preliminary object recognition algorithms and will undergo the stage of culling in the local station.

After the initial process of object recognition, the output, separated into small unit, will comprise of multiple images. This images will be mapped to the respective nodes for processing and their results will be reduced to get the final output.

II. LITERATURE SURVEY

The Authors in [1] proposed a scalable video processing system over the Hadoop network. The system uses FFmpeg for video coding and OpenCV for Image processing. They also demonstrate a face tracking system, which grouped multiple images of the same people together. Video feed captured is stored in the "Hadoop Distributed file system (HDFS)". The system does not states proper security mechanisms and storing such huge data in the HDFS will not be cost efficient. System in [2] used "Nvidia CUDA" enabled Hadoop clusters to improve server performance by using the parallel processing capability of CUDA cores present in Nvidia GPU's. They demonstrated an Adaboost-based face detection algorithm in the Hadoop Network. Although equipping the clusters with Nvidia GPU's might increase the cost of clusters, CUDA cores potentially provides massive improvements in Image processing jobs. Although we aim to implement the system into existing hardware to minimize the cost. The system in [3] used the Hadoop Framework to process astronomical Images. They implemented a "scalable image-processing pipeline" over Hadoop, which provided for cloud computing of astronomical Images. They used an existing C++ Library and JNI to use that library in Hadoop for Image processing. Although they achieve success, many optimizations were not made and Hadoop was not Integrated properly with the C++ Library.

Survey in [4] describes various security services provided in the Hadoop Framework. Security services, which are necessary in the framework such as Authorization, Access Control and Integrity, are discussed including what Hadoop provides and what it does not. Hadoop has multiple security flaws which can be exploited to initialize a replay attack or view the files stored in the HDFS node. Hence as per the scholarly, a good Integrity check method and Authorization control method is necessary.

The object recognition stated in [5] provides an efficient way in recognizing a 3-Dimensional Object from a 2-Dimensional Image. In his stated methodology, “certain features of the object remains constant regardless of the viewing angle”. Extracting these features specifically will save tremendous amount of resources as compared to the older object recognition systems that recreate the entire 3-D objects using Depth Analysis.

Behaviour Recognition can be carried out as stated in [7]. The features will be extracted from video feed and applied to feature descriptors, model events and Event/behaviour models. The output will be mapped from feature space to behaviour label space where a classifier will map it as normal or abnormal.

The system proposed in [8] states an economic, reliable, efficient and scalable surveillance system where data is stored using P2P concept. It avoids load on a single Data Centre and divides the load into multiple Peer Nodes. It also provides authentication as a module between the Peer Nodes and the directory nodes. The system doesn't present any method to implement computer vision and

[9] Proposes an open source Hadoop Video processing Interface to integrate C/C++ applications in the Hadoop Framework. It provides R/W interface for developers to store, retrieve and analyse video data from the HDFS. Using the available security in the Hadoop framework for video data can give poor performance and security was not mentioned in the HVPI.

TensorFlow, a ML System, stated in [10], provides multiple tools to implement multiple training algorithms and optimizations for multiple devices on a large scale. It uses dataflow graphs for computation states and operations that changes those states. TensorFlow can work very well with Hadoop Framework to distribute the processing in the existing hardware.

As depicted in [6], the original eigen faces fails to accurately classify faces when the data is coming from “different angles and light sources” like in our problem.

Table1. Analysis Table

REFERENCE	VARIOUS ALGORITHMS AND LIBRARIES USED	DRAWBACK
[1]	FFmpeg as video coder, OpenCV for Image processing.	Image processing done after storing the video feeds in HDFS.
[2]	JCUDA and JNI for Acceleration and Adaboost Face Recognition.	Cost of Nvidia GPU's can be high.
[3]	Own C++ library and JNI for Interface.	No compiler optimizations.
[4]	Homomorphic encryption, one time pad for authentication, Attribute-group for access control.	Encryption will increase the required processing power for video analysis.
[5]	Perceptual Organization, Probabilistic Ranking, Spatial Correspondence.	Cannot be used for hidden object recognition since the lack of depth analysis.
[6]	TensorFace.	Requires Images from many angles and lightning conditions for accurate recognition.
[7]	Modified Altruistic Vector Quantization Algorithm, Gaussian Mixture Model, Dynamic Bayesian Network, and Markov Model.	Works very well in only classifying abnormal and normal behaviour. Further algorithms are required for specific behaviour recognition.
[8]	P2P and Hadoop.	Doesn't provide any kind of recognition.
[9]	JavaCV, JNI	Lacks security.
[10]	Dataflow graphs.	Cloud process distribution.

III.PROPOSED METHODOLOGY

To provide real time recognition various pre-processing is done to improve Hadoop and neural network performance. The entire process can be categorized in the following phases:-

Video Collection: - The input feed coming from the devices like CCTV will be converted to the HipI Image Bundle (HIB) object using various tools like hib Import, hibInfo. After that, HIB will undergo preprocessing using a video coder like Culler class and FFmpeg. In this stage, various user-defined conditions like the criteria for Image metadata can be applied. Filters like greyscale filter, provides improvements for various face detection algorithms. The images surviving the culling phase will undergo the preliminary object detection phase using object detection algorithms like tensor flow or provided by a library like OpenCV. Weapons, Cars, and Humans will be detected in this phase.

The collected Image will be mapped to MapReduce model using the HibInputFormat class. The header will determine the what data to map to the respective data node in the network.

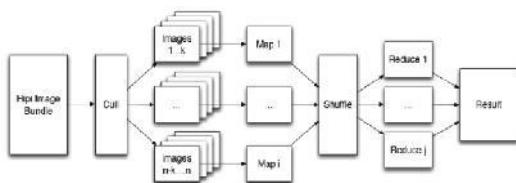


Fig 1. HipI Pipeline courtesy hipi.cs.virginia.edu

Mapping Phase: - Images, which are flagged as humans, will be mapped to the facial recognition and behavior recognition algorithms in the respective data nodes. Images recognized as cars will be mapped to object detection. Various algorithms for recognition in the mapping phase can be derived from OpenCV, which also inherently used Nvidia CUDA and OpenCL for increased performance in the recognition. OpenCV provides Java interface and can be directly used with Hadoop. Although a self developed can be used and if required, will be written in C++ and JNI (Java Native Interface) can used to integrate with hadoop.

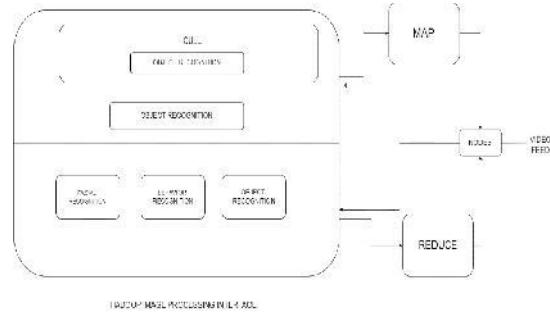


Figure2: Block Diagram

Reduce Phase: - Criminal faces will be detected during facial detection, since the node with the highest confidence value will be declared as winner. Stolen cars will also be detected in the similar fashion. Human behavior will classify and detect specific suspicious behavior.

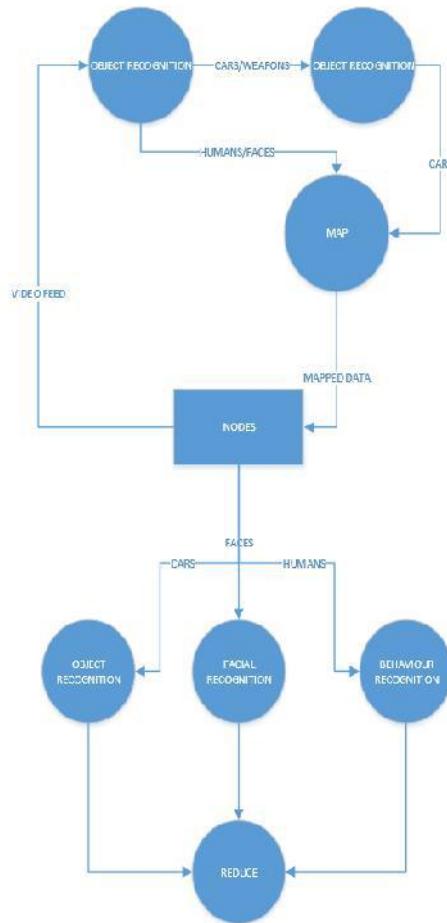


Figure3: Block Diagram of proposed system

IV. CONCLUSION:

Although the above paper only discusses specific applications, the entire architecture is scalable to be implemented in specific environments. The system can find applications in various companies offices, police department and various high security facilities for real time computer vision assistance. The system can be implemented over the existing hardware either as a complement to the existing system or as a substitute to the predefined system. Once enough test samples are collected, various optimizations can be used like different neural networks, more suited to the specific applications. Optimizations can also be made to the JNI to improve further performance. Various pre-processing techniques in the video coder can be applied to improve the neural network performance.

REFERENCES

1. C Ryu, D Lee, M Jang, C Kim(2013). "Extensible video processing framework in apache Hadoop" - IEEE International Conference on Cloud Computing Technology and Science
2. R Malakar, N Vydyanathan "A CUDA-enabled Hadoop cluster for fast distributed image processing, 2013" - Parallel Computing Technologies (PARCOMPTECH), National Conference on. IEEE.
3. K Wiley, A Connolly, S Krughoff "Astronomical image processing with Hadoop, 2011" - Astronomical Data Analysis Software and Systems XX. Vol. 442.
4. MR Jam, LM Khanli, MS Javan , "A survey on security of Hadoop-2014" - Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on. IEEE, 2014.
5. DG Lowe – “Three-dimensional object recognition from single two-dimensional images-1987” – Artificial intelligence 31.3 8: 355-395.
6. MAO Vasilescu, D Terzopoulos – “Multilinear image analysis for facial recognition - Pattern Recognition, 2002”. Proceedings. 16th International Conference on. Vol. 2. IEEE, 2002.
7. OP Popoola, K Wang – “Video-based abnormal human behavior recognition 2012” - IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 : 865-878.
8. YS Wu, YS Chang, TY Juang – “An architecture for video surveillance service based on P2P and cloud computing, 2012” - Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference on. IEEE.
9. X Zhao, H Ma, H Zhang, Y Tang, “HVPI: extending Hadoop to support video analytic applications 2015 - Cloud Computing (CLOUD)”, IEEE 8th International Conference on. IEEE.
10. Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. OSDI, 2016 – “Large-scale machine learning on heterogeneous distributed systems.” arXiv preprint arXiv:1603.04467.

Agricultural Updates via SMS Using Cloud Computing Approach

Rakesh Shirsath

Department of Computer Engineering

Sandip Institute of Technology and Research Center

Nashik, India

rakesh.shirsath@gmail.com

Mandar Shinde

Department of Computer Engineering

Sandip Institute of Technology and Research Center

Nashik, India

mandarshinde681@gmail.com

Mayank Mehta

Department of Computer Engineering

Sandip Institute of Technology and Research Center

Nashik, India

maxmehta007@gmail.com

Yogesh Kolhe

Department of Computer Engineering

Sandip Institute of Technology and Research Center

Nashik, India

yogeshkolhe444@gmail.com

Abstract—India is an agricultural based developing nation. The India's economical benefaction of agriculture to GDP is reducing with the overall economic growth. Some real time challenges that are faced by the Farmers due to much less knowledge about Weather Forecast, Efficient Crop-Production, Sales and regular Market Scenario etc. Cloud Computing is prime factor in agriculture and the most increasing way through which data, applications and resources for computing are provided as kind of service to the Farmer. A Cloud based framework is developed to provide the agricultural updates through SMS (in regional language) that is more useful to farmers to get required information. By considering Data As a Service (DaaS) expert advice for Crop, Fertilizers, Market Scenario etc. is provided to farmers against the database queries. This system will provide easy-to-use interface for both Farmers and concerned agency, update any information on market status of various products regularly.

Index Terms—Cloud Computing, Crop Production, Market Scenario, Weather Forecast.

I. INTRODUCTION

Farmers are the pillars of our nation. Agriculture contributes an vital role of the economical value in India. In the agriculture field, the market is deciding agricultural products value in money. Mostly the Farmers promote their things through village level vendors to local market without knowing their actual price in Indian market. So this brings an economical loss for farmers and also to the customers. Cloud Computing is the most advisable technology due of its cost efficiency and flexibility[1]. Now a days, it is providing technology to make all fields convenient. Here this technology is utilized in agricultural field to design a Cloud Based Framework which provides the updates on cell phone that would be helpful to the farmer to get required information if required. The Cloud Computing is provided with a expert system along with knowledge base and automatic updates, which gives an recognition to Farmers and their products[2]. The networks such as Mobile network, Internet have the emerging power to provide agriculture based information services at affordable, relevant, searchable and up-to-date. The phone technology benefits the customers to get information if required, with physical contactless services. According to the needs of Indian Farmers, various

applications and services are provided for agricultural information[3]. There are different existing systems using SMS which are not that much efficient because farmer get the details which are unrelated to their request also provides limited information. The major challenge in agricultural sector is no storage facility for food which yields to the wastage of agriculture based products resulting loss in the economy of India[4]. So, the easy-to-use interface is provided using mobile phones via SMS to Farmers and updates on market status of various products as demanded[5].

Our Cloud Based Framework is framed to provide the agricultural updates through SMS(in regional language) that is more useful to Farmers for getting required information. By considering Data As a Service(DaaS) expert advice for Crop, Fertilizers, Market Scenario, etc is provided to Farmers against the database queries. Also the enhancement in agricultural sector through Cloud Computing can be achieved with help of a proposed system[6]. This system will provide easy-to-use interface for both Farmers and concerned agency, update any information on market status of various products regularly. The most emerging growth that is happening due to some technology trend and business models[7]. There is expansion in request of Interactive Applications with real time response. Cloud Computing concept has many computers intertwined through the common network like Internet[8]. Cloud computing provides easy to use, on demand, dynamic and steady use of distributed computing resources. It has main characteristics on demand service, very large network access, resource reservation, reliability measured service[9].

In the paper, we will enhance a new system which works on the basis of SMS Controller. The paper is well organized as follows- Section II portrays the literature survey, which is required to easily understand existing work done related to our system. Section III narrates proposed system and explains that how it is different from existing system. In section IV, we explain which possible algorithms can used followed by the conclusion.

II. LITERATURE SURVEY

The work aimed is to provide the characteristics of supply chain, here it is focused on particular short chain. Industries in which this analyzes proceed by the agricultural organizations using available IT innovations by specially considering the cloud. It is providing the role of improving processes and also market information. The concept is induced for short chain. So the overall applications and services require only governing them to check the business needs for cost,

availability and performance agility. Thus, Cloud provides the supportive role by giving flexibility, security and scalability to data management and its applications[1].

The system makes transparency to the farmers, while using the application farmer can access and know about all the food products cultivated in India, simultaneously find the crop advice, weather reports to pretend their crop plantation and upcoming price list. The service will be product and market-specific, which means the farmer can choose accordingly. Also, the farmer will not be loaded with irrelevant information and the updated SMS will be kept as small as possible, to avoid the irrelevant costs. A farmer can easily select the market, as also decide on whether he should hold on his product. In this proposed system, the application is moved into Cloud Computing. The Cloud technology is chosen, as it is a ready to serve high business segment. Cloud Computing helps to configure general purpose, online data center by giving command to support any software application quickly[2].

The application can be accessed by farmer and know about all the food products cultivated in India, simultaneously find the crop advice, weather reports to pretend their crop plantation and upcoming price list. The service will be product market-specific, which means the farmer can choose accordingly. A farmer can pick the market, and also decide on whether he should hold on his product. Now-a-days, the common area of ICT is Cloud Computing as it is providing IT as a service to the cloud users on-demand basis with larger scalability, flexibility and availability. The application of IoT thus provides advantages of pervasive interconnection of devices to be localized, identified and communicated[3].

The suggestions are useful in varied stages of agriculture. It involves production of food, organization of supply chain. Also it will be depending on electronic information. The important feature of this proposed system is that it contains information in different languages, which is very useful. The User can directly select the language version in the Welcome Page itself, so there will be no confusion[4]. Tasks as field definition, task operations, lists and reports all farming use data can be submitted and carried on together in a smart phone at any farm working condition. The Farm Manager is an application which builds the management base to record and browse its tasks, equipments and reports all of them to be performed by the touch of smart phone screen button. The disadvantage is it is applicable only for the small farms if there is any query related to agriculture then the farmer needs to be registered to ask the expert suggestions. The advise officer contains the details of farm land strength. The farmer can view all the details which is included in each and every log-in. At last the fertilizers log-in include what kind of fertilizers used for land as it is based on statistics for previous years data sets matched based on rainfall[5].

All the food products cultivated in India, simultaneously find the crop advice, weather reports to pretend their crop plantation and upcoming price list. The service will be product market-specific, which means the farmer can choose accordingly. The awareness and promotion among the prime stakeholders to acquire the correct and massive information related to agriculture based sector which gives the awareness and have a well established information base for the nation. This will in return lead to a well-connected world[6].

Although Cloud Computing is a new phenomenon it is set as we use Internet. New technologies emerging at rapid rate, each with technology advancement and the potential of making humans lives easier. It facilitates large storage and management to agriculture information at cheaper cost. By applying this technology in agriculture it makes easy to take decision related to crops and lands for farmers. Diverse applications are made and utilized by agriculturists for their particular reason. This applications have diverse utilization according

to its functionalities. Numerous applications are useful for various sort of the cultivating exercises like trimming data, pesticides and data in regards to the accepted procedures of cultivating[7].

III. PROPOSED SYSTEM

In this proposed system, the essential modules are namely as-

- A. Attribute Based Access Controller (ABAC).
- B. Data Filtering and Load Balancing.
- C. Service Module

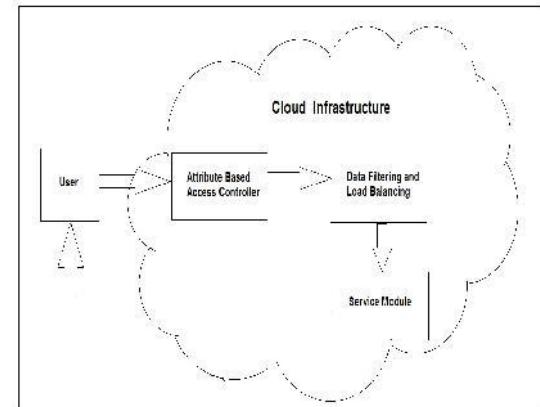


Figure 1. System Architecture

The various components involved are:

A. Attribute Based Access Controller: Attribute Based Access Controller depends upon how to evaluate the available attributes of the object. These operations may include reading, creating, finding, modifying, removing, and executing objects [4]. All Attribute Based Access Controller solutions consist of evaluating attributes and enforce relationships between the attributes. Containing small system, ABAC relies on the assigning the attributes to subjects, objects and the overall development of policy that contains the access rules[5].

B. Data Filtering and Load Balancing: In field of Cloud Computing, the load balancing enhances the supply of workload across different computing resources, as computer, Central Processing Unit etc. Load balancing targets on optimizing resource use, minimize response time. In Data Filtration and Load Balancing first collect Real Time Data Feed and process it, then filter data in adequate area of block and thus send each block to processing mechanism[9]. First processing of data afterwards all other unwanted data will be removed. Thus, filtered information is the association of diverse key value pairs. Then blocks are sent to be processed by data processing unit and thus assign and transmit each distinct data block of processed data to processing steps. Thus, this takes live data then filters and splits them into segments by performing load-balancing algorithm[10].

C. Service Module: SMS Controller comprises of the SMS system, managing interaction between phones, stored data such as contact details, and an end-user's view along with manipulation of the system. Also the data controller handles connections to the database. SMS uses Ultra lightweight Thinlet for UI using Language Translator. It is just a great library, fast to gain information with great internationalization support. The Language Translator is easily translated into different languages. The users are the farmers using

mobile phones. Also it consists of GSM(Global System for Mobile communication),CDMS(Contractor Database Management System) etc. which is used for communication of detailed information to farmers. The detail information of farmers is stored in the database. The Cloud is used to store the information in form of database. Currently, SQLite and MySQL are fully supported. In the future, we are hoping to move database handling to a 3rd party library, which should massively improve the supported platforms.

So, all these individualistic components are necessary for the proper functionality of our system.

IV. ALGORITHM

Algorithm I. Attribute Based Access Control:

Attribute Based Access Control is needed for providing the logical access control to protect information, network devices, or other kind of information technology from unauthorized operations[5].

//Let B 1, B 2 are bilinear group of order p.

//where p is prime number, g is generator group and d is threshold value.

B 1 : e : B 1 x B 2 → B 2 is bilinear mapping.

Step 1 : Generating public key as well as master key

Select t 1 ,...,t n , y from finite field Zq and find

PK=(T 1 =gt 1 ,...,T n =gt n ,Y=e(g,g),y)

//where g is a bilinear group generator B 1 of order p .

//The master key is also generated

MK=(t1,...,tn,y).

Step 2 : Generating private keys

Randomly polynomial q of degree d-1 is selected such that,

q(0)=y.

//Private key is D={Di=g(q(i))(ti)} i AU.

//AU is attribute set.

Step 3 : Encryption

Owner data encrypt a message M B 2 using a attribute's set ACT,

s Zq : CT=(ACT, E=MYs=e(g,g)ys, {Ei=gtis} i AU).

//Set of random numbers.

Step 4 : Decryption

If |AU ∩ ACT| ≥l, then of i AU ∩ACT .

//Selects d attributes to compute values .

(Ei,Di)=e(g,g)q(i)s, Ys=e(g,g)q(0)s=e(g,g)ys.

//The message is M=E/Ys.

Thus for the authentication, private keys are produced for the principle of secret sharing.

Algorithm II. Filtration and Load Balancing Algorithm :

First collect huge amount of Real Time Data Feed then process it and thus filter data in adequate amount of size block. and thus start processing mechanism of each block[9].

Step 1: Filter related data.

//It firstly processes the data.

//All other unwanted data will be deleted.

Step 2: To distribute the data into key value pair.

//The filtered information is association of different key value pair.

Step 3: To transfer the unprocessed data without been processing to aggregation step.

//The data processing unit forwards the block to be processed.

Step 4: The processing in data processing unit.

//Thus, assign and transmit data block of processed data to processing steps and send the data via data processing unit.

This algorithm requires the input as current data and then filters or splits them into segments and apply load-balancing algorithm[10].

V. CONCLUSION

Due to lack of actual price of product, farmers faces economical loss through farming. Cloud computing enables easy to use, on demand, dynamic and steady use of different distributed computing resources. Basically our aim is to provide free services to the growing community of farmers. Proposed system consists of application which provides text message provider in regional languages to farmers to resolve the concerned queries. It provides easy-to-use interface for both farmers and concerned agency, also updates any information on market status of various products regularly. Due to cost effectiveness of proposed system, it can be successfully implemented in various rural areas.

REFERENCES

- [1] Praveen B. and Viswesh M., "Agriculture Updates via SMS A Cloud Computing Approach", International Journal of Innovation, Management and Technology, 2013, Vol. 4, No. 5.
- [2] Bhagyashri Deyannavar S., "A Survey on Cloud Based Framework for Providing Updates on Agricultural Products", International Journal of Emerging Technology in Computer Science Electronics (IETCSE), 2015, Vol. 14, No. 2.
- [3] S.Divya1, M.Mathivani,"Agricultural Updates Using Multitenant Cloud Infrastructure In Distributed Environment", International Journal of Advanced Research in Computer and Communication Engineering, March 2017, Vol. 6, Issue 3.
- [4] Vincent C. Hu,David Ferraiolo,Rick Kuhn,"Guide to Attribute Based Access Control (ABAC) Definition and Considerations",NIST Special Publication,2016,800-162.
- [5] Amit Sahai and Brent Waters, "Fuzzy Identity-Based Encryption," in EUROCRYPT, Aarhus, 2005, pp. 457-473.
- [6] Dr. S. Manimekalai1, Asst. Professor in School of Computer Science,"A Cognitive Approach to Mobile Application", International Journal of Emerging Trends Technology in Computer Science (IJETTCS), December 2013, Volume 2, Issue 6.
- [7] Kokula Krishna Hari K , "Agriculture Updates via SMS A Cloud Computing Approach", International Journal of Innovation, Management and Technology, 2013, Vol. 3, No. 6.
- [8] Lantzos, T., Koykoyris, G., Salampasis, "Farm Manager: an Android application for the management of small farms", Procedia Technology, 2013, 8, 587-592.
- [9] -Rakesh. S. Shirath, Vaibhav A.Desale, "Big Data Analytical Architecture for Real-Time Applications", IJSRNSC, August 2017, Volume-5, Issue-4.

- [10] Rakesh Patel, Mili. Patel Lecturer, Department of Information Technology ,”Application of Cloud Computing in Agricultural Development of Rural India”, International Journal of Computer Science and Information Technologies,2013,Vol. 4.
- [11] Prashant Paud,”A Review of: Cloud Centric IoT based Framework for Supply Chain Management in Precision Agriculture” ,International journal of advance research in computer science and management studies, November 2014, Volume 2,Issue 11.
- [12] Yanxin Zhu, Di Wu and Sujian Li,”Cloud Computing and Agricultural Development of China”, IJCSI International Journal of Computer Science Issues,January 2013, Vol. 10, Issue 1, No 1.
- [13] R. Buyya, J. Broberg, and A. Goscinski, ”Cloud computing principles and paradigms”, A John Wiley and Sons, INC, Publication, 2010, pp. 6.1-7.6, 15.1-16.7.
- [14] Sukhpal Singh. and Inderveer channa, ”Agro-Info:Cloud Based Autonomic System For Delivering Agriculture As a Service”, Int. J. Curr. Eng. Technol.,2013,vol.3, no. 5, pp. 20842087.
- [15] Mitsuyoshi, Eiji Kawashima, ”An Application of Cloud Computing To Agriculture Field”, J. Netw. Comput. Appl.,2010, vol. 4, no. 2, pp. 19972307.
- [16] Major singh and Harinder kaur, ”Cloud computing in agriculture Field”, J. Netw. Comput. Appl.,2000, vol. 4, no. 2, pp. 19972307.

A Survey on Visual Concept-Detection

Prajakta Bhosale, Nita Patil

Datta Meghe College of Engineering,
Mumbai, Maharashtra, India

S.D. Sawarkar

Datta Meghe College of Engineering,
Mumbai, Maharashtra, India

Abstract — Video search from huge database as per user perception is complex task. Video retrieval has number of applications and making it promising research topic. This paper offers an overview of general strategies in visual concept-based video retrieval, focusing on methods for video structure analysis, including shot boundary detection, key frame extraction and scene segmentation, extraction of features and concept detection.

Keywords- Feature extraction, concept detection, video retrieval.

I. INTRODUCTION

Video collections are becoming widely available, raising the need for effective access to video content. Video information indexing and retrieval [1] are required to describe, store, and organize multimedia information and to assist people in finding multimedia resources conveniently and quickly. Dynamic video is an important form of multimedia information. Videos have the following characteristics: 1) much richer content than individual images; 2) huge amount of raw data; and 3) very little prior structure. These characteristics make the indexing and retrieval of videos quite difficult. In the past, video databases have been relatively small, and indexing and retrieval have been based on keywords annotated manually. More recently, these databases have become much larger and concept-based indexing and retrieval is required, based on the automatic analysis of videos with the minimum of human participation. Traditional content-based video retrieval system cannot retrieve video shots semantically related with the user's intention. One way to facilitate access is concept-based video retrieval, where visual concepts are detected in video. A recent trend in concept-based video retrieval has been to search for generic methods that learn to detect concepts using examples [2, 3, 4]. Given a video collection with concept annotations, we can use a retrieval approach similar to text retrieval, where concepts are considered textual labels that can be indexed and retrieved using a standard text retrieval engine. A difference with text retrieval is that users' textual queries have to be translated into visual concepts, a step called concept selection. We explore different methods to address the problem of assessing concept selection in video retrieval system. This paper offers overview of strategies in concept selection in video and using it in video retrieval. It includes distribution- based concept selection, benchmarking, Ontology-Enriched Semantic Spaces, contextual estimates, multimodality fusion. The general overview on the overall process of concept detection framework which is outlined in Fig. 1. The framework includes the following: 1) structure analysis: to detect shot boundaries, extract key frames, and segment scenes; 2) feature extraction from segmented video units (shots or scenes): These features include static features in key frames, object features, motion features, etc.; 3) Concept detection: Modern approaches enable a semantic search by pooling a set of concept detectors (e.g., car and building) to extract semantics from low-level features, and thus forming a semantic space to facilitate high-level understanding of

user queries. Such search methodology is usually referred to as concept-based video search. The remainder of this paper is organized as follows: Section II briefly reviews the work related to video structure analysis. Section III addresses feature extraction. Section IV concept selection methods.

II VIDEO STRUCTURE ANALYSIS

Videos are arranged according to a descending hierarchy of video clips, scenes, shots, and frames. Video structure analysis divides videos into shot boundary detection, key frame extraction, and scene segmentation.

A. Shot Boundary Detection

A shot is a consecutive sequence of frames captured by a camera action that takes place between start and stop operations, which mark the shot boundaries. There are strong content correlations between frames in a shot. Therefore, shots are considered to be the fundamental units to organize the contents of video sequences and the primitives for higher level semantic annotation and retrieval tasks. Generally, shot boundaries are classified as cut in which the transition between successive shots is abrupt and gradual transitions which include dissolve, fade in, fade out, wipe, etc., stretching over a number of frames. Cut detection is easier than gradual transition detection. In shot boundary detection we first extract visual features from each frame, then measure similarities between frames using the extracted features, and, finally, detect shot boundaries between frames that are dissimilar.

The features used for shot boundary detection include color histogram or block color histogram, edge change ratio, motion vectors scale invariant feature transform , corner points , information saliency map, etc.

In second step i.e. measuring similarity between frames using the extracted features include pair-wise similarity measures the similarities between consecutive frames and window similarity measures that measure similarities between frames within a window. Using the measured similarities between frames, shot boundaries can be detected. Current shot boundary detection approaches can be classified into threshold-based and statistical learning-based.

1) Threshold-Based Approach: The threshold-based approach detects shot boundaries by comparing the measured pair-wise similarities between frames with a predefined threshold [13]. When a similarity is less than the threshold, a boundary is detected. The threshold can be global, adaptive, or global and adaptive combined. 1) The global threshold-based algorithms use the same threshold, which is generally set empirically, over the whole video, as in. 2) The adaptive threshold-based algorithms compute the threshold locally within a

sliding window. 3) Global and adaptive combined algorithms adjust local thresholds, taking into account the values of the global thresholds.

2) Statistical Learning-Based Approach: In statistical learning-based approach frames are classified as shot change or no shot change depending on the features that they contain. Supervised learning and unsupervised learning are both used. a) Supervised learning-based classifiers: The most commonly used supervised classifiers for shot boundary detection are the support vector machine (SVM) and Adaboost.

a) SVM : Use the SVM as a two-class classifier to separate cuts from non cuts and then use the SVM to classify the frames using these features into three categories: cut, gradual transition, and others.

b) Adaboost: The color and motion features are roughly classified first using a fuzzy classifier, and then each frame is classified as a cut, gradual, or no change frame using the Adaboost classifier. The main merit of the Adaboost boundary classifiers is that a large number of features can be handled.

3). Unsupervised learning-based algorithms

The unsupervised learning-based shot boundary detection algorithms are classified into frame similarity-based and frame-based. The frame similarity-based algorithms cluster the measurements of similarity between pairs of frames into two clusters: the cluster with lower values of the similarities corresponds to shot boundaries and the cluster with higher values of the similarities corresponds to non boundaries.

Shot boundary detection approaches can be classified into uncompressed domain-based and compressed domain-based. To avoid time-consuming video decompression, the features available in the compressed domain.

B. Key Frame Extraction

The features used for key frame extraction include colors (particularly the color histogram), edges, shapes, optical flow, MPEG-7 motion descriptors. Current approaches to extract key frames are classified into six categories: sequential comparison-based, global comparison-based, reference frame-based, clustering based, curve simplification-based, and object/event-based.

1) Sequential Comparison Between Frames: In these algorithms, frames subsequent to a previously extracted key frame are sequentially compared with the key frame until a frame which is very different from the key frame is obtained.

2) Global Comparison Between Frames: The algorithms based on global differences between frames in a shot distribute key frames by minimizing a predefined objective function that depends on the application.

a) Even temporal variance: These algorithms select key frames in a shot such that the shot segments, each of which is represented by a key frame, have equal temporal variance.

b) Maximum coverage: These algorithms extract key frames by maximizing their representation coverage, which is the number of frames that the key frames can represent.

c) Minimum correlation: These algorithms extract key frames to minimize the sum of correlations between key frames (especially successive key frames), making key frames as uncorrelated with each other as possible.

d) Minimum reconstruction error: These algorithms extract key frames to minimize the sum of the differences between each frame and its corresponding predicted frame reconstructed from the set of key frames using interpolation.

3) Reference Frame: These algorithms generate a reference frame and then extract key frames by comparing the frames in the shot with the reference frame.

4) Clustering: These algorithms cluster frames and then choose frames closest to the cluster centers as the key frames.

5) Curve Simplification: These algorithms represent each frame in a shot as a point in the feature space.

6) Objects/Events: These algorithms jointly consider key frame extraction and object/event detection in order to ensure that the extracted key frames contain information about objects or events.

C. Scene Segmentation

Scenes are identified or segmented out by grouping successive shots with similar content into a meaningful semantic unit. The grouping may be based on information from texts, images, or the audio track in the video.

According to shot representation, scene segmentation approaches can be classified into three categories: key frame based, audio and visual information integration-based, and background-based.

1) Key Frame-Based Approach: This approach represents each video shot by a set of key frames from which features are extracted. Temporally close shots with similar features are grouped into a scene.

2) Audio and Vision Integration-Based Approach: This approach selects a shot boundary where the visual and audio contents change simultaneously as a scene boundary.

3) Background-Based Approach: This approach segments scenes under the assumption that shots belonging to the same scene often have similar backgrounds. According to the processing method, current scene segmentation approaches can be divided into four categories: merging based, splitting-based, statistical model-based, and shot boundary classification-based.

a) Merging-based approach: This approach gradually merges similar shots to form a scene in a bottom-up style.

b) Splitting-based approach: This approach splits the whole video into separate coherent scenes using a top-down style.

c) Statistical model-based approach: This approach constructs statistical models of shots to segment scenes.

d) Shot boundary classification-based approach: In this approach, features of shot boundaries are extracted and then used to classify shot boundaries into scene boundaries and non scene boundaries.

III. FEATURE EXTRACTION

To extract features as per video structural analysis, we focus on the visual features suitable for video indexing and retrieval. These mainly include features of key frames, objects, and motions. Auditory features and text features are not covered.

A. Static Features of Key Frames

The key frames of a video reflect the characteristics of the video to some extent. Traditional image retrieval techniques can be applied to key frames to achieve video retrieval. The static key frame features classified as color-based, texture-based, and shape-based.

1) Color-Based Features:

Color-based features include color histograms, color moments, color correlograms, a mixture of Gaussian models, etc. Color features can be extracted from the entire image or from image blocks into which the entire image is partitioned.

2) Texture-Based Features: Texture features in common use include Tamura features, simultaneous autoregressive models, orientation features, wavelet transformation-based texture features, co-occurrence matrices, etc.

3) Shape-Based Features: Shape-based features that describe object shapes in the image can be extracted from object contours or regions. A common approach is to detect edges in images and then describe the distribution of the edges using a histogram.

B. Object Features

Object features include the dominant color, texture, size, etc., of the image regions corresponding to the objects.

C. Motion Features

Motion information represents the visual content with temporal variation. Video motion includes background motion caused by camera motion and foreground motion caused by moving objects.

Object-based motion features can be further classified into statistics-based, trajectory-based, and objects' spatial relationships-based.

1) Statistics-Based: Statistical features of the motions of points in frames in a video are extracted to model the distribution of global or local motions in the video.

2) Trajectory-Based: Trajectory-based features are extracted by modeling the motion trajectories of objects in videos.

3) Objects' Relationship-Based: These features describe spatial relationships between objects.

IV. CONCEPT SELECTION METHODS

We identify three broad approaches to concept selection: automatic, human association, and generation from extensively annotated video data.

1. Automatic concept selection

We use the term 'automatic concept selection' to describe the concept selection algorithms used in video retrieval systems to automatically translate a query to the system concept lexicon, usually returning a weighted list of concepts as a result.

2. Utilizing human associations

Knowledge about the world is implicit in human minds, and retrieval systems can exploit this knowledge by asking humans to select appropriate concepts for individual queries.

3. Utilizing labeled video collections

Recent efforts to label large video collections have allowed researchers to investigate generative concept selection, using data sets manually annotated with respect to both queries and concepts. Rather than these basic techniques researchers developed their own methods. Some of them are as follows:

4. Distribution based concept selection:

In distribution-based concept selection (DBCS), we select the most discriminative, concepts for video retrieval. The targeted concepts are those concepts whose distributions of detection score fluctuate widely between the relevant and irrelevant collections, but remain stable within both. DBCS takes into account the detection variation of data collection by assessing the variance of detection scores as one of selection criteria.

5. Benchmarking method:

In this method re-usable benchmarks are created. The use of these benchmarks allows evaluating concept selection independently from other components of video retrieval systems. Two benchmarks are used for assessing concept selection, one human-generated, the other back-generated from a video collection annotated with concepts.

6. Ontology-Enriched Semantic Spaces:

In these semantic spaces are constructed to measure concept similarity globally. Each concept is represented as a vector for similarity measurement purposes. The first space is named Ontology-enriched Semantic Space (OSS), second space is called Ontology-enriched Orthogonal Semantic Space.

7. Contextual Estimates

For each individual concept, the prior probability of the concept is incorporated with detection score of an individual SVM detector.

8. Multi-modal Fusion:

In this method two or more individual techniques are used for concept detection and combine to find result.

ACKNOWLEDGMENT

Authors would like to thanks DMCE for infrastructure and facilities provided.

REFERENCES

- [1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 2, no. 1, pp. 1–19, Feb. 2006.
- [2] M. Naphade, J. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3):86{
- [3] C. G. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. W. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *ACM Multimedia '06*, pages 421{430 .
- [4] M. Worring and G. Schreiber. Semantic image and video indexing in broad domains. *IEEE Trans. Multimedia*, 9(5), 2007
- [5] Hu, Weiming, et al. "A survey on visual content-based video indexing and retrieval." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.6 (2011):797-819.

- [6] Huurnink, B., Hofmann, K., & De Rijke, M. (2008, October). Assessing concept selection for video retrieval. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval* (pp. 459-466). ACM.
- [7] Cao, Juan, et al. "Distribution-based concept selection for concept-based video retrieval." Proceedings of the 17th ACM international conference on Multimedia ACM, 2009.
- [8] Wei, Xiao-Yong, Chong-Wah Ngo, and Yu-Gang Jiang. "Selection of concept detectors for video search by ontology-enriched semantic spaces." *IEEE Transactions on Multimedia*10.6 (2008): 1085-1096.
- [9] Aytar, Yusuf, O. Bilal Orhan, and Mubarak Shah. "Improving semantic concept detection and retrieval using contextual estimates." *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007.
- [10] Wei, Xiao-Yong, Yu-Gang Jiang, and Chong-Wah Ngo. "Concept-driven multi-modality fusion for video search." *IEEE Transactions on Circuits and Systems for Video Technology*21.1 (2011): 62-73.

Reliable Human Identification using Iris as a Biometric

Jayshree H. Sali

Instrumentation Engineering

TCOE Boisar

Mumbai, India

jayshree.sali454@gmail.com

Akram Pathan

Electrical Engineering

TCOE Boisar

Mumbai, India

akramkhan452@gmail.com

Rakesh S. Kushwah

Electrical Engineering

TCOE Boisar

Mumbai, India

rakeshkushwah027@gmail.com

Abstract—Among various methods of personal identification iris recognition is regarded as the most reliable and accurate system available. The purpose of this paper is to represent iris recognition algorithm so as to prove its reliability as a biometric on the basis of its performance. The images selected for study are from CAS IA iris database. The classification is done on the basis of the most repeatedly occurring class using KNN classifier. Euclidian distance metric was employed to check the similarity between two iris images and the two images said to be matched if its value is less than or equal to threshold. The system was tested and shown an overall accuracy of 97.5 % with false rejection rate of 3% and false acceptance rate of 2%.

Keywords — *Iris Recognition, Hough Transform (HT), Euclidean distance(ED).*

I. INTRODUCTION

With rapid development of our society more advanced identification systems are in demand. Iris recognition system is gaining highest proportion among all the biometric identification systems. The most important characteristic of iris is its uniqueness in nature which leads to develop more and more accurate identification system.

The first working model of iris recognition was implemented by Professor John Daugman of Cambridge University in the 1990 [1] [2]. The Daugmans system is patented and many commercial developers are using it. A large number of studies have tested the Daugman system as well as commercially it has been used by many users and proved that it has zero failure rate when tested with millions of images. Few years later many other developers have designed the iris recognition system and acquired a great success. The few notable of them are Wildes et al. [3], Boles and Boashash [4]. The Wildes et al. system also has a flawless performance [3], whereas the Lim et al. system achieves an accuracy rate of 98.4 %. Though many developers have developed the algorithm but they could not achieve the accuracy and speed when compared with the Iris recognition system developed by Daugman.

The iris is a dark annular portion of the eye which is surrounded by muscle, tissue, blood vessels, cells and outer boundary within the eye as shown in figure.1 resulting in random and unique pattern for each iris. Due to this unique combination it is more suitable for biometric identification. In addition it also remains stable over the period of time. This unique pattern contains some redundant features and if utilized directly in the classifier it affects the performance of iris

recognition system resulting in low efficiency. The useful features can be extracted from digital eye image with help of

developed mathematical algorithms and the results can be stored in the form of code into a biometric template. The biometric template of digital eye image is usually created by the system. A database for different digital eye images is generated in the form of code and stored as biometric template for future comparisons. With the help of developed matching algorithms this biometric template is compared with previously stored template in the database in order to get the verification of the individual. Iris recognition technique consists of three basic stages.

1. Preprocessing
2. Feature Extraction
3. Image Matching

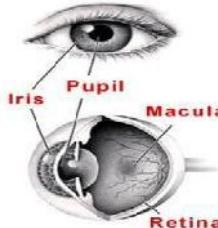


Figure.1 Eye Image

II. PREPROCESSING

Segmentation:

Iris is an inner portion between pupil and sclera. To separate iris portion from the digital eye image segmentation is used. A circular Hough Transform based automatic segmentation approach is used for inner and outer boundary location. It removes the redundant regions like eyelashes and eyelids present in lower and upper region of image. So sometimes specular reflections may occur inside the iris region. The segmentation method must be capable to partition these noises and locate inner and outer boundary of the image. Generally pupil boundary which is an inner boundary of an iris portion is easily located using binarization. But it requires a good quality image.

i) Pupil Detection

In Hough method first gradient image is created from intensity image. The iris image is convolved with the sobel filters to convert into intensity image. sobel operator calculate the gradient of the image intensity at each point. Using canny edge detection technique edges of gradient image is found. To obtain the center the given formulas are used.

$$xc = x - r * \cos(\theta)$$

$$yc = y - r * \sin(\theta)$$

The accumulator stores the values of x_c and y_c for a particular value of r . The highest value stored in the accumulator counter is considered as the centre of the pupil as well as the radius.

ii) Outer boundary detection

Using histogram equalization Image is enhanced to have sharp variation at image boundaries and outer boundary can be easily detected. From the pupil center circles of different radius are drawn and their intensities lying over the periphery of the circle are boundary.

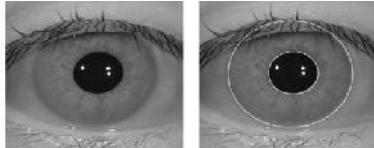


Figure 2: Inner and outer boundary detection.

Normalizat ion:

Normalizat ion is the process which converts the segmented iris image into polar form which helps to generate binarized code. This process is known as Unwrapping. Daugman's

Rubber sheet model is used for this purpose. Iris boundary and pupil boundary generally does not have concentric circles. This condition gives different reference points for conversion into polar form. Remapping method is used for conversion of Cartesian scale to polar scale considering pupil centre as reference point. This algorithm remaps each point of the iris image to polar coordinates (r, θ) .

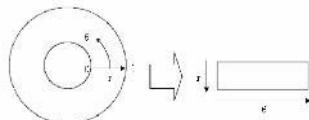


Figure 3 Daug man's rubber sheet model

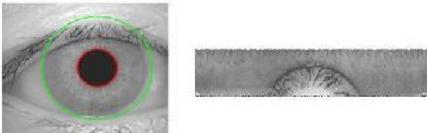


Figure 4. Normalized Iris into polar form

III. FEATURE EXTRACTION

To extract the texture information from an iris image wavelet transform algorithm is used. The important characteristic of wavelet function is that it extracts the local frequency information of an image, eliminates the redundant information, reduces the noise from the image and reconstructs the iris image in the form of binary vector. The input image is decomposed using wavelet transform. This is carried out in four levels with high pass and low pass filters. When wavelet

function is applied the image is divided in to four sub regions. These components are approximation image component (LL), horizontal detail co mponent (HL), vert ical detail co mponent (LH), diagonal detail co mponent (HH). The forth level component contains higher energy level therefore it is selected as an input image for the next level. The output of the DWT is then encoded into biometric template with phase quantization method and the iris code is generated as shown in figure 5 which is saved in a database for comparison between iris temp lates.



Figure 5.: Iris code

IV. IMAGE MATCHING

a) Classification Phase:

KNN algorithm is preferred particularly for dimensionally high input data. KNN algorith m has lower error rate compared to other classifiers. It is a simplest form for classification and calculation among the others. In dimensional space the training patterns are plotted according to their observed feature values and labeling is done by their known class. [8] Within the same space an unlabelled test pattern is also plotted. The classification is done on the basis of the most repeatedly occurring class among its training patterns and its nearest neighbor patterns. Euclidian d istance metric is the most popular for KNN classification which measures similarity between feature vectors.

b) Matching algorith m

In matching algorith m the degree of similarity between two feature vectors is determined. For distances comparison and computation of feature vectors Euclidean distance algorithm is used in our system which decreases the computational cost.

$$ED = \frac{1}{N} \sqrt{\sum_{i=1}^N (iris_angle A(i) - iris_angle B(i))^2}$$

differentiates the same iris images from different iris database. The user is genuine if the matching score is less than or equal to reference threshold.

V. EXPERIMENTAL RESULTS

The database used for study was taken from CASIA database and the performance verification of the system was carried out with two set of images in two intervals of 100 and 250 images respectively. It is observed that Hough transform performs better as compared to other localizat ion techniques and system has efficiently detected inner and outer boundaries of all images. The result shows that the iris recognition system performed perfect recognition with 100 eye images with zero percent false registration. The results are plotted with false acceptance rate and false rejection rate and found overall accuracy of 97.5 % with FRR of 3% and FAR 2% as shown in figure 6 and figure 7.

There may not be 100% matching Even if the same iris images

are compared, with the samples of iris images present in the database. Similarly, the matching score will have exactly the same value even if two very different iris images are compared. Solution to this problem is the decision of correct value of reference threshold value which basically

We compared the system with other systems on the basis of accuracy. The results show that the proposed algorithm has improved the recognition rate when compared with other systems proposed by Ma[6], Boles[7], Daugman[2]. They obtained 89.37%, 92.64 and 99.99 accuracy respectively.

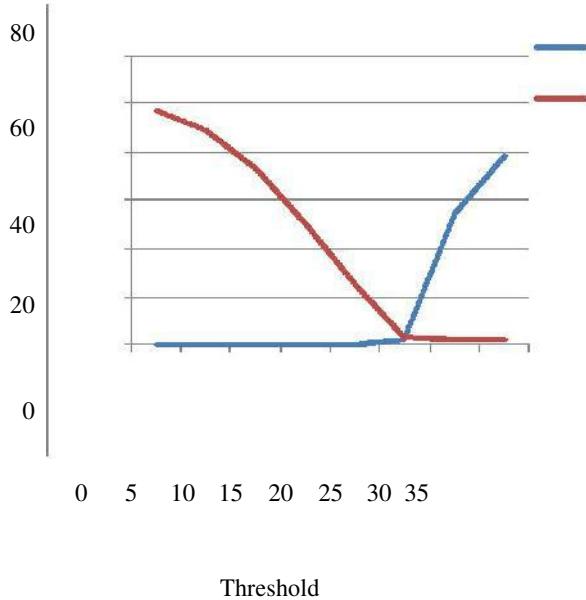


Figure 6: FA R/FRR – Threshold plot

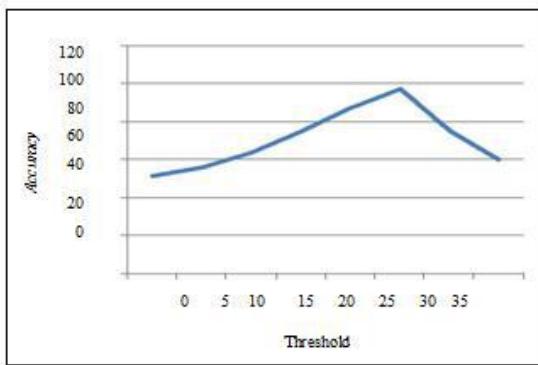


Figure 7: Accuracy-Threshold plot

VI. CONCLUSION

The iris recognition system that was developed proved to be a highly accurate and efficient system that can be used as a biometric. It is proved that iris recognition is one of the most reliable methods available today in the biometrics field. The accuracy achieved by the system was very good.

J.Daugman	,"Ho	REFERENCES	work",IEEE
n			
			Trans.Comput.,syst.,VideoTechnol., vol.14,no.1,pp.21-30,2004
			J.Daugman .,"How is iris recognition work",IEEE Trans.Comput.,VideoTechnol., vol.14,no.1,pp.21- 30,2004.
			J.Daugman ."New method iniris iris recognition",IEEE Trans.Syst., Man,Cybern.,B,vol.37,no.5,pp. p1616771,2007.
			R. Wildes. Iris recognition: "An emerging biometric technology"
			R. Wildes. Iris recognition: "An emerging biometric technology"
			Proceedings of the IEEE, Vol. 85, No. 5, 1997.
			B. B. Bhowmik, "A human identification technique using images of the iris and wavelet transform". IEEE Transactions on Signal Processing, Vol. 46, No. 4, 1998.
			W. Boles, B. Boashash. "A human identification technique using images", C. H. Daouk, L. A. El-Esber, F. D. Kamoun and M. A. Al Alaoui
			"Iris Recognition" IEEE ISSPIT 2002 page 559
			of the iris and wavelet transform", IEEE Transactions on Signal Processing, Vol. 46, No. 4, 1998.
			L. Ma, and T. Tisse, Personal Recognition Based on Iris Texture Analysis", IEEE Trans. on PAMI, Vol. 25, No. 12, 2003, pp. 519-533
			W. Boles, B. Boashash, "A human identification technique using images of the iris and wavelet transform",IEEE Transaction on Signal Processing, 1998, 46(4): 1185-118.
			C. H. Daouk, L. A. El-Esber, F. D. Kamoun and M. A. Al Alaoui "Iris Recognition" IEEE ISSPIT 2002 page 559
			Arun Saman, Chandanayoti Kumar "Iris Verification Using Legendre Moments and KNN Classifier" ISSN August 2013.
			L. Ma, and T. Tisse, "Person Recognition Based on Iris Recognition", IEEE Trans. on PAMI, Vol. 25, No. 12, 2003, pp. 519 -533
			AzadeBastani, "Iris Recognition based on Wavelet Transform and Probabilistic Neural Network", IJCTE, volume, 14, number 3, Aug 2014.
			Moments and KNN Classifiers", ISSN August 2013.
			International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 2 (2015), pp. 1-10

A New Fighter Alarm Based On Interaction

Hamidreza Hajiaghai

Department of Computer

Damavand Branch, Islamic Azad University

Mehdi Gheisari

Department of Computer

Damavand Branch, Islamic Azad University

Abstract- Alarm or robbery alarm system is a devise that declare passing from privacy with creating noise and informing different ways such as sms, phone. Alarms on the markets have different forms and varieties, however, but generally they have common structure and their main purpose is similar to each other. There is any kind of alarm on the markets and each of them uses in particular places such as: automobile alarms, bank alarms, house alarms. This article considers and presents one different alarm that it's aim is making more possible security in public and private especially military sites.

Key words: Telephone, automatic fighter alarm, call control, messaging systems

I. INTRODUCTION

The first alarm system was invented by a man in Boston, August Rasel pap, in 21 June 1853.nowadays pap's invention may seem simply but it is very affection against rivals. The function of this system is that door and window linked together independently by parallel tropics, if the door or window was open,

system's electric tropics was close and sudden action causes activating Electromagnetic alarm. Nowadays, these systems are used at Residential, commercial, industrial, and military to protecting against robbery or property damage and personal protection and they also can combine with Television surveillance systems to recording rival action automatically. In addition, this

alarm has very capabilities such as: automated call, automated system to send and receive SMS, Telephony control system and most important of all, fighter alarm – which this system can fight physically and prevent to entering places. In this article, we consider new alarm based on interaction. The most important protective tools are sensors to controlling and recognizing robbers in alarm systems. And these systems are useless without sensors. In this article, we consider one of most important of them.

a) High tech new alarm systems - 20 century

We witnessed significant developments of alarm techno in 20 century.

When designing of emergency calls box was affordable after World War II, it was used to controlling medical, police and fire station services. At a result, people security improved

across the country. Engineers combined the first motion detectors on own alarms in 1970.with the rise of democracy, alarm systems was known as a Standard safety features in a building, in the 80's and 90's.finally, the first wireless alarm system entered the markets and alarm techno changed entirely in practical level well and The inevitable mess of cables was away. Nowadays, even we can protect complex places entirely by Motion detectors and video surveillance technology. It can be said still technology innovation have to wonderpeople.

b) PIR

Most of alarm systems based on PIR are designed to when a man or great warm-blood animal moves in detector functional zone of PIR, an alarm or spotlight kindled or a door opened or other types of electro-mechanical systems can be activated. Infrared detectors are following electric made by one or two crystals, one optical filter and a FET transistor. Some of crystals and specific ceramics produce load exporting heat that it can be called as

“Electric Follower”. Modern Electric infrared detectors followers are made by connection two small electric ceramics follower and reveres polarity such as PIS201S and E600STO and output of this system is buffered by a JFET.

The function of this detector is that if a body moves among crystal vision of electric follower a part of infrared energy is radiated and is shone on surface of crystal and this function causes minor changes in heat and consequently it changes the output voltage. Changing the voltage give an order to the alarm that it cause activating safety systems such as alarms.

II. DESIGNING, IMPLEMENTATION AND EVALUATING OF PROPOSED ALARM

Fighter alarm has different using of safety of places and avoiding any possible theft. Most important feature, of course, is its certain difference with advanced alarm in the markets that causes this system was registered as an

invention and presented also as a research paper. Besides this alarm, there is a system that war weapons placed on it and do the operation such as Colt. As advanced alarms, this alarm has the ability of linking to different sensors such as smoke, motor, gas leak that they can be wire or wireless. This device is capable of sending occurring events through different sms to different people. It also calls with relevantagencies and presents its address orally. Also it can capture the environment and send to them online and we can hear environment voices and talk to robber.

Other facilities of fighter alarm are as follow:

A) phone controlling system

This system enables the user to allow changing the created situation in any location and device settings. User can call system and with entering security codes create variable changes such activating or inactivating some sensors, SMS system, and fighter system. The system sends a SMS when user set the system in order to ensuring of setting system. This system doesn't need anyapplication.

B) SMS system

SMS systems of alarms do, that is, if one or more sensors react as the result of physical elements, system will send constant SMS to reserved numbers in system when sensors commend alarm. But SMS system of this alarm has differences such as: the order of sensors is different from each other and there is different SMS foranyusedsensor.Forexample, when the door number one reacts as the result of physical movement, then it orders to alarm system and so the alarm system sends SMS to specified people that its text is well to events, for example the robber isentered from the door number two

or the window.

C) Fighter Alarm System

Fighter system has a Colt. System will react when robber does any physical movement in the front of its camera. This system is influenced by physical factors on its camera cleverly and automatically the camera order to trigger gun to guide toward robber. Then, recorded sound activates and orders to robber leaving the site before 3 seconds, for example. If the robber submitted and left the situation, the system will inactivate. But if he does not leave the situation after 3 seconds, the gun will shoot burglar. This system, of course, uses in military sites more than others and also it uses in public places such as banks, with permission from the appropriate agency.we can separate this fighter alarm and use it in military places separately. One certain cable of fighter system links to computer that it canshow confirmed SMS to user after he sets the setting and this can assure user from his changing. The most important and practical system is used in proposal alarm is fighter system. This system can prevent entering robber through area and fight him and get him out of place and it can kill him. This system is very important to protect the safety of position.

REFERENCES

- [1]. Alarm- krav spesifikasjon for Amoco Norway Oil Company, Val hall CCR Upgrade Project,IFE/HR/F-99/1118.
- [2]. Requirement Specification for the HAMBO Alarm System, IFE/HR/F-2000/1141
- [3]. Alarm Systems: A Guide To Design, Management and Procurement, The Engineering Equipment and Materials Users Association (EEMUA) publication No. 191
- [4]. NORSOK Standard I-002
- [5]. www.ptil.no
- [6]. www.homeadvisor.com
- [7]. www.teachmealarms.com

SQLIProtectParser for Preventing Sql Injection Attack

Neha Patwari

Department Of Information Technology

Thakur College of Engineering & Technology, Mumbai

Nishtha Mathur

Department Of Information Technology

Thakur College of Engineering & Technology, Mumbai

Ratna Nayak

Department Of Information Technology

Thakur College of Engineering & Technology, Mumbai

Shruti Mathur

Department Of Information Technology

Thakur College of Engineering & Technology, Mumbai

Purvi Sankhe

Department Of Information Technology

Thakur College of Engineering & Technology, Mumbai

Abstract—With the changing of globalization, the rise and generality of web application have gained a focal and significant part in the areas of innovation and progressions. Web applications encase the working between a client and the administrations gave by the server, which contains a database as its backend. The client can get to the required data through sending a demand as content to the web server, which is deciphered by the server side content to build a SQL. The query is sent to the database which reacts keeping in mind the end goal to produce a HTML page that is sent back to the client. Since the working of web application is a dynamic and entangled issue, certain dangers to the database security have been enlisted. One such disturbing risk is the predominance of SQL Injection Attack. Thus a dynamic calculation is given in this paper for anticipating SQL Injection Attacks which depends on setting free sentence structures and compiler parsing procedures. The paper endeavors to show the documentation of a SQLIProtectParser for the avoidance of SQL Injection Attacks. This Parser decides the formation of queries and thinks about whether the queries are practically identical or not. This parser has been utilized on an example web application and the outcomes have turned out to be certain majors to forestall SQL Injection Attack

Keywords-SQL;SQLInjectionAttack;SQLIProtectParser; Prevention

I. INTRODUCTION

There has been a quick progression in data innovation because of the across the board utilization of the web since the previous couple of years. The basic man today utilizes the web with various purposes, for example, to be utilized as a part of the field of instruction, for cash exchanges and different incalculable exercises. The safe sites store the very delicate data alongside non-basic information in their database frameworks such that the Owner of the data can get to it rapidly while attackers of the unapproved clients are hindered in their endeavors to approach the data.

Consequently, we need to comprehend the engineering of web application; a web application acknowledges demands from clients with a specific end goal to accumulate data from a database. It is expected by database that the info is right and in this manner utilizes it to get to the database by making a SQL. These web applications turn out to be basically inclined to SQL injection attackers since these don't

check the legitimacy of the client queries before submitting them to assemble the information. For instance, attackers imagining as authentic client use malicious made information content which contains SQL guidelines keeping in mind the end goal to create SQL queries on the web application back-end. Whenever web application forms the queries, the acknowledged pernicious query may rupture security net of the underscored database. As an outcome of the query there happens a despicable working of the database parser which brings about the arrival of the sensitive data [1]. With a specific end goal to approach the touchy data from the database a general break-in technique is to first make a query which will degenerate the working of the database parser, and forward the utilization of this query to the focused on database. This kind of approach keeping in mind the end goal to approach the private data is known as SQL injection. Presently SQL injection has turned into a common event because of the simple access of the database by means of the web. It is similarly important to have a profound comprehension of the kinds of correspondence which happens amid a specific session in the middle of a web application and a client with a specific end goal to have a superior comprehension of SQL injection.

II. RELATED WORK

A. Syntactic Structure Enforcement

Different procedures manage input approval by authorizing that all information will take the syntactic position of literals. Tie factors and parameters in put away techniques can be utilized as placeholders for literals inside questions, with the goal that whatever they hold will be dealt with as literals and not as discretionary code. SQLrand, an as of late proposed guideline set randomization for SQL in web application.

It depends on an intermediary to interpret directions powerfully, so SQL catchphrases entered as info won't come to the SQL server as watchwords. The principle disservices of such a framework are its unpredictable setup and security of the randomization key. Halfond and Orso address SQL injection attackers through building model of legitimate inquiries and after that guaranteeing that created questions comply with this model by means of runtime observing , following a comparable way to deal with Wagner and Dean's work on Intrusion Detection .By means of Static Analysis . The exactness of this system is liable to both the accuracy of the statically developed model and the tokenizing method utilized. Since how their model is generated, user inputs are statically syntactic positions. These methods for authorizing syntactic structure don't reach out to applications that acknowledge or

recover inquiries or inquiry sections, for example, those that recover put away questions from constant stockpiling.

B. Static and Runtime Checking

Some genuine web applications have vulnerabilities, despite the fact that measures, for example, those specified above are utilized. Vulnerabilities exist as a result of insufficiency of the strategy, uncalled for usage, incomplete utilization, or some mix of these. Hence, blackbox testing devices have been worked for web database applications. One from the examination group is called WAVES (Web Application Vulnerability and Error Scanner). A few business items likewise exist. While testing can be valuable practically speaking for discovering vulnerabilities, it can't be utilized to make security ensures. Therefore, a few methods in view of static investigation or runtime checking have been proposed, a large portion of which depend on the thought of taintedness,. Specifically, there are two late procedures utilizing static examination to track the column of untrusted contribution through a program: one in view of a sort and one in view of a focuses to investigation (utilizing an exact focuses to investigation for Java also, arrangements determined in PQL). The two frameworks trust client, so they don't give solid security ensure. There is likewise late work on runtime pollute following.

III. OVERVIEW OF SQL INJECTION

A web application is one by which a client can get to the administrations gave by the web server while chipping away at a customer machine, which contains a database. Example is online email id where the client enters a login name and password to get to the email account. As the client presses the submit button a URL is made and is sent to the web server. The server side of the content interprets the client contribution because of which a dynamic SQL query is made. It is submitted to the database and HTML pages are produced because of the query which is sent back to the client. A specific area of the database query code is put together by the malicious attackers to the server, while reacting with the relating result some touchy data is uncovered by the server. This is ordered as SQL injection attack. A SQL injection attack contains injection of a SQL query to the application through the input from the client. If fruitful SQL injection can interpret and change the information in the database (Insert/Update/Delete), it means that an SQL injection attack happens. When the intended effect of an SQL query is changed by an attacker on inserting new SQL keywords or operators into the query, following are the qualities of SQL injection attacks:

- i) Threat Modeling ii) Attack Intent iii) Assets

The concept motivating a SQL injection is simple above all attacks like these can be executed and mastered with easiness. To exploit the SQL injection weakness the basic prerequisite for the attacker is to make out the working of the web application. Malicious SQL commands are often inserted fastidiously into the content of the standards permitting the assailant to trick the net application so that a malicious query can be forwarded to the database.

IV. TYPES OF SQL INJECTION ATTACKS

A. First Order Attacks

All type of attacks, if performed directly in text field and provides necessary information or data, then this type of attack is called First Order Attack or Direct injection.

In the case of direct injection the SQL query will use each variable submitted as such without any changes. For example attempt to take parameter's valid value and appending a space along with the word

"OR" with this. In case if an error is generated by this, a direct injection is possible.

First order is performed by SELECT query which is used in application for retrieving information.

Tautologies Queries

The aim of this type of attack is to insert code that may be in one or more conditional statements due to which the statements are always evaluated as to be true. The results of this type of attack take place due to the way in which the application uses the result of the query. The most common purpose is to skip validation route and extract data. An attacker exploits an injectable area under this type of injection which is utilized in a query's WHERE conditional. The transformation of the conditional into a tautology results in returning all the rows in the database table being targeted by the query. For the attack to be fruitful the code must either display all of the returned records or must perform some action so that at least one record is returned.

Illegal/Logically Incorrect Queries

This category of attack allows gathering the related information as per the type and structure of the back-end database of a Web application.

The main aim of this attack is to collect information for further attacks and is treated as a beginning step. These attacks pinpoint a weakness due to which the application servers returns the default error page which often contains over description. The vulnerable or injectable parameters can be revealed to the attacker due to the simple fact that error messages are being generated.

The additional error information which was fundamentally aimed at assisting the programmer to repair or correct their application further allows the attacker to access information related to the schema of the back-end database. During working on this type of attack, he tries to insert statements which can result in syntax error, type conversion or could create logical error into the database. The injectable parameters can be detected by using the syntax errors. The deduction of the data types of certain columns or the separation of the data can be done by using the type errors. The names of the tables and columns causing the errors can often be revealed by logical errors.

Union Queries

In such attacks the parameters which are not strong are exploited by the attacker with a view to transform the data set returned for a particular query.

This technique allows the attacker to form the application, giving back data from a table not from the one which intended by the developer but from another unintended table.

The attacker performs it by making a statement in the way: ' UNION SELECT < remaining of injected query >.

Since the second/injected query is totally controlled by the attackers, this query can be used by them in order to retrieve information from a particular table. This attack results in the form of a dataset from the database which is the collective result of the original query and the injected query.

B. Second Order Attacks

In this type of attack when the malicious code is injected into the web based application instead of being immediately executed it is stored by the web application i.e. it is first stored in the database to be retrieved, rendered or executed by the victim. This category of attack happens because of the notion that when the data is contained in the database, it is often supposed to be clean and need not be checked again. While

due to the regular use of the data in the queries, it is still able harm the web application. This type of attack happens in case where the filtration process is ignored during the process of data insertion in search page. We should apply filtration for special characters before storing data in databases, which no special characters are allowed for inserting in databases. It is inherently performed by INSERT basics which are used in application. INSERT keyword is used to add information in the database. In case of web application this keyword is used for user registrations, bulletin boards inclusion, adding items to shopping carts, etc. While trying to INSERT injection it could result in the flooding of the rows in the database having single quotes and SQL keywords. As per the at tentativeness of the administrator it can be evaluated that what is to be done with the information. For example the user is on a site on which user registration of some kind is allowed.

A format is provided in which the user has to enter name, address, phone number, etc. As the information is submitted in the format a page is generated where this information is displayed along with an option to edit the information. This is what is required by the user. Thus after the process of insertion the required data can be modified and updated. Thus in case some malicious data is inserted in the database by the attacker, the data can be updated as per the desire of the attacker.

Piggybacked Query attack is example of Second Order Attack[5].

Piggybacked Queries

In this attack, attacker tries to inject queries in the original query. These kinds of attacks in

contrast to the other type of attacks instead of modifying the original intended query tries to insert new and distinct query that “piggy-back” on the original query. This results in multiple SQL queries to the database. The initiating query is executed as normal while the remaining queries are injected queries, and being executed along with the initiating query.

The attack of this category is highly critical. In case an attacker succeeds in this attack he can virtually insert any sort of SQL command in the additional queries and is able to execute them along with the initiating query. This kind of attack vulnerability is often due to the possession of a database configuration via which multiple statements can be inserted in a single string.

V. PROPOSED METHODOLOGY

The methodology which has been used to protect the SQL injection attacks is the merging of SQLIProtectParser with the application therefore protecting against any attacks. Firstly SQLIProtect Parser has been built which is used to find out the structure of the query. Then limitations of the method are identified. Finally, the solution to overcome the problems has been proposed making the system fully efficient.

A. Approach

The developer built a data structure for the parsed representation of the statement, which is called a parser. For parsing, we need the grammar language of statement. In this method, by parsing two statements and comparing their parser functionality, it leads to conclusion that the two queries are equal. When sql is injected successfully in database query, the parser of the intended SQL query and the resulting SQL query is generated after mismatch of attacker's input.

The SQL Query is:

```
SELECT * FROM login WHERE login name=' ' AND password=' ';
```

Web applications have SQL injection vulnerabilities because inputs are not sanitized which they use to construct structured output.

If an attacker passes name = ' OR 1=1 as the login name, all login name in the database will be returned and displayed, reason being transformation of entire WHERE clause into a tautology of code injected in the conditional statement('

OR 1=1 --). The conditional used by database to evaluate each row and decide the rows to return to the application. Since conditional is a tautology, query evaluates each row in the table as true and returns all of them to application. The problem is reckoned by taking into consideration its cause:

The detailed information of the program is that the substrings are taken from user input and the substrings are restrained syntactically. The concept is to restrict queries in which the input substring modifies the syntactic structure of the remaining query. Such queries are called SQL injection attacks in the perspective of database back-ends.

The user's intake is visualized by using meta-character displayed as '(|' and '|)'. It allegorizes the commencement and ending of each input string. This meta character follow the string through assignments, concatenations, etc., thus as a query is ready to be transferred to database, it contains matching pair of markers identifying the substrings from input. We should reject to introduce input substrings from modification of the syntactic structure of the remaining of the query. For this grammar for queries as per the standard grammar for SQL queries is build up. In the grammar, the only productions in which '(|' and '|)' occur have the following form:

Non terminal ::= '(|' symbol '|)'

Where symbol is either a terminal or non-terminal

For query to be in the language of this grammar, the substrings surrounded by '(|' and '|)' must be syntactic. A parser generator is used to build a parser for grammar and each query is attempted to be parsed. In case the query is parsed successfully, it meets the syntactic constraints and is legitimate. Conversely, it fails the syntactic constraints and may be a SQL injection attack.

After SQLIProtectParser is built using the grammar of the output language and plan of action is specified that permitted syntactic forms, it remains on the web server and intercepts generated queries. Each input needs to be propagated in form of some query, notwithstanding the input's source, gets amplified with the meta-characters '(|' and '|)'. Then query is generated by the application, which SQLIProtectParser attempts to parse. If a query parses successfully, SQLI Protect Parser sends it to the database without the meta-character. Otherwise, the query is block out.

Example:

```
SELECT * FROM login WHERE name='shyam ' OR 1=1 -- ;
SELECT * FROM login WHERE name=(('shyam' OR 1=1 -- ));
SELECT * FROM login WHERE name=(('shyam' OR 1=1 -- ));
```

Limitation

This solution can be conquer in either of two ways:

1. If the attacker is somehow able to detect the delimiter used, it would require only a little changes of the query to break this protection.

Eg: Query with routine delimiter:

```
SELECT * FROM Table WHERE Condition = '(| $VARIABLE |)'
```

Attack string (\$VARIABLE) becomes : xyz | OR |

2. The attacker may simply use a brute force attack to guess the correct delimiter combination, try all possible combination.

Eg: Query with routine delimiter:

```
SELECT * FROM Table WHERE Condition = '(| $VARIABLE |)'
```

Attack string (\$VARIABLE) becomes : /) OR #) OR !) OR |) OR (/ OR (# OR (! OR (

VI. RESULT

Original solution where we utilize static delimiter moved up to go around potential security leaks. Hence forth execution by dynamically changing the delimiter mix for each factor field and not utilizing the same delimiter mix for two sequential variable fields or in same field in application. Because of this modification to the first calculation, the attackers should accurately figure the correct arrangement of delimiters used to sidestep the parser's security framework. Since the delimiter mix will be cycled arbitrarily this won't be effortlessly conceivable. By this proposed technique static delimiter has been made dynamic. This arrangement influences the parser more to secure than before (fig. 1).

Sample query string becomes:

```
SELECT * FROM Table WHERE id = (| shyam'|)
and pwd = (# 'efg' #)
```

The essential structure of work where the client input is translated by the web application. In the web application it has been utilized the idea of dynamic delimiter with the goal that the attackers can't figure the arrangement of the delimiter; here even the client has no clue about the grouping of the delimiter. Subsequently in the given application the restriction of the static delimiter has been eliminated.

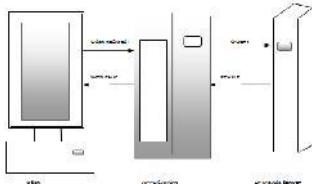


Fig 1: Methodology

If the client puts any contribution by utilizing the delimiter for instance x|) OR '| 1 = 1 , then it will be checked at the application itself and the mistake is accounted for here itself. Presently from application query is sent to the parser (fig. 2).

The Parser decides the structure of SQL query and input variable. Parser looks at that both queries (implies query at the application and the query at the Parser) are practically identical or not. Incase both the queries are practically identical then it reaches the database then reaction is taken from the database, which is produced as a HTML Page and is send to the client (fig.3) [1].



Fig 2:Some Capture of Methodology

VII. CONCLUSION

This paper presents the first overview of SQL injection attacks in web application. According to the presented paper an effective technique has been developed for preventing SQL injection attacks. The implementation on web application and parser on java CC [10] proved effective under testing. Here have been diligent efforts in applying parser on web application and produces output.

The result of evaluation and test proves that the proposed method is an effective technique to prevent SQL Injection Attacks.

In this work it has been managed to prevent SQL injection attacks through:

- Tautologies Queries
- Union Queries
- Illegal/Logically Incorrect Queries
- Piggybacked Queries

•SQLIPProtectParser has been built for SQL constraints

Following are the two goals for future works: The parser is to be more comprehensive for maximum number of SQL commands. The technique can be used to avoid cross-site scripting.

REFERENCES

- [1] Zhendong Su. The essence of command injection attacks in web applications. pages 372–382. ACM Press, 2006.
- [2] J.Viegas William G.J.Holfond,"A classification of sql injection attacks and countermeasures," In IEEE, 2009.
- [3] Chris Anley. Advanced sql injection in sql server applications. In An NGSSoftware Insight Security Research (NISR) Publication, 2002.
- [4] SQL Injection, Are Your Web Applications Vulnerable? <http://www.securitydocs.com/library/2656/>.
- [5] J.Viegas William G.J.Holfond,"A classification of sql injection attacks and countermeasures," In IEEE, 2009.
- [6] SQL Injection, Are Your Web Applications Vulnerable? <http://www.securitydocs.com/library/2656/>.
- [7] SQL Injection Attack and Defense. <http://www.securitydocs.com/library/3587/>.
- [8] D.K. Bhattacharya Debasish Das, Utpal Sharma," An approach to detection of sql injection attack based on dynamic query matching," In International Journal of Computer Application(0975-8887) volume 1-No.25,2010., 2010.
- [9] Khaled Elshazly1, Yasser Fouad2, Mohamed Saleh3, Adel Sewisy4, "A Survey of SQL Injection Attack Detection and Prevention"
- [10] JAVACC FAQ. <http://www.engr.mun.ca/~theo/ JavaCC-FAQ/javacc-faq-moz.htm>.

Biometric Identification And Real Time Control And Security System In Smart Car (Bircs3)

Riya Parikh

Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai

Pooja Kammala

Department of Information Technology,
Thakur College of Engineering & Technology
Mumbai

Manasi Kadam

Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai

Neha Patwari

Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai

Abstract- In the era of globalization the vehicle market is booming. There is rapid increase in the number of vehicles. In the last two decades number of vehicles running on road has increased in multiple of population. As the number of vehicle increases, the number of theft attempts has also increased all around the world, hence security and protection of the vehicles has gained a great importance nowadays. So to protect vehicles from the theft we have designed a security system for the vehicle and named it as Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3). This includes i) Authentication (Face Detection System), ii) prevention (Buzzer ringing and Ignition On and off) and iii) Android Application (Engage with vehicle). The system has Mobile Phone that is installed in the vehicle. The vehicle stuff stolen can be stopped by using GPS full-length of mobile phone & this information is used by owner of the vehicle for remoter processing. The owner sends a notification to mobile which is used in the vehicle that has stolen which in turn controls the engine of vehicle by locking it immediately. The engine can be unlocked by only by the sending the password to microcontroller by Authorized person only. The goal behind the design is to develop security for vehicle and Embedded system to communicate with engine of vehicle.

Key Words – Vehicle, Security, Owner, Theft, Identification, Android, GPS

I. INTRODUCTION

A huge number of cars are lost every year in the state and a large number of automobiles are additionally recuperated by the Police from when they get the thief or not notwithstanding when the offenders leave the vehicles they have stolen after they have utilized them. The typical issue with the recouped vehicles achieving the real proprietors is that the vehicle require not be found in an indistinguishable ward from one in which the protest was propelled. Along these lines, when a vehicle is recouped, typically the Police attempt to follow out the real proprietor of

the vehicle from the RTO in view of the permit and case number. In any case, this is a protracted and tedious process for the RTO to follow out the real proprietors from the records and advise back to the Police stations. In light of these deferrals, vehicles that are recouped all lengthy time-frame to really achieve their proprietors. In spite of the different advancements that have been acquainted as of late with distinguish auto burglaries and following it, It was accounted for that the same number of autos were stolen yearly on the planet. This venture comprises of an android based remote vehicle framework will give viable, ongoing vehicle area, mapping and detailing this data esteem and include by enhancing the level of administration gave. A vehicle following framework will advise where your vehicle is and where it has been, to what extent it has been. The framework utilizes geographic position and time data from the Global Positioning Satellites. The framework has an "On-Board Module" which dwells in the vehicle to be followed and a "Base Station" that screens information from the variegated vehicles [5].

II. LITERATURE SURVEY

Vehicle security [3][4] is unchangingly been an important priority in the automobile industry. Various techniques like inside locking system with watchtower were one of the security parameter, which could only protect versus thefts only when the vehicle was stationery. However, to alimony in touch with a remote vehicle and track its other aspects like speed and location are stuff ripened and tested Today 's generation phones are not only capable of sending mails, making phone or video calls but moreover have the sufficiency to tenancy other smart phones. In this project we introduce a new efficient technique to dis-engage a remote vehicle using android technology. The parameters used for measuring the result was the squatter image placid from the users and stored them in database and then using these images as a way to demonstrate the users [1].

A. EXISTING SYSTEM

Vehicle security is always being an important priority in the automobile industry. Various techniques like central locking system with alarm were one of the security parameter, which

could only protect against thefts only when the vehicle was stationery. Untracking Vehicle Tracking Unit has the worthiness to integrate the GPS tracking system with existing vehicle watchtower or provide watchtower features when someone is tampering with owner vehicle. It allows detecting the security threat surpassing the vehicle is driven yonder and gives the worthiness to track the vehicle over the internet. The existing system has, Car watchtower techniques are used to prevent the car theft with the help of variegated type of sensors like pressure, tilt and shock & door locking system, and GPS systems which provides a wholesale layout of geographical address. If Owner vehicle has been stolen then he/she must be file FIR to the police station.

B. DRAWBACKS OF EXISTING SYSTEM

The existing system does not provide any security to two wheelers vehicles. The forfeit and maintenance of the system is high. Owner must require statement to the police and wait for whoopee that will be taken from the police. The GPS system does not provide street-wise address. Owner doesn't have knowledge well-nigh his vehicle until the police take action.

C. PROPOSED SYSTEM

The proposed system is Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3) a new and self-contained product that gives complement to the existing system. The product combines the installation of an electronic device in a vehicle, with purpose-designed android software (application) at least at one operational base to enable

III . MOTIVATION

Face recognition has been a sought without problem of biometrics and it has a variety of applications in modern life. The problems of squatter recognition attract researchers working in biometrics, pattern-recognition held and computer vision. Several squatter recognition algorithms are moreover used in many variegated applications untied from biometrics, such as video compressions, indexing's etc. They can moreover be used to classify multimedia content, to indulge fast and efficient searching for material that is of interest to the user. A squatter recognition system can be of unconfined help in forensic sciences, identification for law enforcement, surveillance, hallmark for financial and security system, and giving preferential wangle to authorized users i.e. wangle tenancy for secured areas etc. The problem of squatter recognition has gained plane increasingly importance without the recent increase in the terrorism related incidents. In today's fast pacing world increasingly and increasingly people are moving into the smart phone technology. This is due to various applications in the smart phones that allows users to tenancy and automate stuff virtually them at the printing of the button, reducing the time they spend on the work they consider unproductive.

IV. METHODOLOGY

The proposed solution, Biometric Identification and Real time Control And Security System in Smart Car (BIRCS3) is mainly used for securing car in case of theft. If the person who enter in the car is owner then the security system will be enabled that is the scanner will scan the image of a person because as he/she is an owner his image will be already stored in database. So the system will consider him/her as a authorized person then the car will start automatically.

If the person is relative, by using android application the owner will give access by locking or unlocking the car. Then scanner will scan the image of a relative and store that information in database. Whenever in future that person wants to use car he will be considered as authorized person by system. If the person is unknown (stranger/thief), the security system will be enabled i.e., the scanner will scan the image, if the scanned image is not matching with the images stored in the database the system will consider the person as unauthorized and then immediately the control system will enabled which will lock the doors and windows of the car and the alarm will ring and the scanned image of the unauthorized person will be send to the owner to his mobile app and to the police. Android application is for authorized person that should be installed in user mobile phone.

The person is authorized then following features will be provided by using android application. Such as a navigation application system where a user can search for a location that he or she wants to visit, messaging system where an application will offer the colleague/friend contact number to send the estimated delay time due to traffic.

Hardware:

- a) *Power Supply*
- b) *USB To TTL Type USBMOD3*
- c) *Arduino Atmega328*
- d) *Motor driver IC L293D*
- e) *Dc motor*
- f) *Buzzer*

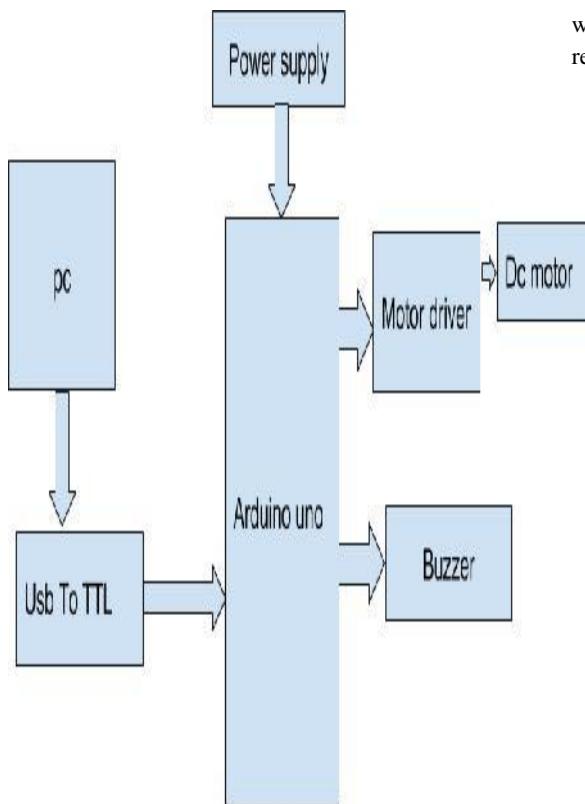


Figure 1: Block diagram of hardware kit

The Hardware kit is connected in following steps as shown in Figure 1 :

1. The Arduinouno is connected to PC using USB To TTL.
2. The power supply is connected to Arduinouno.
3. The buzzer and Dc motor is connected to Arduino for authorization
4. The DC motor is connected to Motor Driver for direction i.e. to move in clockwise or anticlockwise.

Software:-

Android Studio
MATLAB

The Global Positioning System (GPS) is a satellite based navigation system consists of a network of 24 satellites located into orbit. The system provides essential information to military, starchy and commercial users virtually the world and which is freely wieldy to anyone with a GPS receiver. works in any weather circumstances at anywhere in the world. Normally no subscription fees or system charges toutilize GPS. A GPS receiver must be locked on to the signal of at least three satellites to estimate 2D position (latitude and longitude) and track movement. With four or increasingly satellites in sight, the receiver can determine the user's 3Dposition (latitude, longitude and altitude). Once the vehicle position has been determined, the GPS unit can determine other information like, speedloftiness to destination, time and other. GPS receiver is used for this research

work to shift the vehicle location and provide information to responsible person.

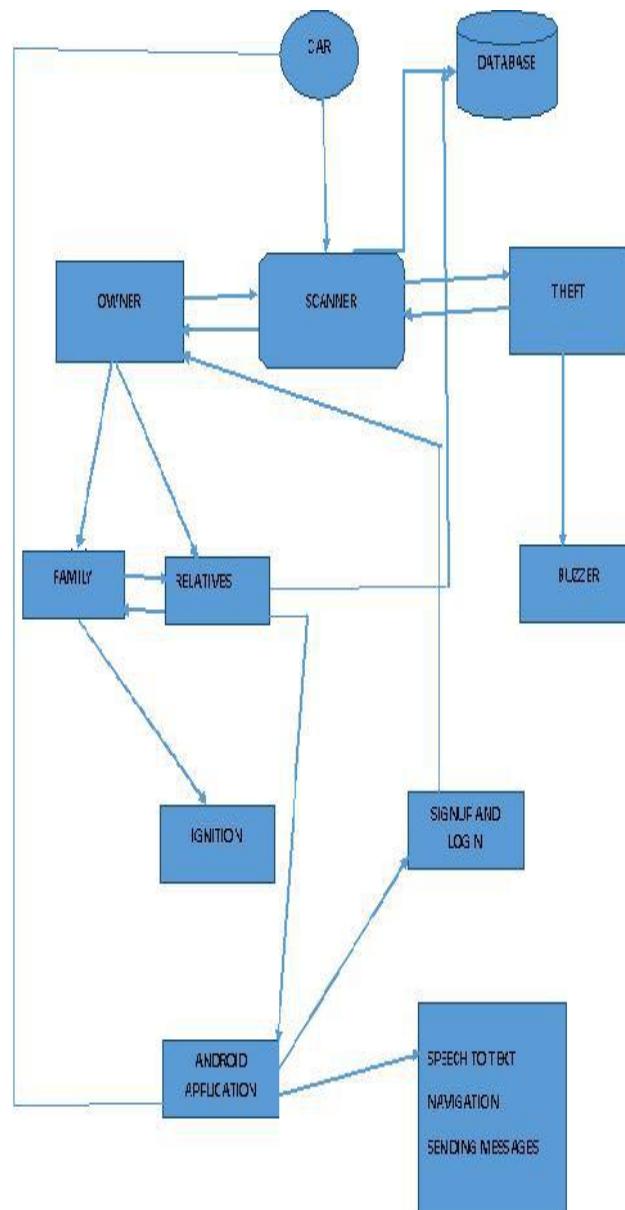


Figure 2: Flow chart for Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3)

- a) GPS Technology
- b) ATmega328

The ATmega328 is a single tweedle microcontroller created by Atmel and belongs to the mega AVRseries. The Atmel 8-bit AVRRISC-based microcontroller combines 32 KBISP winc memory with read-write abilities, 1 KBEEPROM, 2 KBSRAM, 23 stipulated purpose I/O lines, 32 unstipulated purpose working registers, three flexible timer/counters with

compare modes, internal and external interrupts, (one after the other)programmable USART, abyte-oriented 2-wire (one after the other)(connecting point/way of interacting with something), SPI(one after the other)port, 6-channel 10-bit A/D converter ,programmable watch dog timer with internal oscillator, and fives of tware selectable power saving modes. The device operates between 1.8-5.5 volts. The device(accomplishes or gains with effort)throughput sunes capable 1 MIP Sper MHz). Android (Automated Numeration of Data (understood/made real/achieved) by Much-improved Image Detection) Android is an operating system for mobile devices suchas Smartphone and table computers. It is ripened by the Open Handset Friendly partnership led by Google. Android consists of a kerne lbased on the Linux kernel, with middleware ,libraries and Computer Engineering and SmartSystems APIs written in Candusings of ware runnin go nanusing (solid basic structure on which bigger things can be built) which in cludes Java- compatible libraries based on Apache Harmony. Android uses the Dalvik virtual machine with just-in-imetcollection/creationtorunDalvikdex-code(DalvikExecutable),which is usually translated from Java byte code.

d) Matlab

MATLAB is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C and FORTRAN. The Image Processing Toolbox is a hodgepodge of functions that proffer the capabilities of the MATLAB's numeric



Figure 3 Image Processing Toolbox

computing environment. The toolbox supports a wide range of image processing operations, including: Geometric operations, Neighborhood and woodcut operations, linear filtering and filter design, Transforms, Image wringer and enhancement, Binary image operations – Region of interest operations

V. RESULT



Figure 4 Hardware kit

The Figure no:3 shows that If the image in the trained database is matched with the image in the test database then the ignition will start and if the image doesn't match then the ignition will stop and buzzer will ring. The Figure no:4 represent the hardware kit where the Arduino Atmega328 is connected to the computer using the USB To TTL Type USBMOD3 and programming has been implemented in the matlab.

VI. CONCLUSION

Thus the low cost proposed system acts as an anti-theft and stolen vehicle recovery system which can be used to track a vehicle by simply sending a notification to the proposed embedded device in the vehicle. With the help of android application in both the cases of personal as well as business purpose improves safety, security, communication medium, performance monitoring and recovery of vehicle. Updating this setup is very easy which makes it open to future requirements without the need of rebuilding everything from scratch, which also makes it more efficient.

VII. ACKNOWLEDGMENT

The people we would to thanks are many for their help in presenting this project. We specially want to thanks our Principal Dr. B. K. Mishra, our HOD-IT Dr. Rajesh Bansode and our project guide Mrs. NehaPatwari whose support have inspired and shaped much of the content of this project moreover guiding us on this project work and spending hours working to modernize the content of this project. Moreover we want to thanks to all the staff members who help us to improve.

REFERENCES

- [1] Jian Xiao, Haidong Feng."A Low-cost Extendable Framework for Embedded Smart Car Security System ,"*Proceedings of the 2009*

- IEEE International Conference on Networking, Sensing and Control,*
March 26, 2009, pp: 2320–088X.
- [2] Raquib B., Soumyajit R., Parthib M., Subhajit B., Abhishek M., Sauvik D." Implementation of MATLAB based object detection technique on Arduino Board and iROBOT CREATE ,"*IJSRP*, Vol. 4, Issue 1, Jan 2014.
 - [3] Joseph A. O'Sullivan, Robert P., "Advances in Security Technologies: Imaging, Anomaly Detection, and Target and Biometric Recognition ,"*Microwave Symposium IEEE/MTT-S International Volume*, 2012.
 - [4] Jayashri. M. Ahire, Swapnali.S.Baste, Krutika. S. Bhadke , Pratibha. G. Shelake - "Android App for Stolen Vehicle Tracking and Engine-Disengaging System ,"*"IJRITCC February 2015*, Available <http://www.ijritcc.org>
 - [5] J.-P. Hubaux, S. C Apkun, and J. Luo, "The security and privacy of smart vehicles," *IEEE Security & Privacy Magazine*, vol. 2, no. 3, pp. 49–55.
 - [6] W. Stallings, — "Cryptography and Network Security ,"*Prentice-Hall*, Englewood Cliffs, NJ, USA, 4th edition
 - [7] <http://www.androidauthority.com/best-car-apps-android>

Sign Language Recognition Using Computer Vision

Radhika Gokani

Information Technology Department

TCET

Mumbai, India

radzzg@gmail.com

Dinesh Gowda

Information Technology Department

TCET

Mumbai, India

dineshgowda03896@gmail.com

Bhavesh Mishra

Information Technology Department

TCET

Mumbai, India

bhaveshmishra11496@gmail.com

Sudhir Dhekane,

Information Technology Department

TCET

Mumbai, India

sudhir.dhekane@thakureducation.org

Abstract-

The blind, deaf & mute community have been experiencing problems in communication for decades. Since gestures are powerful means of communication among peoples, the problems of such community can be solved by providing an output. For mute peoples, the signs can be converted to speech output. Blind peoples can hear the signs voice output. Deaf peoples can read the text display of signs.

The sign language recognition system will enable them to comprehend signs in the form of speech and text display. The training data set of images that is used has 5 gestures, each with 50 variations of a single gesture with different lighting conditions. The purpose of this is to improve the accuracy of classification.

Results reported by different researchers in their research papers shows different percent of accuracy based on the algorithm and methodology used. American Sign Language Recognition using Support vector machine and back propagation algorithm shows highest percent of accuracy.

Letters Accuracy achieved using SVM was 92.882 and back propagation accuracy was 96.882 for letter A. Our proposed system used Contour algorithm to detect and extract image called as feature extraction which showed 70% of accuracy while research papers used Back Propagation Neural Network Technique and Glove based systems which showed 98.61% accuracy.

Keyword:SVM;sign language recognition; Contour algorithm

I. INTRODUCTION

It is important to establish communication with Deaf and Dumb people through hand gestures or signs. Gestures are basically the physical action form performed by a person to convey some meaningful information. Gestures are a powerful means of communication among humans. This Paper will help us to recognize gestures and translate them to meaningful description and voice output. Hand gesture recognition, pattern recognitions, image processing, feature extraction will be carried out to capture image/gesture. The motivation for developing such helpful application came from the fact that it would prove to be of utmost

importance for socially aiding people and how it would help increasingly for social awareness as well.

A. BACKGROUND

A robotic hand with the capability of spelling words using the manual alphabet was developed in 1977. Fingers spelling hands were created to Assist deaf/blind individuals by replicating the hand on-hand interpreting of the manual alphabet of a human (Jaffe, 1994).A simplistic glove designed by Ryan Patterson was developed in 2002 in the United States Now Sign Language detection is done with the help of feature extraction and neural network algorithms .Nowadays gesture detection is mostly done through feature extraction algorithms and classification algorithms like regression, neural network, Scale invariance Fourier transform etc. A real-time sign language translator is an important milestone in facilitating communication between the dumb and deaf community and the general public. Computer recognition of sign language is an important research problem for enabling communication with hearing impaired people. The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars. This is because normal people are usually unaware of these grammars. As a result, it has been seen that communication of a dumb person are only limited within his/her family or the deaf community.

B. SCOPE

Our model is based on decision tree and pattern matching algorithm. Accuracy is based on how clear and proper image or hand gesture is obtained through webcam. Description of gesture will be displayed on computer screen based on trained dataset Input and a voice output of gesture will be provided. Back propagation algorithm will be used to train algorithm for detection purpose.

C. IMPORTANCE & MOTIVATION

The motivation for developing such helpful application came from the fact that it would prove to be of utmost importance for socially aiding people and how it would help increasingly for social awareness as well. The remarkable ability of the human vision is the gesture recognition, it is noticeable mainly in deaf people when they communicating with each other via sign language and with hearing people as well.

II. OVERVIEW

Computer recognition of sign language is an important research problem for enabling communication with hearing

impaired people. The problem arises when dumb or deaf people try to express themselves to other people with the help of these sign language grammars. This is because normal people are usually unaware of these grammars. As a result, it has been seen that communication of a dumb person are only limited within his/her family or the deaf. The remarkable ability of the human vision is the gesture recognition, it is noticeable mainly in deaf people when they communicating with each other via sign language and with hearing people as well. Gesture will be displayed on computer screen based on trained dataset Input and a voice output of gesture will be provide language translator is an important milestone in facilitating communication between the dumb and deaf community and the general public

III. LITERATURE SURVEY

The paper presents an algorithm of Hand Gesture Recognition by using Dynamic Time Warping methodology. The system consists of three modules: real time detection of face region and two hand regions, tracking the hands trajectory both in terms of direction among consecutive frames as well as distance from the centre of the frame and gesture recognition based on analyzing variations in the hand locations along with the centre of the face. Experimental results show that the accuracy is 90% in recognizing 24 gestures based on Indian Sign Language [1].

IV. PROPOSED SOLUTION

This paper introduces an efficient and fast algorithm for identification of the number of fingers opened in a gesture representing an alphabet. The basic objective of this paper is to develop a computer based detection system that will enable dumb people significantly to communicate with all other people.

V. METHODOLOGY

. By this, with the number of fingers extended we can perform and label it with different function and text and speech output. Capturing of image is the input of the proposed system and output of the system will be recognizing given inputs and provide a display

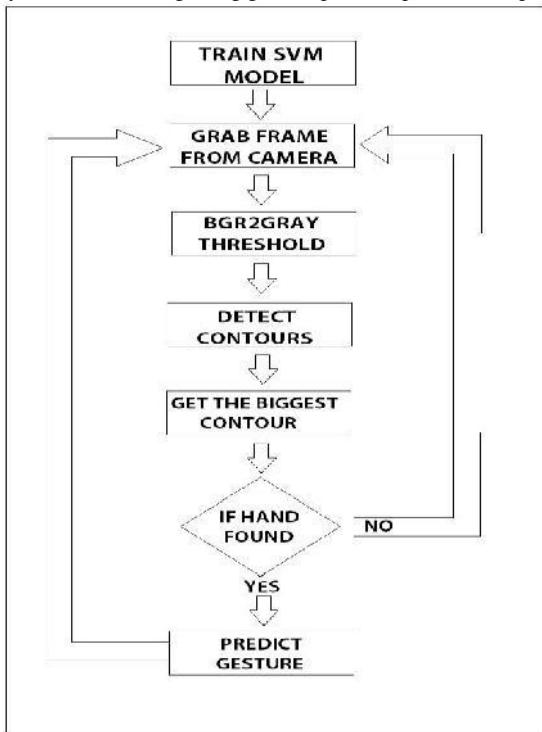


Figure 1: Flowchart

output description of input on monitor and also providing voice output of the gesture. Processing of the input will be done with the help of Contour Algorithm which will find the points and then match it with the already fed database and produce the image. The original image is blurred using Gaussian blurring to remove noise. We then create thresholding for image segmentation. grayscale. It then converts it to binary for processing of Thresholding allows the only particular color to be highlighted as white and other colors as black. OpenCV automatically calculates a bimodal image from its image histogram. With the help of contours detection, we draw a convex points and defects points. Figure 1 shows how the sign detection system first captures the images of the hand and converts them to The convex points are generally the edges of gestures. Then convexity defects which are the deepest point of deviation on the contour

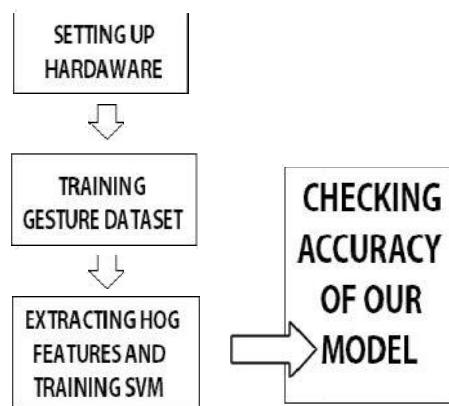


Figure 2: shows the process of training the model using training data and implementing the model for getting the testing data.

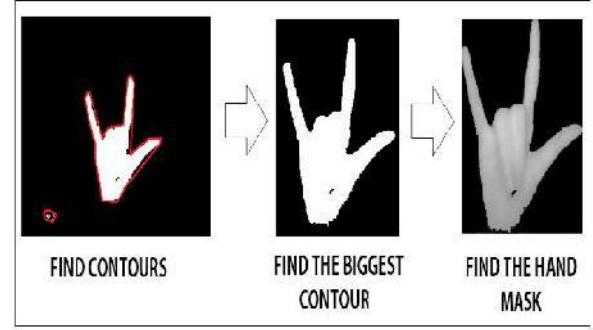


Figure 3(a):Steps

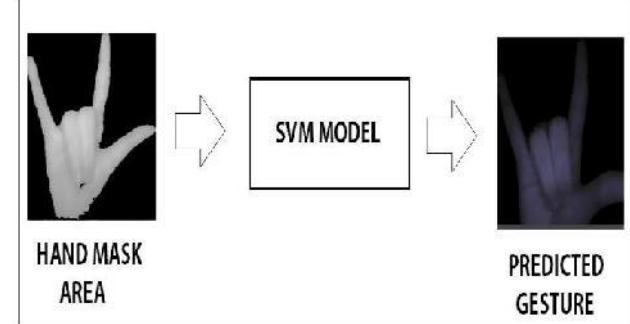


Figure 3(b):Steps

Figure 3(a) shows how the contour algorithm captures hand gestures and finds out the key points of hand gestures. Figure 3(b) shows how

VI. RESULTS

After implementing the algorithms, the following results are obtained

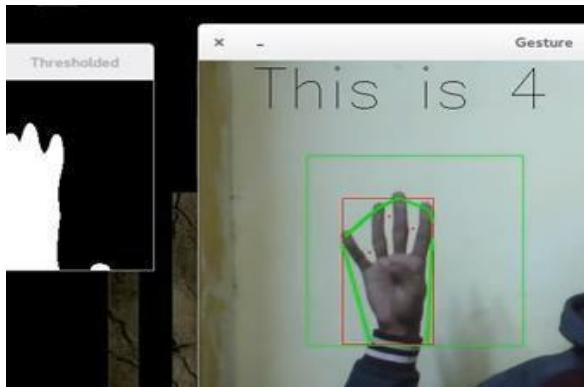


Figure 4: Display of input

Figure 4 shows the input detection of sign gesture of 4 and shows the description of gesture.



Detected Sign displayed as “Hi”

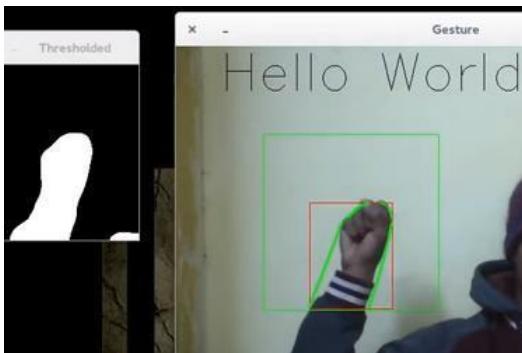


Figure 6: Detected Sign displayed as “Hello World”

Figure 5 and 6 shows detected sign gesture input and output description of two different signs.

the detected hand mask area is converted to the predicted point gesture with help of Support Vector Machine.

VIII. CONCLUSION

We have achieved how to detect sign gestures by implementing the contour algorithm. Our system has achieved 70% of accuracy which first detects and then extracts keypoints which is referred as Feature Extraction process. We have demonstrated this with help of few different signs.

REFERENCES

- [1] WashefAhmed,kunalchanda,somamitra , "Dynamic time warping based sign language detection" published in IEEE in 2016 ba
- [2] Ellen walker," Recognizing American Sign Language Letters: A Machine Learning Experience in an Introductory AI Course" published in Florida Artificial Intelligence Research Society Conference 2015.
- [3] SabahetaDogic,Gunaykarlipublished“Backpropogation based sign language recognition” in year 2015.
- [4] Kamalpreet Sharma, Naresh Kumar Gard published “American Sign Language Recognition using Support vector machine and back propagation algorithm” in year 2014
- [5] Sulfayanti, Dewiani, Armin Lawi Published “Sign language detection based on vector angle & Euclidean distance” in year 2016.
- [6] Sakshi Goyal, Ishita Sharma ,Shanu Sharma published “Sign Language Recognition System For Deaf And Dumb People” in Year 2015
- [7] KjrlfdrnFCM , “Visual interpretation of hand gestures as a practical interface modality,” Ph.D. thesis, Columbia University, New York,1997 accessed on 10/9/17.
- [8] Yaniar, Ninar.S. 2011. Comparison distance measurements on Facial Recognition Process. Proceeding Seminar TugasAkhirJurusanTeknikElektro FTI -ITS. Surabaya.(in Bahasa Indonesia) accessed on 19/9/17.
- [9] David G.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 2004, pp 1-28. Accessed on 21/9/17.

Design of Guest Tagging System For Surveillance Using IoT Analytics

Prachi Jain <i>IT Department</i> <i>Thakur College of Engineering and Technology</i> <i>University of Mumbai, India</i>	Lokesh Ucheniya <i>IT Department</i> <i>Thakur College of Engineering and Technology</i> <i>University of Mumbai, India</i>	Satyendra Yadav <i>IT Department</i> <i>Thakur College of Engineering and Technology</i> <i>University of Mumbai, India</i>	Akash Singh <i>IT Department</i> <i>Thakur College of Engineering and Technology</i> <i>University of Mumbai, India</i>
--	--	--	--

Deven Shah
IT Department
Thakur College of Engineering and Technology
University of Mumbai, India

Abstract - Govardhan Eco-village needs to find an eco-friendly cost-effective solution to tag the different categories of guests arriving, and granting them various permission access to different zones in the village. Every type of guest needs to be served in a more convenient manner so as to maximize the level of user experience at the village which is an attraction place for hundreds of devotees. After solving the above problem, we realized Crowd management is a very complex, challenging and costly exercise. If not managed or controlled effectively, it can result into crowd disasters aka stampede like situations. The recent Elphinstone Rail Stampede in Mumbai, 29 September 2017 is an apt example to illustrate the same.

The aim of our proposed innovation is to provide a crowd management framework application which predicts decisions on how to distribute the crowd evenly in such spatially constrained areas of the village and also provide maximum user experience in such scenario.

Keywords - Artificial intelligence, Business Intelligence, Crowd Risk, Crowd Disaster, IoT Analytics, Mass Gatherings, Neural Networks, Prepared Plan Evacuation, Precaution Plan Evacuation, Rescue Plan Evacuation, Simulation, Stampede, Splunk, Supervised Learning

I. INTRODUCTION

Govardhan Ecovillage home to a Wellness retreat and sustainable farming community is situated near Mumbai. Considering the present scenario at Govardhan Ecovillage, it has a manual way of authorizing the visitors by providing them simple Identity cards for authentication purposes. Being an 85-acre farming community, Govardhan Ecovillage (GEV), does not have an adequate way of controlling the access of the visitors and requires a system which can automatically manage the visiting crowd. Around a few hundred guests visit daily.

While finding the solution we realized, large accumulation of crowds at a specific area can prove to be dangerous. If the crowd is big enough, it can exert enough pressure to bend steel. Crowd management is essential to weaken this "Crowd Risk".

Crowd Disasters are tragically common in various parts of the world, including India, where heavy population density and a high tolerance for crowded spaces create gnarly conditions that can easily turn deadly. Re

Some Aspects of Crowd Management are:

1. *Planning*— Spatial, temporal and event related dimensions have to be considered in the planning process. The space which is going to host the crowd has to be studied first. Additional planning is required if any sudden surge of activity is anticipated at any point of time. Besides, special attention has to be given if there is a specific event happening in that space-time context.

Sudden unforeseen emergencies (like a fire) can also trigger incidents like a human choking, stampede. Even when such incidents cannot be predicted, they should always be anticipated and prepared for

- Enlarging or creating extra space for the event, counters, extending the time for the event, planning for large crowds, and anticipating whether any particular event can trigger sudden or unusual movement are the possible solutions.

II. PROBLEM DEFINITION

The objective of the problem definition is to find an eco-friendly cost-effective solution to tag the different categories of guests that come to Govardhan Eco Village, namely daily visitors, residential guests, unregistered day visitors, villagers and VIPs. Each category of guests will have access to different zones in the village.

Every type of guest needs to be served in a more convenient manner so as to maximize the level of user experience at the village which is an attraction place for hundreds of devotees.

III. PROPOSED SOLUTION

A. Design of Proposed Solution

The proposed solution uses a RFID tag and Long-Range UHF Passive reader to record the current location and timestamp of the guests (visitors). RFID Reader with a range of about 25 m will be present at a center of each prominent location of the Govardhan Eco-Village, thus covering a diameter of about 50 m. Each RFID Tag will be mapped with the details of the Visitor to whom it is assigned.

RFID reader will be used to record the RFID Tag's identity and current timestamp. RFID Reader will send this data along with its deployed location to a centralized database (Server) so that any kind of user can be tracked and their access can be controlled to different zones of the area. Long range communication will be implanted by line of sight communication whereas short range by using the radio frequency communication in walky-talkies which is already used in the village.

Real Time Analytics will be performed on data acquired so as to manage the flow of users in different zones of area effectively, reduce the waiting time of guests and serve the visitors in a more convenient manner.

For e.g., if the number of users is more in the Prasad Ghar of the village than its holding capacity we can move them to another area for other services.

Currently the entire scenario is simulated as a smart city model on Cup Carbon U-One 3.8, a Smart City and IoT Wireless Sensor Network Simulator Tool.

Also, Data Visualization in the form of charts and graphs will be generated automatically to provide knowledge which in turn will help in making more intelligent, optimized decisions.

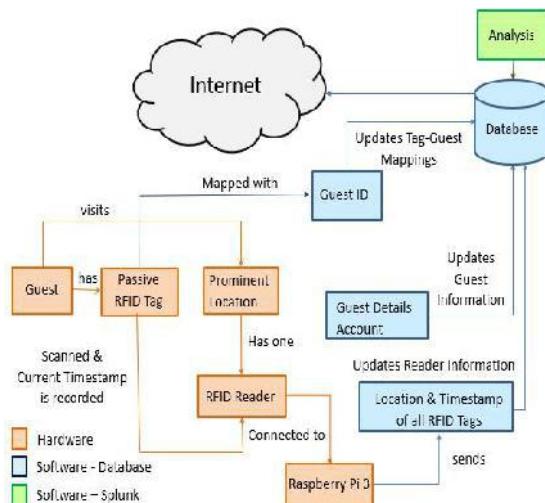


Fig. 1. Workflow of Guest Tagging System Design for Surveillance using IoT

A Guest visits a prominent location. Each guest is provided with the Passive RFID tag which is mapped with guest details. RFID reader operated via a Raspberry Pi at every prominent location of the village reads each tag and sends the tag's

identity, current timestamp recorded and location to the database. Universal forwarders are used to collect all data from the database and provide it to Splunk for Data Processing and Analysis.

B. Importance of the Proposed Solution

Proposed solution incorporates the following features:

1. Crowd Management and Control
 - Crowd management refers to situations where the evacuations places for the crowd is relatively more and therefore the crowd gets various options to ease itself.
 - Crowd Control refers to scenarios where evacuations places for the crowd is relatively less and therefore the crowd has few or one option(s) to ease itself. Hence, crowd is forced to follow a particular plan.
2. Prevention, Avoidance and Detection of Crowd Disasters (Stampede like phenomenon).
3. Real time with Archived Data Analytics

The proposed innovation provides real time analysis and predictions useful to generate faster response time and effectively manage and control the crowd. Also, archived data is used for analytics purpose to generate more accurate predictions.

4. Track or Monitor Individual User Movements

At the granular level, with the support of high computational power we even can able to track individual person movements as well.
5. High Scalability and Interoperability

The proposed application can be integrated to any RFID System which track its users automatically or needs to swipe its RFID card such as Railway Stations which uses NFC cards, Attendee Tracking in any mass-gathering event etc. The proposed application can also be integrated to location aware mobile system, Wi-Fi membership etc.

6. Predicts and recognizes various Knowledge Patterns of a crowded environment visually on the dashboard such as:
 - Peak time of crowd gatherings.
 - Average waiting time of the user in a queue.
 - Average time spent by the user in an area.
 - Most popular and populated places under the area being monitored.
 - Preferences of the crowd for the services offered under any RFID System (or other means like location aware mobile system, Wi-Fi membership) etc.

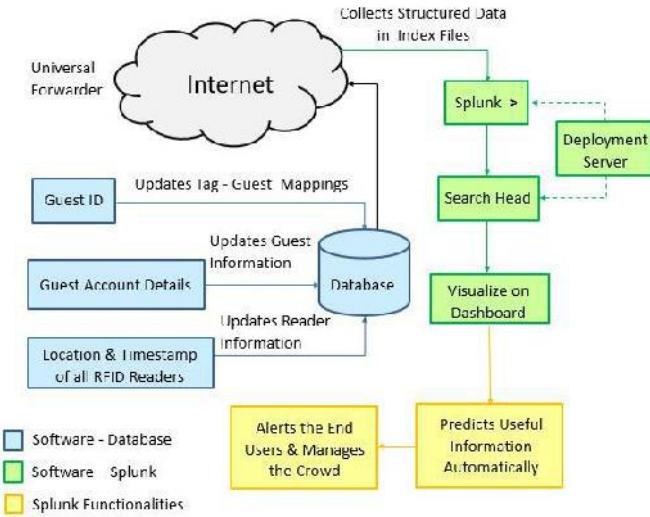


Fig. 2. Workflow of Guest Tagging System Analysis for Surveillance using IoT.

All the collected data forwarded to Splunk is indexed and a prediction algorithm is deployed to distribute the crowd adequately and provide maximum user experience. The output of the prediction algorithm are alerts, statistics etc. sent to GEV administrators.

IV. METHODOLOGY

The aim of our proposed innovation is to provide a crowd management framework application which predicts decisions on how to distribute the crowd evenly in spatially constrained areas and also provide maximum user experience in such scenario in order to prevent, detect and avoid Crowd Disasters (stampede like situation).

The proposed application will be in the form of two-layer neural network.

1. In public places, we need to get the information about number of people in particular time and collect it by various means, like RFID (as part of Smart card based ticket), location aware mobile system, Wi-Fi membership etc.
2. Multidimensional input data from any RFID system or other means as discussed, about its user, location, current timestamp and RFID tag/card or other means as discussed above is mapped with user details and composes of the input layer of the neural network.
3. The first stage trains the neural network to be a simulator of the area which requires crowd management. Each area is allocated a weight/cost based on its actual crowd holding capacity and the current crowd is monitored under the area based on supervised learning process.
4. The second stage enables the neural network controller to learn to predict crowd management or evacuation plans by using simulator as a guide. It controls the actual dynamic system.
 - a. The control process consists of feeding the current state vector to the controller which in turn provides an

evacuation plan for Crowd Disaster prevention, avoidance and detection – Prepared Plan Evacuation, Precaution Plan Evacuation and Rescue Plan Evacuation respectively.

b. Prepared Plan Evacuations are mainly based on the geographical and dimensional structure of areas to arrange evacuation entrances and exits reasonably, inspect hardware regularly, etc.

c. Precaution Plan Evacuations is mainly based on real-time monitoring. When the monitoring system finds that the population density in one region has reached a certain threshold, or there are abnormal movements among people, actions will be taken immediately to evacuate the crowd and thus avoiding a stampede.

d. Rescue Plan Evacuations refers to the situations where accident hasn't been effectively prevented during precaution evacuation. Once the monitoring system has caught a stampede, the arranged rescue evacuation plans should be immediately activated, and various rescue forces should be organized to rescue people.

4. The learning process continues as the simulator and controller improves and tracks the dynamic system.

Thus, the proposed Crowd Monitoring cum Management system predicts Crowd Disasters like stampede, helps in activating arranged rescue evacuation plans immediately and organizes various rescue forces to mobilize people and ease the crowd.

At the granular level, with the support of high computational power we even can able to track individual person movements.

- The project maintains two levels of users: -

o Administrator Level- Owner or Staff at GEV

o User Level- the Guests/ Visitors and the data-entry level operators

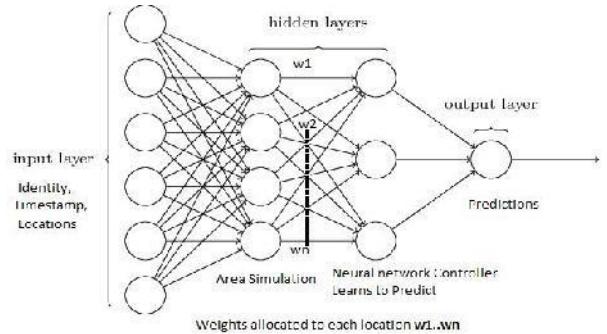


Fig. 3. Neural Network designing aka Prediction Algorithm

V. RESULTS AND DISCUSSION



Fig. 4. Simulation of 500 routes and 5 Readers to design the Smart City for Guest Tagging System on CupCarbon One 3.8



Fig. 5. Simulation of 500 Mobile devices to design the Smart City for Guest Tagging System on CupCarbon One 3.8 for a Simulation Time of around 100 Seconds

ID	Name	Address	Gender	Age
1	ARUN KHARA	AKOLA	M	51
2	ASHOK MARU	MUMBAI	M	49
3	RAMESH PARDHI	KHAJURHAO	M	31
4	SANJAY SHAH	KOPARGAON	M	15
5	AYUSH DOSHI	DHULE	M	9
6	MILIND SAWANT	LIJARDA	M	22
7	OMKAR JANGAM	MADRAS	M	33
8	SHALINI AYYAR	KOPARGAON	F	29
9	NAMITA JANI	UDUPIAD	I	1
10	PRIYA MADNANI	MANGLORE	F	22
11	SHREYA PATKAR	AGRA	F	50
12	DANINI KOHLE	INDORE	F	61
13	JANAK BHIMANI	DIVVAS	M	34
14	NARAY ANAND	UJJAIN	M	51
15	SANJEEV BHASIN	NIMACH	M	40
16	RAJESH KHAN	RATIAM	M	15
17	MILAN DESPANDEY	NAGDA	M	42
18	NAMAN GOUR	SURAT	M	49
19	AYAJ JAIN	DARODA	M	11
20	RAJESH BALAI	BARUCH	M	13
21	NITIN NANE	PALI	M	25
22	SUDHIL MANIK	JAMISLDPUR	M	21
23	PRakash REDKAR	BHILWARA	M	53
24	AYAJ JAIN	BHILWANDI	M	15
25	HASHMI KALDIYA	VALGHAR	F	24
26	RAMITA RAI	IAI SAON	F	31
27	FUSHPA MANI	BHUSAWAL	F	20

Fig. 6. 500 Guest Details namely Name, Address, Gender,Age mapped with 500 RFID identities in comma separated format.

TABLE 1. LOG FILES AND ITS ATTRIBUTES GENERATED WHILE SIMULATION OF SMART CITY ENVIRONMENT FROM REDAERS

ID	Position x	Position y	Locations (A, B, C, D, E)	Current Timestamp
1	-4.45	48.39	A	37.094064

a. Sample of 501, 502, 503, 504, 505 Readers dataset layout. (Table footnote)

Log File Content
1 #4-4.488806343078A#48.3976105F713476#2#0.0940F40000000000001 2 #4-4.48982768722534#48.39865058E05447#2#1.094C6399999999995 3 #4-4.48982768722534#48.39865058E05447#2#1.094C6399999999995 4 #4-4.488806343078A#48.3976105E713476#2#3.094C6399999999996 5 #4-4.48982768722534#48.39865058E05447#2#4.094C6399999999990 6 #4-4.4891166687C1172#48.399776E2D202664#2#5.094C6399999999997 7 #4-4.489202999396E#48.398062217377E#4#6.094C6399999999997 8 #4-4.488806343078A#48.3976105E713476#2#7.094C6399999999995 9 #4-4.48982768722534#48.39865058E05447#2#8.094C6399999999996 10 #4-4.4891166687C1172#48.399776E2D202664#2#9.094C6399999999992 11 #4-4.48982768722534#48.39865058E05447#2#10.094C6399999999999 12 #4-4.4891166687C1172#48.399776E2D202664#2#11.094C6399999999992 13 #4-4.49322270243C20#40.35403005C0003C#2#12.094C6399999999992 14 #4-4.49322270243C20#40.35865058E05447#2#13.094C6399999999992 15 #4-4.4891166687C1172#48.399776E2D202664#2#14.094C639999999994 16 #4-4.491511181945#48.399334F#2#14.094C639999999994 17 #4-4.49322270243C20#40.35433365E00335#2#16.094C639999999996 18 #4-4.48982768722534#48.39865058E05447#2#17.094C639999999996 19 #4-4.4891166687C1172#48.399776E2D202664#2#18.094C639999999996 20 #4-4.49151101094#40.359C46G394550#2#19.094C639999999992 21 #4-4.48787012371826#2#18.005738526977#2#20.0C10E3999C00998 22 #4-4.49613332748413#48.396599C21498214#2#21.0C40E3999C00998 23 #4-4.48982768722534#48.39865058E05447#2#22.094C63999999999 24 #4-4.4951C35522724#48.397396F#19R373#2#23.094C639999999985 25 #4-4.4891166687C1172#48.399776E2D202664#2#24.094C63999999999 26 #4-4.4915118181945#48.399346G#2#25.094C6399999999975 27 #4-4.4938575602722174#48.3908973E93221#2#26.094C63999999998 28 #4-4.4902795021C713#40.393094C7396201#2#27.094C63999999990 29 #4-4.48787012371826#2#18.005738526977#2#28.C010E3999C00999 30 #4-4.49613332748413#48.396599C21498214#2#29.C510E3999C009992 31 #4-4.4937704570225#48.398018617357#2#30.C640E3999C009992 32 #4-4.48982768722534#48.39865058E05447#2#31.094C63999999996 33 #4-4.4951133592224#48.39J396E119K3#3#32.094C63999999996 34 #4-4.4891166687C1172#48.399776E2D202664#2#33.094C63999999996 35 #4-4.4910S1101094#40.395204G599455#2#34.094C63999999996 36 #4-4.48831127716C#15#48.39138123581117#2#35.094C63999999996 37 #4-4.493857560272217#48.3908973E93221#2#36.094064 38 #4-4.4982795021C713#48.393094C7396231#2#37.094064

Fig. 7. Log file Generated from READER 1

```

1 #=-4.49332372436523#48.39433365600335#D#0.0940640000000001
2 #=-4.49400901794433#48.39525976109966#D#1.1257306666666666
3 #=-4.49400901794433#48.39525976109966#D#2.1573973333333325
4 26#-4.492979049682617#48.38041070655736#B#3.1573973333333316
5 26#-4.4940090179443364#48.39525976109966#D#4.1573973333333331
6 26#-4.492979049682617#48.38041070655736#B#5.157397333333333
7 29#-4.4940090179443328#48.39525976109966#D#6.157397333333333
8 32#-4.19220973486329#48.39395276009991#7.1573973333333329
9 59#-4.49400901794433#48.39525976109966#D#8.1573973333333329
10 6#-4.49400901794433#48.39525976109966#D#9.1573973333333325
11 26#-4.492979049682617#48.38041070655736#B#10.1573973333333321
12 29#-4.492979049682617#48.38041070655736#B#11.1573973333333325
13 32#-4.492206573486328#48.39395276009991#12.1573973333333325
14 49#-4.493588538818#48.39159798625123#B#13.1573973333333325
15 58#-4.49033976011963#48.38860313420963#D#14.157397333333329
16 81#-4.49238625005120963#D#15.1573973333333329
17 59#-4.4904041290283#48.3895888849852#16.1573973333333329
18 26#-4.492979049682617#48.38041070655736#B#17.15739733333333
19 29#-4.49405193388574#48.38041070655736#B#18.15739733333333
20 32#-4.492206573486328#48.39395276009991#19.157397333333325
21 49#-4.490599632096328#48.39395276009991#20.15739733333325
22 48#-4.49860331420963#48.39395276009991#21.15739733333318
23 58#-4.49033976011963#48.38916106396578#B#22.15739733333314
24 71#-4.49048959716797#48.38916106396578#B#23.15739733333318
25 61#-4.494976413189674#48.39125601716999#B#24.15739733333333
26 59#-4.4950501290283#48.3895888849852#16.157397333333307
27 26#-4.492979049682617#48.38041070655736#B#26.157397333333314
28 29#-4.49405193388574#48.38041070655736#B#27.15739733333314
29 32#-4.492206573486328#48.39395276009991#28.15739733333329
30 49#-4.49860331420963#48.39395276009991#29.15739733333325
31 48#-4.4980331420963#48.39395276009991#30.15739733333325
32 45#-4.494976413189674#48.39125601716999#B#31.15739733333333
33 58#-4.49033976011963#48.38916106396578#B#32.15739733333333
34 71#-4.49048959716797#48.38916106396578#B#33.15739733333333
35 81#-4.493386745452881#48.39125601716999#B#34.15739733333333
36 59#-4.49051290283#48.3895888849852#16.15739733333333
37 80#-4.49315071105957#48.387597043953#D#36.15739733333336
38 26#-4.492979049682617#48.38041070655736#B#37.15739733333336

```

Fig. 8. Log file Generated from READER 2

```

1 #=-4.488515853881836#48.39159798625123#C#0.0957306666666666
2 #=-4.4920249930964#48.39005221737764#C#2.1607306666666666
3 61#-4.49132666015625#48.38991916693571#C#3.1607306666666655
4 #=-4.4920249930964#49.39005221737764#D#4.1607306666666664
5 61#-4.4913492666015625#49.38991916693571#C#5.1607306666666664
6 #=-4.49040959716797#48.389161406396579#C#6.1607306666666664
7 #=-4.49040959716797#48.389161406396579#B#7.1607306666666664
8 #=-4.4980331420963#48.393900681221105#B#8.157397333333325
9 #=-4.494976413189674#48.39125601716999#B#9.157397333333333
10 28#-4.48776403557666#48.389303899658#10.16073066666665
11 7#-4.48920249938964#48.39005221737764#C#11.16073066666659
12 8#-4.49303017561615#48.39085705036145#B#12.16073066666659
13 56#-4.48958873787793#48.38932935490003#C#13.16073066666659
14 61#-4.4913492666015625#48.38991916693571#C#14.16073066666662
15 73#-4.48915958405451#48.3890331211975#C#15.16073066666662
16 28#-4.48776403557666#48.390303899658#42#16.16073066666662
17 7#-4.48920249938964#48.39005221737764#C#17.16073066666662
18 8#-4.49403047561615#48.39085705036145#C#18.16073066666662
19 56#-4.48958873787793#48.38932935490003#C#19.16073066666662
20 61#-4.4913492666015625#48.38991916693571#C#20.16073066666665
21 73#-4.49091598405451#48.3890331211975#C#21.16073066666640
22 76#-4.48654178004687#48.3885623933397#C#23.16073066666665
23 15#-4.48654178004687#48.3885623933397#C#23.16073066666665
24 28#-4.48776403557666#48.3939052976109996#D#24.16073066666665
25 56#-4.49058873787793#48.3939052976109996#D#25.16073066666664
26 #=-4.49339453207459#48.38894791981599#D#26.16073066666640
27 #=-4.49092024993964#48.39085705036145#D#27.16073066666640
28 #=-4.49403047561615#48.39085705036145#D#28.16073066666640
29 56#-4.48958873787793#48.38932935490003#C#29.16073066666665
30 61#-4.4913492666015625#48.38991916693571#C#30.16073066666666
31 55#-4.4933237243652#48.39433365600335#C#31.16073066666666
32 73#-4.49091598405451#48.3890331211975#C#32.16073066666666
33 76#-4.48654178004687#48.3885623933397#C#33.16073066666666
34 15#-4.49396102600998#48.39525976109996#C#34.16073066666666
35 28#-4.48776403557666#48.389303899658#42#35.16073066666666
36 3#-4.483945569720459#48.38884791981599#C#36.16073066666666
37 7#-4.48920249938964#48.39005221737764#C#37.16073066666666
38 8#-4.49403047561615#48.39085705036145#C#38.16073066666666

```

Fig. 9. Log file Generated from READER 3

```

1 #=-4.484288692474365#48.39722589721657#D#0.0957306666666667
2 #=-4.484288692474365#48.39722589721657#D#1.1257306666666667
3 #=-4.484288692474365#48.39722589721657#D#2.1257306666666667
4 #=-4.484288692474365#48.39722589721657#D#3.1257306666666667
5 #=-4.484288692474365#48.39722589721657#D#4.1257306666666667
6 #=-4.484288692474365#48.39722589721657#D#5.1257306666666667
7 #=-4.484288692474365#48.39722589721657#D#6.1257306666666667
8 #=-4.484288692474365#48.39722589721657#D#7.1257306666666667
9 #=-4.484288692474365#48.39722589721657#D#8.1257306666666667
10 #=-4.484288692474365#48.39722589721657#D#9.1257306666666667
11 #=-4.484288692474365#48.39722589721657#D#10.1257306666666667
12 #=-4.484288692474365#48.39722589721657#D#11.1257306666666667
13 #=-4.484288692474365#48.39722589721657#D#12.1257306666666667
14 #=-4.484288692474365#48.39722589721657#D#13.1257306666666667
15 #=-4.484288692474365#48.39722589721657#D#14.1257306666666667
16 #=-4.484288692474365#48.39722589721657#D#15.1257306666666667
17 #=-4.484288692474365#48.39722589721657#D#16.1257306666666667
18 #=-4.484288692474365#48.39722589721657#D#17.1257306666666667
19 #=-4.484288692474365#48.39722589721657#D#18.1257306666666667
20 #=-4.484288692474365#48.39722589721657#D#19.1257306666666667
21 #=-4.484288692474365#48.39722589721657#D#20.1257306666666667
22 #=-4.484288692474365#48.39722589721657#D#21.1257306666666667
23 #=-4.484288692474365#48.39722589721657#D#22.1257306666666667
24 #=-4.484288692474365#48.39722589721657#D#23.1257306666666667
25 #=-4.484288692474365#48.39722589721657#D#24.1257306666666667
26 #=-4.484288692474365#48.39722589721657#D#25.1257306666666667
27 #=-4.484288692474365#48.39722589721657#D#26.1257306666666667
28 #=-4.484288692474365#48.39722589721657#D#27.1257306666666667
29 #=-4.484288692474365#48.39722589721657#D#28.1257306666666667
30 #=-4.484288692474365#48.39722589721657#D#29.1257306666666667
31 #=-4.484288692474365#48.39722589721657#D#30.1257306666666667
32 #=-4.484288692474365#48.39722589721657#D#31.1257306666666667
33 #=-4.484288692474365#48.39722589721657#D#32.1257306666666667
34 #=-4.484288692474365#48.39722589721657#D#33.1257306666666667
35 #=-4.484288692474365#48.39722589721657#D#34.1257306666666667
36 #=-4.484288692474365#48.39722589721657#D#35.1257306666666667
37 #=-4.484288692474365#48.39722589721657#D#36.1257306666666667
38 #=-4.484288692474365#48.39722589721657#D#37.1257306666666667

```

Fig. 10. Log file Generated from READER 4

```

1 #=-4.1745E5151660156#48.3968E2595178725#E#3.1032066666666665
2 #=-4.1745E5151660156#48.3968E2595178725#E#4.1257306666666665
3 14#-4.1747E20391363852#48.3900101598033386#E#5.1257306666666665
4 1#-4.1745E5151660156#48.3968E2595178725#E#6.1257306666666665
5 14#-4.1747E20391363852#48.3900101598033386#E#7.1257306666666665
6 1#-4.1745E5151660156#48.3968E2595178725#E#8.1257306666666665
7 1#-4.1747E501606860619#48.398E7935E50155#E#9.1257306666666665
8 14#-4.1747E20391363852#48.3900101598033386#E#10.1257306666666665
9 1#-4.1747E501606860619#48.398E7935E50155#E#11.1257306666666665
10 1#-4.1747E20391363852#48.3900101598033386#E#12.1257306666666665
11 1#-4.1747E501606860619#48.398E7935E50155#E#13.1257306666666665
12 1#-4.1747E501606860619#48.398E7935E50155#E#14.1257306666666665
13 2#-4.4747E51545131855#48.39134021129603#E#15.1257306666666665
14 1#-4.4769E50169619#48.398E7935E50155#E#16.1257306666666665
15 1#-4.4769E50169619#48.398E7935E50155#E#17.1257306666666665
16 1#-4.4769E50169619#48.398E7935E50155#E#18.1257306666666665
17 1#-4.4769E50169619#48.398E7935E50155#E#19.1257306666666665
18 1#-4.4769E50169619#48.398E7935E50155#E#20.1257306666666665
19 1#-4.4769E50169619#48.398E7935E50155#E#21.1257306666666665
20 1#-4.4769E50169619#48.398E7935E50155#E#22.1257306666666665
21 1#-4.4769E50169619#48.398E7935E50155#E#23.1257306666666665
22 1#-4.4769E50169619#48.398E7935E50155#E#24.1257306666666665
23 1#-4.4769E50169619#48.398E7935E50155#E#25.1257306666666665
24 1#-4.4769E50169619#48.398E7935E50155#E#26.1257306666666665
25 1#-4.4769E50169619#48.398E7935E50155#E#27.1257306666666665
26 1#-4.4769E50169619#48.398E7935E50155#E#28.1257306666666665
27 1#-4.4769E50169619#48.398E7935E50155#E#29.1257306666666665
28 1#-4.4769E50169619#48.398E7935E50155#E#30.1257306666666665
29 1#-4.4769E50169619#48.398E7935E50155#E#31.1257306666666665
30 1#-4.4769E50169619#48.398E7935E50155#E#32.1257306666666665
31 1#-4.4769E50169619#48.398E7935E50155#E#33.1257306666666665
32 1#-4.4769E50169619#48.398E7935E50155#E#34.1257306666666665
33 1#-4.4769E50169619#48.398E7935E50155#E#35.1257306666666665
34 1#-4.4769E50169619#48.398E7935E50155#E#36.1257306666666665
35 1#-4.484288692474365#48.39722589721657#E#37.1257306666666665
36 1#-4.4769E50169619#48.398E7935E50155#E#38.1257306666666665
37 1#-4.49400501794433#48.39525976109966#E#39.1257306666666665
38 1#-4.49400501794433#48.39525976109966#E#40.1257306666666665

```

Fig. 11. Log file Generated from READER 5

VI. CONCLUSION AND FUTURE WORK

Crowd Monitoring and Management in large scale gathering are complex and multifactorial phenomenon which are not so easy to solve. We need automated systems as companions to help solve these issues in the most optimized manner. And save the great havoc- loss or damage to human life.

- The proposed solution tries to achieve both by providing assistance to the people serving and managing the crowd with accurate statistical knowledge and most optimized decision.
- In terms of performance neural networks provide accurate results with proper initialization of thresholds on the nodes, effective and accurate study of conditions constrained on the neural network.
- In terms of security, this method collects all guest entries who visit the village on remote server, so any theft or stranger will be detected.
- Manual entry work will be automated, so time, cost & man power will be reduced. All guest's entries will be forwarded online through internet. Previously entry work was in an offline mode where simple guest ids were given and noted in a register.
- Analysis will be performed on data acquired so as to manage the flow of guests in different zones effectively, reduce the waiting time and serve the visitors in a more convenient manner.

Thus, the proposed solution will be implemented to maximize the user experience during crowd management at large scale events and prevent stampede like phenomenon.

The proposed solution is more inclined towards the social and functional requirements of crowd management strategic approaches- preventing, detecting and avoiding stampede like phenomenon which creates a great havoc – loss of human life in large scale gathering events / areas. Also, it would be an added feature to maximize the user experience during crowd management at large scale events

The system can be scaled to various Crowd Management phenomenon like Stampede which can occur in any densely crowded areas such as Mumbai Local Stations, Pilgrimages like Hajj, Kumbh Mela, Festivals like Ganpati, Durga Puja, Mass Gatherings, Live Shows etc.

We know it's not that easy to manage a crowd of say 1 million, but when it comes to saving lives, some measures need to be taken.

VII. ACKNOWLEDGMENT

We would like to express our heartfelt gratitude towards our mentor Dr. Deven Shah for his unconditional guidance and

support throughout the development of the proposed solution and for always being there for us. It has been a pleasure experience working with him.

Special thanks to Govardhan Eco-Village for providing us the problem definition and allowing us to visit your place for requirement gathering and pursue our innovative idea.

A big thank you to all our friends, family, teachers and well-wishers notably Mrs. Neha Kapadia, Mr. Aaditya Desai, Sudesh, Shashi and Akanksha for providing their invaluable feedback and warm support always.

REFERENCES

- [1] RFID Handbook: Fundamentals And Applications In Contactless Smart Cards, Radio Frequency Identification And Near-Field Communication, Third Edition WILEY Publications
- [2] Designing the Internet of Things: Adrian McEwen, Hakim Cassimally WILEY Publications
- [3] Identification and tracking of persons using RFID-tagged items: United States Patent Application 20020165758 Kind Code A1 Hind, John R. ; et al. November 7, 2002
- [4] <http://www.spychips.com/documents/ATT00075.pdf>
- [5] The Use of RFID for Human Identification: A DRAFT REPORT from DHS Emerging Applications and Technology Subcommittee to the Full Data Privacy and Integrity Advisory Committee Version 1.0
- [6] https://www.dhs.gov/xlibrary/assets/privacy/privacy_advc_om_rpt_rfidi_draft.pdf
Web References:
- [7] https://www.splunk.com/web_assets/pdfs/secure/Splunk_a_nd_MapReduce.pdf
- [8] <https://www.splunk.com/pdfs/solution-guides/splunk-guide-to-operational-intelligence.pdf>
- [9] <http://www.corerfid.com/Files/White%20Papers/033%20Health%20Fact%20Sheet.pdf>
- [10] <https://www.ecovillage.org.in>
- [11] https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol_4/cs1_1/report.html
- [12] https://www.researchgate.net/post/Neural_Network_V_s_Fuzzy_Logic_V_s_Genetic_Algorithm
- [13] <https://www.quora.com/Crowd-Control-Is-it-possible-to-prevent-large-human-stampedes>
- [14] <http://blog.kkaggarwal.com/2013/12/crowd-management-and-stampede-prevention/>
- [15] <http://decibelmanagement.com/crowd-dynamics-how-to-prevent-and-survive-a-human-stampede-or-crowd-crush/>
- [16] <https://spectrum.ieee.org/tech-talk/consumer-electronics/portable-devices/hajj-pilgrims-to-get-electronic-bracelets-in-wake-of-deadly-disaster>
- [17] <http://epmonthly.com/article/human-stampede-an-unexamined-threat/>

Homomorphic Encryption for Cloud Security

Prathamesh Padwal

*Department of Information Technology
Thakur College Of Engineering, and Technology
Mumbai, India
padwalpratham02@gmail.com*

Jaykumar Yadav

*Department of Information Technology
Thakur College Of Engineering, and Technology
Mumbai, India
kumarinject@gmail.com*

Misal Dave

*Department of Information Technology
Thakur College Of Engineering, and Technology
Mumbai, India
misal.dave@gmail.com*

Vijaykumar Yele

*Department of Information Technology
Thakur College Of Engineering, and Technology
Mumbai, India
vijaypyele@gmail.com*

Abstract – Existing data protection mechanisms such as encryption was failed in securing the data from the attackers. It does not verify whether the user was authorized or not. Cloud computing security does not focus on ways of secure the data from unauthorized access. Encryption does not provide much security to our data. This files does not have much security. So, hacker gains access the documents. Twitter incident is one example of a data theft attack in the Cloud. This system will provided with Login authentication for access of cloud for secure authentication and to maintain integrity. If authentication fails i.e. illegitimate user tries to access the cloud then Decoy files will be provided.

Keywords: *Cloud Computing;Encryption*

I. INTRODUCTION

In today's worlds the small as well as big -big organizations are using cloud computing technology to protect their data and to use the cloud resources as and when they need. Cloud is a subscription based service. Cloud computing is a shared pool of resources. The way of use computers and store our personal and business information can arises new data security challenges. Encryption mechanisms not protect the data in the cloud from unauthorized access. The traditional database systems are usually deployed in closed environment where user can

access the system only through a restricted network or internet.

With the fast growth of W.W.W user can access virtually any database for which they have proper access right from anywhere in the world. By registering into cloud the users are ready to get the resources from cloud providers and the organization can access their data from anywhere and at any time when they need. But this comfortness comes with certain type of risk like security and privacy.

To overcome this problem by using a new technique called as fog computing. Fog computing provides security in cloud environment in a greater extend to get the benefit of this technique a user need to get registered with the fog. Once the user is ready by filling up the sign up form he will get the message or

email that he is ready to take the services from fog computing.

II. BACKGROUND

Cloud has become essential part of our life that are almost reliable on cloud for data storage. Due to this reliability there are many threats for cloud and the researches which are going on is basically concentrated on the encryption. Encryption is insufficient to prevent the cloud for the data theft attack. Main problem on cloud in insider data theft and the encryption is no capable to prevent the cloud. For this reason proposing whole new approach which

is decoy information technology in check data access in the cloud and observe abnormal information access patterns and when the abnormal patterns are detected that perform series of security questions and flood the intruder's database with decoy information and dummy files and thus can prevent cloud's data. Secured cloud computing with decoy documents extends the Cloud Computing approach to the boundary of the network, thus enabling a new kind of applications and services. Identifying masquerades has become very difficult. Many planned approaches rely on encryption by auditing a variety of sources which are not able. Few years back there was a Twitter incident which is a best example for data theft attack from the cloud service provider. This incident exposed the security problems in cloud computing as it could make the customers of Twitter to lose their sensitive data and documents.

III. MOTIVATION

While this particular attack was launched by an outsider, stealing a customer's admin passwords is much easier if perpetrated by a malicious insider. Rocha and Correia outline how easy passwords may be stolen by a malicious insider of the Cloud service provider. The authors also demonstrated how Cloud customers' private keys might be stolen, and how their confidential data might be extracted from a hard disk. After stealing a customer's password and private key, the malicious insider get access to all customer data, while the customer has no means of detecting this unauthorized access. Much research in Cloud computing security has focused on ways of preventing unauthorized and illegitimate access to data by developing sophisticated access control and encryption mechanisms. However these mechanisms have not been able to prevent data compromise. Van Dijk and Juels have shown that fully homomorphic encryption, often acclaimed as the solution to such threats, is not a sufficient data protection mechanism when used alone.

IV. SCOPE

In our future work, this security system as that have explained is applicable for single cloud ownership system. If the cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud. Cloud computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud.

V. PROBLEM DEFINITION

Data protection mechanisms such as encryption have failed in preventing data theft attacks, especially those perpetrated by an insider to cloud provider. Different Access Control Methodologies are used and single authentication system is used. This will provide multiple authentication system. Much research in cloud computing security has focused on ways of preventing unauthorized and illegitimate access to data by developing sophisticated access control and encryption mechanism.

A. Phase 1: (Planning, Analysis, Designing and Implementation)

Analysis: Getting clear idea of the project title and doing research on it will get our definition and after that then it will first create literature survey of the project and do the whole documentation

Planning: After analysis it will first study about it and do some research on it for our better understanding of the project and getting a rough idea about the project about what would be our problem definition.

Designing: Then it will construct the design of the project and according to that it will list down all the requirements needed for construction of the prototype of our project.

Implementation: After acquiring the requirements it will proceed to the construction of the application by using HTML, Java, MySQL and other technologies.

B. Phase 2: (Testing and Deployment)

Testing: After the prototype is ready it will test the Login system and encryption of data and it will check if it supports mechanism or not. If not it will solve the issues regarding to it and try other various test cases.

Deployment: After, complete integration and testing of project real time running and operation of the system will be done. Cloud user is expected to validate his/her authentication and transmit & receive the encrypted data securely.

VI. METHODOLOGY

Numerous proposals for cloud-based services describe methods to store documents, files, and media in a remote service that may be accessed wherever a user may connect to the Internet. A particularly vexing problem before such services are broadly accepted concerns guarantees for

securing a user's data in a manner where that guarantees only the user and no one else can gain access to that data. The problem of providing security of confidential information remains a core security problem that, to date has not provided the levels of assurance most people desire. Many proposals have been made to secure remote data in the Cloud using encryption and standard access controls. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including insider attacks, misconfigured services, faulty implementations, buggy code, and the creative construction of effective and sophisticated attacks not envisioned by the implementers of security procedures. Building a trustworthy cloud computing environment is not enough, because accidents continue to happen, and when they do, and information gets lost, there is no way to get it back. One needs to prepare for such accidents. The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack.

VII. FEASIBILITY STUDY

Technical Feasibility: The system works on network layer of OSI model. The system implemented using Java, C, Html, etc. as programming and scripting languages. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms.

Economic Feasibility: The cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud. Cloud computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud. Even after the development, the organization will not be in a condition to invest more in the organization .Therefore, the system is economically feasible.

Legal Feasibility- The project is develop under legal license of windows, Linux and other software's required. The data will not be used for some illegal purpose and will be encrypted using algorithms. Data will not be misused.

Operational Feasibility- An overview of the process portrayed that the proposed Homomorphic encryption in cloud would be beneficial in faster retrieval of data,

secure data, and trusted data, integration of data and encourage accountability.

Social Feasibility- Easy to make allocation and accurately with this help student/people will able to make and submit the project to the industry/university on time and help them to protect their cloud data.

VIII. FLOW DIAGRAM

Figure 1 is a flowchart diagram. Flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem.

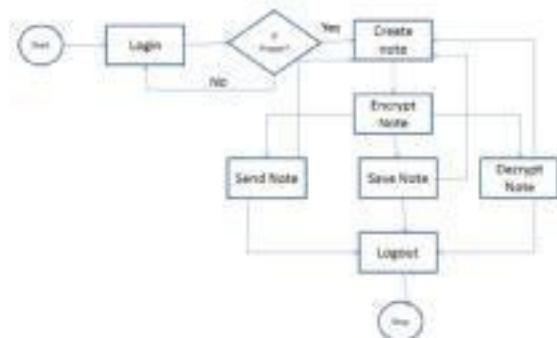


Fig 1: Flow Chart Diagram for Homomorphic Encryption for Cloud Security.

IX. USE CASE DIAGRAM

Figure 2 is a use case diagram of the system. Use case diagrams are usually referred to as behavior diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors).

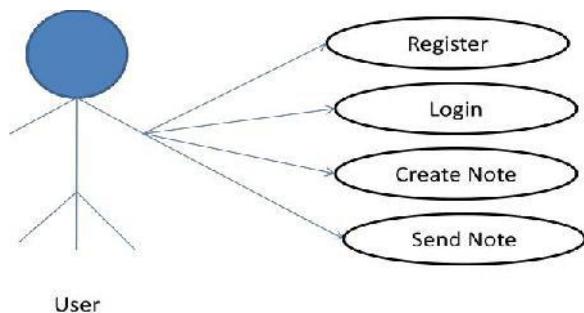


Fig.2: Use Case Diagram for Homomorphism Encryption for Cloud Security.

X. LITERATURE SURVEY

Data protection mechanisms such as encryption have failed in preventing data theft attacks, especially those perpetrated by an insider to cloud provider. According to Ref [1], only SSL/TLS protocols are used for data transmission. This will provide Login Authentication System and Decoy file system if authentication fails. According to Ref [2],[3],[4]; different encryption algorithms are used such as RSA, Paillier, AES, etc. Due to inefficiency and slow working of pallier and RSA using AES algorithm for 256 bits in our project. According to Ref [5], different Access Control Methodologies are used and single authentication system is used. This will provide multiple authentication system. Much research in cloud computing security has focused on ways of preventing unauthorized and illegitimate access to data by developing sophisticated access control and encryption mechanism.

XI. EXPECTED RESULT

a. Outputs

Encryption: On creating a note it encrypts the note using AES encryption method. Which is later been used to send to any of the registered users, where in it is sent in encrypted format.

Decryption: On receiving the note user requires to decrypt the note with appropriate key which is sent on email. Hence if user puts correct key the user can see decrypted.

b. Outcomes:

Phase 1:

Analysis: Creating the synopsis, literature survey and feasibility study for proposed system.

Planning: Proposing statement of work, scope definition for planning the prototype from the problem definition.

Designing: Creating a visually appealing frontend for the application system.

Implementation:

Various modules will be created like:

1. Login System
2. Decoy File System
3. AES algorithm of key length 256 bits is used for encryption

Phase 2:

Testing: The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning and perform integration testing system testing.

Deployment:

1. Real time deployment of project
2. Cloud user will validate his/her authentication and transmit & receive the encrypted data securely.

XII. FUTURE SCOPE

In our future work, this security system as it have explained is applicable only for single cloud ownership system. If the cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud architecture. Cloud computing is the future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud.

A lot of effort will be put in return to provision the appropriate security to make business on cloud environments. Future researches should aim to provide new architectures and techniques to harden the different resources shared between tenants. The hypervisor is the most critical component of virtualized environments. If compromised, the host and guest OSs could potentially be compromised too.

XIII. CONCLUSION

In this position paper, it presents a novel approach to securing personal and business data in the Cloud. It proposes monitoring data access patterns by profiling user behavior to determine if and when a malicious insider illegitimately accesses someone's documents in a Cloud service.

Decoy documents stored in the Cloud alongside the user's real data also serve as sensors to detect illegitimate access. Once unauthorized data access or exposure is suspected, and later verified, with challenge questions for instance, it inundates the malicious insider with bogus information in order to dilute the user's real data.

Such preventive attacks that rely on disinformation technology, could provide unprecedented levels of security in the Cloud and in social networks. The basic idea is that it can limit the damage of stolen data if it decreases the value of that stolen information to the attacker.

REFERENCES

- [1.] Rabi Prasad Padh, Manas Ranjan Patra, Suresh Chandra Satapathy "Cloud Computing:Security Issues and Research Challenges", 2011.
- [2.] MahaTebba, Said El Haji, "Secure Cloud Computing through Homomorphic Encryption", 2013.
- [3.] Iram Ahmad, Archana Khandekar, "Homomorphic Encryption Method Applied to Cloud Computing", 2014.
- [4.] Vishal R. Pancholi, Dr. Badeshi P. Patel, "Enhancement of Cloud Computing Security with Secure Data Storage using AES", 2016.
- [5.] Jayshree Khandagale, Deepak Fodse, Poonam Sul, Prita Patil, "FOG Computing", 2016.
- [6.] Tim Mather, Subra Kumaraswamy, Shahed Latif, "Cloud Security and Privacy:An enterprise perspective on risk and compliance", 2009.
- [7.] Ihsan Jabbar, Saad Najim, "Fully Homomorphic Encryption to Secure Cloud Computing", 2016.

Authentication and Access Control

Akanksha Jain

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
akankshajain1502@gmail.com*

Pooja Jain

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
poojajain112@gmail.com*

Shivani Rege

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
regeshivani23@gmail.com*

Sudhir Dhekane

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
sudhirdhekane1987@gmail.com*

Abstract: -

Authentication and access control of computer applications is today's need in the technical education institutes. It is common practice that students performing various subject labs where they are utilizing computer applications actually required to perform task rarely, rather they are busy utilizing the application irreverent w.r.t lab task. So a system is needed to be designed which takes care of the authentication and access control. This can be achieved by allowing authenticated users to use the computer labs and actual application and blocking unnecessary application to valid users. After the process of authentication is done it is necessary to give access rights to the authenticated users according to their lab requirement.

In this work, the system designed would be tested on the college labs. This system will have a login page which would test the authenticity of the user and then allow access to the user if the credentials are valid. The students would be allowed to access applications that are necessary for that particular lab session. This access of applications would be dependent on the timetable of the labs.

The system would be implemented in college labs. There would be three type of login i.e. teacher login, student login, guest login. Teachers would be allowed to access all applications as well as change the timetable fed in server computer. The student would only be able to access applications according to the timetable of the labs. The applications required for that particular lab would only be accessible. This would ensure that all the students are performing the intended task.

Keywords: Authentication, Access control.

I. INTRODUCTION

Access Control and Authentication is basically allowing access rights to particular users after they are validated with the user after they are validated with the user identity database. Nowadays, Security of data is of utmost important as the data present in the computer can be misused by unauthorized intrusions. An intruder can modify and change the program source codes and can also use your pictures or email accounts to create derogatory content such as, fake misleading and offensive social accounts. Thus, access to data, accounts etc should be given to only

authenticated users. After authentication process, the access to resources should be limited too. Giving access to all resources to each authenticated user would lead to misuse and harm to data. Therefore access rights play an important role in any company, educational institutions etc. which classify the authenticated users and grant access to resources accordingly.

II. RELATED WORK

According to the author the system is used to give all access to admin and then it also defined the user groups that can be characterized by security levels, roles or any partition of users. A user can access any data that has lower or equal security level and i.e. accessible by the roles to which the user is assigned. All type of access mode are controlled in this phase [3].

In this paper, the Author has specified that the system is having different access right to different level of users. It specifies the need of access control in particular task. It also specifies that usage of different access control which help the user to give different right to different user module[4]

Here the author has written that system specifies the need of new access control as Content Based Access Control where each user is allowed by RBAC rule to access a large set of data objects. In CBAC, each user is allowed by an MLS or RBAC rule to access a large set of data objects, while the CBAC rule imposes an additional layer of restrictions that the user could only access "a subset" of the designated records[5].

III. PROPOSED SYSTEM

Access to applications should be given to authenticated students and professors. After authentication is done, only the applications that are required for the practical could be accessed and others should be blocked .This access and blocking of applications would be according to the timetable of that particular lab schedule, which would be fed in the server computer. This server machine would accordingly adjust the access rights of the computers connected via LAN setup in a particular lab.

IV. METHODOLOGY

The users will be authenticated by validating details that are stored in the database .After authentication work is done then access is granted according to the user role modules

i) Student Module:-The student would be given access to applications that are required for implementation of the practical session of a particular subject. Other applications would be blocked. If the student requires a certain application then the teacher module would grant access accordingly

ii) Teacher Module:- Teacher has all rights to access all the Software application available on the computer. If in case any student wants another software application other than the allotted application, teacher would grant access by modifying the server

iii) Guest Module:- This module would be granted access to all the applications but the session would expire in 30 minutes and again the user would be redirected to the login page and if again access required ,should be requested to teacher module.

The below figure 1 is showing workflow of proposed system.

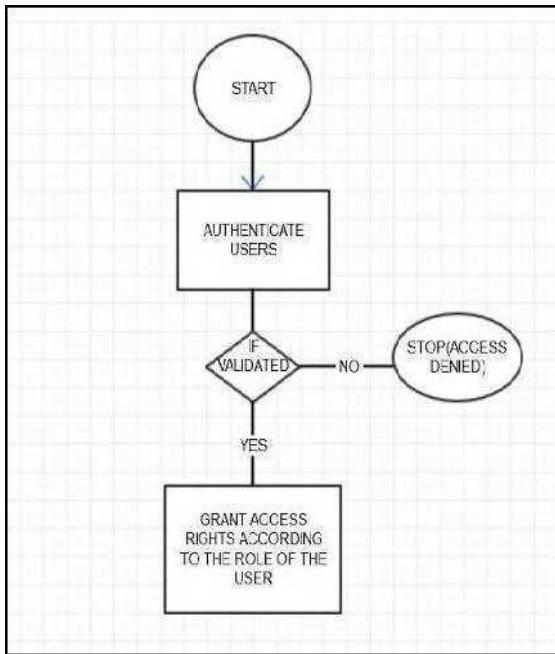


Figure 1: -Flow Chart for authentication and access control system

The coding language that will be used to implement this work is java (session handling).

V. EXPECTED RESULTS

Student should get access of all needy application with respect to the lab which he is going to perform unnecessary applications must be denied.Teacher should assign necessary other (except assigned) application access to student on his demand if it's relevant.

Teacher should continue the authentication and access of all application to guest user as per his demand and necessary once limited session period expires.

No application or computer usage to any invalid i.e. non authorization user during lab session. Let us see the execution environment in background with different cases as follow.

Table I: JAVA LAB (Case 1)

SUBJECT/ ACCESS	JAVA	MYSQL	MATLAB	VB
STUDENT	yes	no	no	no
TEACHER	yes	yes	yes	yes
GUEST	yes	yes	yes	yes

This access control table lists the access rights of different modules. Here the example taken is the java lab and hence the applications needed to execute java code are accessible to students, others are blocked. Next access table gives the access rights of database lab.

Table II: DATABASE LAB (Case 2)

SUBJECT/ ACCESS	JAVA	MYSQL	MATLAB	VB
STUDENT	no	yes	no	no
TEACHER	yes	yes	yes	yes
GUEST	yes	yes	yes	yes

*Yes denotes access granted

*No-access denied

VI. CONCLUSION

The system will be providing following features:

Authenticated users and provide access application. Allocate access rights to different users according to the role i.e. student and teacher would be given different access rights. The access rights granted to students would be set according to the lab subject.

If any change in access rights is required that could be modified & allocated by the teacher module.

This system will avoid the use of unnecessary applications by the students while performing his/her task in lab session i.e. Student cannot use any other software other than the application allocated to him.

REFERENCES

- [1] G.A Champine "Project Athena as a distributed computer system", IEEE computer, vol. 23, no. 9, pp. 40-51, Sept. 1990.
- [2] R.M Needham "Using encryption for authentication in large networks of computers".
- [3] Min-A Jeong , Jung-Ja Kim , Yonggwan Won "database security system as a Parallel and Distributed Computing, Applications and Technologies" IEEE computer, 20 October 2003.
- [4] Qing-hai Bai , "Cross Strait Quad-Regional Radio Science and Wireless Technology Conference " IEEE computer, Harbin, China, 10 October 2011.
- [5] Wenrong Zeng,Yuhao Yang, andBo Luo " Big Data (Big Data)," 2014 IEEE International Conference, Washington, DC, USA, 08 January 2015
- [6] Jeff Voas, Piero P. Bonissone, Yoshihiro Ohba "IEEE Consumer Communications & Networking Conference" IEEE Communications Society , Las Vegas, USA, 12-15 January 2018.
- [7] Mohamed Fathy , Marianne Azer , Moustapha Bahgat , Ayman Yehia "Security access control research trends" Wireless and Mobile Networking Conference(WMNC), 27 June 2013.

Web Server Log Analysis

Sangeeta Vhatkar

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India
vhatkarsangeeta@gmail.com*

Vishal Patel

*Department Of Information Technology
Thakur college of Engineering and Technology
Mumbai, India
p.vishal4996@gmail.com*

Shubham More

*Department of Information Technology
Thakur college of Engineering and Technology
Mumbai, India
moreshubham333@gmail.com*

Harsh Rawal

*Department Of Information Technology
Thakur College of Engineering and Technology
Mumbai, India
rawalharsh48@gmail.com*

Abstract— Web server log analysis is a novel and peculiar field constantly formed and changed by the convergence of various Web technologies. Due to its interdisciplinary character, the diversity of issues it addresses, and the variety and number of Web applications, it is the subject of many distinctive and diverse research methodologies.[1]About Log Files - Current software application often produce (or can be configured to produce) some auxiliary text files known as log files. The main problem with log files being the vast amount of unprocessed data which needs to be processed for analysis. Such files are used during various stages of software development, mainly for debugging and profiling purposes Use of log files helps testing by making debugging easier. It allows to follow the logic of the program, at high level, without having to run it in debug mode. Nowadays, log files are commonly used also at customer's installations for the purpose of permanent software monitoring and/or fine-tuning. Log files became a standard part of large application and are essential in operating systems, computer networks and distributed systems.

Keywords:Web Server; Log Files;Computer Network

I. INTRODUCTION

Web-based applications have become prevalent in the ubiquitously-connected world. The recent increase in demand for huge data storage and high processing speed has led to the era of cloud computing, yet, most cloud-based systems and applications are accessed through the Internet web. Web security therefore has remained as top priority for Internet and Cloud-Services Providers (ISP/CSP). In order to improve ISP/CSP user confidence and to protect web-based systems and applications, strong, effective security mechanisms must be deployed. Many architecture designs have been proposed and many cryptographic algorithms implemented; yet intruders continue to gain access to web based application, to stealconfidential information, and to make unwanted modifications – including recent intrusions made to Target, the

Home Depot, BlueCross Insurance, and many hospital and government informationsystems. Two major techniques have been used in log analysis: pattern matching and machine learning. While the pattern matching method may work dynamically, only known patterns can be recognized, yet new types of injections may be created when only small changes are made to existing patterns Machine learning also has its limitation, since classification in machine learning algorithms works with probabilities, it may not be able to correctly classify SQL injections that combine groups of words each was classified with high probability as non-SQL injection. Most existing log analysis methods for SQL injection detection are based on either patternmatching or machine learning.Log files containing approximately 20000 Log entries or more are taken into account as the example dataset in our project. Entries such as the IP addresses, category of Operating system, Origin of IP address and the threat concerned with the related IP are processed. Log files are often the only way how to identify and locate an error in software, because log file analysis is not affected by any time-based issues known as probe effect. This is the opposite to an analysis of a running program, when the analytical process can interfere with time-critical or resource-critical conditions within the analyzed program. Log files are often very large and can have complex structure. Although the process of generating log files is quite elementary and unexacting, log file analysis could be a tremendous task that requires enormous computational resources, long time and enlightened procedures. This often leads to a common situation, when log files are continuously generated and occupy valuable space on storage devices, but nobody uses them and utilizes enclosed information. The proposed work till date has substantial outcomes such as the log file provided as the input is mined for the appropriate data and kept ready for further use. Therefore, with the growing speed of data in the web, a framework is needed to process and analyze the data for vulnerabilities. Here we elucidate the analysis of log files using pattern matching framework which incorporates the major preprocessing task and session identification algorithm to handle vast amount of log data. From the results it is concluded that processing a huge file in distributed fashion

II. SCOPE

Proactive monitoring - Move from reactive to proactive real-time log monitoring and view app performance, system behavior, and unusual activity across the stack. [2] Monitor key resources and metrics, and eliminate small issues before they turn into big problems.

Troubleshooting - Trace issues down to their root cause by analyzing them in the context of the entire stack. See how your components interact, identify correlations, and share findings with experts across team boundaries to resolve problems fast.

Data analysis and optimization - Analyze and visualize your data to answer key questions, track SLA compliance, and identify anomalies. Our project automatically recognizes common log formats and gives you a structured summary of all your parsed logs. Team collaboration and integration - Building and running complex systems requires tight coordination between development, operations, and product.

III. MOTIVATION

The web pages are one of the most important advertisement tools in international area for foundation, institutions, etc. Therefore, the suitability to W3C standards, content and design of web pages are very important for system administrator and Web designer. Especially, the number of visitors is acceptable as the measure of the affectivity and quality for a commercial foundation or a university. So web analyzers have to analysis their server log files to determine systems error. Instead of tracking the behavior of overall users (interested or not interested) in order to redesign the web site to support users. The data mining techniques like Association, Clustering, and Classification can be applied only on to the group of interested regular users to find frequently accessed patterns which results in less time consumption and less memory utilization with high accuracy and performance.

IV. PROBLEM DEFINITION (PHASE WISE)

Web Server Log analysis system is a kind of application which will help to detect and avoid the attacks which might happen on the server by using the log files of the server itself, this log files can be used to accurately give us an idea so as to where and how the attack is been initiated and is taking place.

A. PHASE 1: (Planning, Analysis, Design,Coding)

Planning: This involves how to analyze the log files from the obtained data. To use the available database to detect what kind of attack takes place on the server at the present moment.

Analysis: Critical study, analysis and review of feasibility for proposed solution. If a genuine user is an attacker (unauthorized person) a warning will be issued to the server manager (authorized person) then the preferred action can be taken by him.

Coding: Open source algorithms can be implemented using various coding techniques including java, PHP etc.

Design: The application will get an access to the log files of the server which will then indicate what kind of attack is been executed and what action can be taken.

PHASE 2: (Integration, Testing,Deployment)

Integration – Integrating of various proposed modules such as Log detection module whether it detects attacks from various destinations.

Testing – Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.

Deployment – Give the completed prototype to available server for security purpose in the active servers.

V. METHODOLOGY

We present the application of the proposed methodology for analysing of the web log files. In this study, we have developed an expert system to assist the web designer and web administrator to improve their website by determining occurred link connections in the website. Firstly, we have obtained access log files which are recorded in web server of the First University. The obtained log files were analysed by proposed web usage mining methodology in SAS software. We present an overview of the tasks for each step and discuss the challenges involved. The architecture consist of three main tasks for performing web usage mining: pre-processing, pattern discovery and pattern analysis.

An important task in web usage mining application is the creation of a suitable pre-processed usage data set. This process is usually complex and critical to the successful extraction of useful from the log files in web usage mining. Purpose of the pre-processing is to offer a structural, reliable and integrated data source for pattern

Data cleaning is the first step performed in the pre-processing of web usage mining. In the raw logs, not all the log entries are valid for pattern analysis. We only want to keep the entries that carry relevant information. Therefore, the data cleaning step is used to eliminate the irrelevant entries from the access log files.

Transaction identification is to create meaningful clusters of references for each user. Each user session is considered either as a single transaction consisting of many page references or a set of many single-page reference transaction.

A session can be described as the group of activities performed by a user from the moment he entered the website to the moment he left it. Therefore, session identification is the process of segmenting the access log of each user into individual access sessions.

VI. FEASIBILITY STUDY

Technical Feasibility -Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

Economic Feasibility -As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

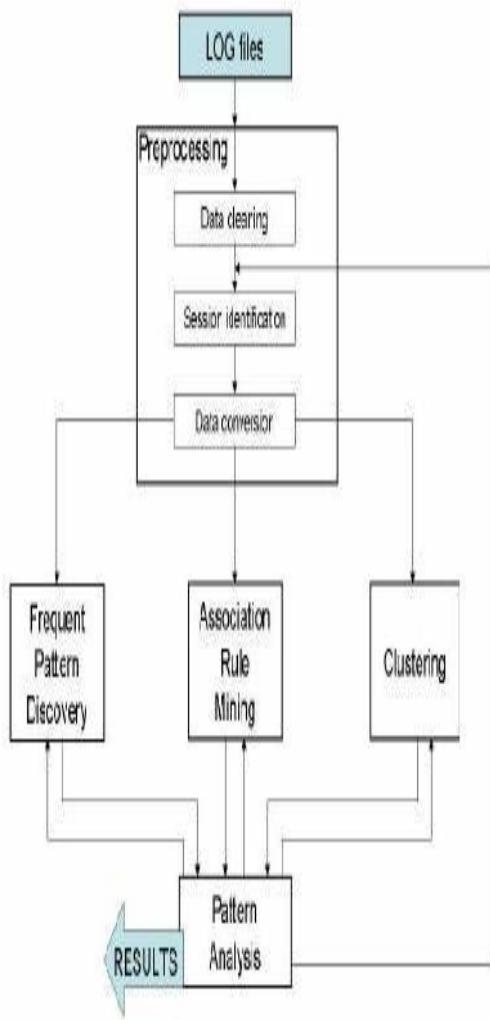
Legal Feasibility- Legal issues can affect a system's acceptance by bottom users, its performance, or the decisions on whether to use it in the first place so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any log analysing system.

Operational Feasibility -It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The

proposed system is acceptable to users. So the proposed system is operationally feasible.

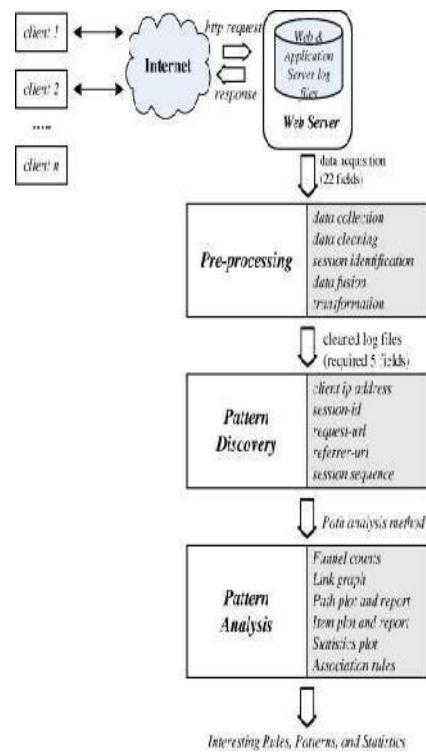
Social Feasibility -The acceptability of a log analysis system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness.

VII. FLOW DIAGRAM



filling out a form regarding their interests, for example when registering on the web portal. The clustering of the users can be accomplished based on the forms. On the other hand, the clustering can be made based on the information gained from the log data collected during the user was navigating through the portal. Different types of user data can be collected using these methods, for example (I) characteristics of the user (age, gender, etc.), (ii) preferences and interests of the user, (iii) user's behavior pattern. The aim of clustering web pages is to have groups of pages that have similar content. This information can be useful for search engines or for applications that create dynamic index pages. The last step of the whole web usage mining process is to analyze the patterns found during the pattern discovery step. The irrelevant patterns have to be filtered out, and the resulted patterns or clusters have to be validated. Some visualization techniques can help this process for the user.

VIII. BLOCK DIAGRAM



A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to document, study, and plan, improve and communicate often complex processes in clear, easy-to-understand diagrams. Figure 1 shows the flow chart for Web server log analysis using patternanalysis.In the preprocessing phase the data have to be collected from the different places it is stored (client side, server side, proxy servers). After identifying the users, the click-streams of each user has to be split into sessions. In general, the timeout for determining session is set to 30 minute.The pattern discovery phase means applying data mining techniques on the preprocessed log data. [3]It can be frequent pattern mining, association rule mining or clustering. In this paper we are dealing only with the task of clustering web usage log. In web usage mining there are two types of clusters to be discovered: usage clusters and page clusters.

The aim of clustering users is to establish groups of users having similar browsing behavior. The users can be clustered based on several information. In the one hand, the user can be requested

- **Data collection and pre-processing:**

An important task in web usage mining application is the creation of a suitable pre-processed usage data set. This process is usually complex and critical to the successful extraction of useful from the log files in web usage mining. Purpose of the pre-processing is to offer a structural, reliable and integrated data source for pattern.

- **Data cleaning:**

Data cleaning is the first step performed in the pre-processing of web usage mining. In the raw logs, not all the log entries are valid for pattern analysis. We only want to keep the entries that carry relevant information. Therefore, the data cleaning step is used to eliminate the irrelevant entries from the access log files.

- **Transaction identification:**

The aim of transaction identification is to create meaningful clusters of references for each user. Cooley et al. (1999a) propose a general model for transaction identification. In their model, each user session is considered either as a single transaction consisting of many page references or a set of many single-page reference transaction.

- **Session identification:**

A session can be described as the group of activities performed by a user from the moment he entered the website to the moment he left it. Therefore, session identification is the process of segmenting the access log of each user into individual access sessions.

- **Pattern discovery:**

In order to extract patterns of usage from web log files are used data mining techniques for web usage mining. [4]Pattern discovery is the key process of the web mining, which includes the algorithms and techniques from several research areas, such as data mining, machine learning, statistics and pattern recognition. The techniques such as statistical analysis, association rules, clustering, classification, sequential pattern and dependency modelling are used to discover rules and patterns

IX. EXPECTED RESULT

This application will provide the security to any server using the log files associated with the server. This can accurately detect the attack

that takes place also the details related to the attacker such as the ip of attacker, type of attack and other attacker attributes.

We have to plan that stored data i.e. attack log type in database (Trained database) needs to match with the database which also contains the same log data as trained database i.e., test database. If the trained database matched with the test database following action will be performed first, if user is authorized person access will be granted and secondly if user is an unauthorized person then threat will be detected.

Design by using Log Analysis Algorithm all the connection followed by using Database to detect what kind of threat or attack can be executed, and Warning is associated to Dashboard output and it is used for indication purpose and after all the interconnection is done the server is used for keeping it protected from threats.

Various modules for coding will be created like:

1. Application code having log analysis algorithm developed over proper code.
2. Dashboard output for warning the server manager.
3. Possible action that can be taken by the server manager. For integration, constructing the various modules of the proposed solution to integrate them into a prototype. The integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stress testing.

The benefits of Analysis on after deploying the proposed prototype include:

1. Proper and efficient way for securing server.
2. This system can be used in various multiple servers.
3. The cost is efficient.
4. The owner has full control on his/her server.
5. Attacker can be identified easily.
6. It is feasible for each and every car.

X. FUTURE SCOPE

Instead of tracking the behavior of overall users (interested or not interested) in order to redesign the web site to support users. The data mining techniques like Association, Clustering, and Classification can be applied only on to the group of interested regular users to find frequently accessed patterns which results in less time consumption and less memory utilization with high accuracy and performance.

XI. CONCLUSION

The normal process of log information retrieval includes store process and then visualize, however, the proposed platform does preprocess-store and directly visualize, thereby decreasing overall latency. Making use of the proposed platform, real time data analytics can be made possible on large data sets, thus facilitating prompt insights into the data. The platform can take in any type of log files, which gives it a generic capability to analyze different logs simultaneously. The platform can also be extended as a base module to serve other applications for decision making on the basis of real-time analytics. The clusters, which are created on the fly depending on the patterns identified, help shortening the search time. Thus the searching is done in a single cluster instead of the whole data store. This tremendously reduces the lag between request and response. As an added functionality the platform can be extended to serve any type

of data other than logs, if the data matches the regular expression in system. Thus, the proposed solution will be implemented.

REFERENCES

- [1] Jan Valdman, Log File Analysis, University of West Bohemia in Pilsen, July 2001.
- [2]<https://www.capterra.ch/directory/30856/log-management-software>
- [3] <http://studyres.com/doc/4204560/a-result-evolution-approach-for-web-usage-mining-using-fu...>
- [4]https://mafiadoc.com/extraction-of-interesting-patterns-through-association-rule-resul-da_59c712131723ddf68040fa71.html
- [5] <https://www.lucidchart.com/pages/what-is-a-flowchart-tutorial>
- [6] SanthoshPininti, Sindh Usha and Teng-Sheng Moh "Detecting Web Attacks Using Multi-Stage Log Analysis."
- [7] K. R. Suneetha "Identifying User Behavior by Analyzing Web Server Access Log File"
- [8] AmrutaAmbre, NarendraShekokar "Insider threat Detection using Log analysis and Event Correlation."
- [9] Kurt Thearling "An Introduction to Data Mining", Computer Society of India Communications, Oct 2006.

Cart Free Shopping

Nishant Gandhi

*Dept. of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India
nishantgandhi18@gmail.com*

Ajinkya Padave

*Dept. of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India
a2padave@gmail.com*

Kajal Gogri

*Dept. of Information Technology
Thakur College of Engineering &
Technology
Mumbai, India
gogrikajal@gmail.com*

Rahul Neve

*Dept. of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
rahulneve@gmail.com*

Abstract

Nowadays shopping is either done online or the customer visiting the supermarkets or mall personally. There are many threats occurring in an online shopping so people hesitate to shop online and prefer going to the markets or malls but there they also face many problems like unavailability of the products or waiting in long queues etc. People sometimes face connectivity issues or credit card fraud, security breaches like the phishing and pharming attack, leak of user data, inconvenience of the products delivered, and time required for the delivery. In this paper we have tried to integrate both the methods and arrive to a solution where quality of the product will be maintained along with the time factor being controlled and security achieved at the higher level.

Parameters that the customer consider while shopping are quality of the product, availability of the product, time being saved and connectivity with the server, no waiting for loading of webpages and secure payment systems. We have tried to cover all these parameters and provide an effective solution to them by making the android app which will almost solve all the parameters that the customer consider while shopping. The time factor is mainly achieved by making use of QR code instead of barcode which decreases the scan time and the app functions smoothly for all android smartphones.

Using QR codes and data mining knowledge this project has help to scan the products much faster than the bar codes. The number of data characters has increased as that from limit of 20 characters. The authentication of user helps to maintain the user data security and help them to view their previous shopping list that is created when they scan proceed for checkout, this list can be viewed later on when they return for shopping. This features of online shopping in traditional shopping has given us the proper blend of technology and the traditional methods to provide the best for the customers.

Keywords-phishing, pharming, security, QR code , barcode

Traditional shopping is a tedious and time-consuming job. Although the growing trend of online shopping has reduced some load, there is still some difference in actually going to

shops, and hand picking products to get the feel of their quality and features, that cannot be experienced online. Customers also feel wary to carry out online purchases due to fear of less secure transaction process that may lead to hacking of user's sensitive data, insecurity of credit/debit cards, unreliability or breach of privacy. The project aims at removing flaws of both kinds of shopping, and bridge the gap between physical and virtual world. In traditional shopping, the customer has to wait in long queues at the cash counter. The cashier scans barcode for every individual product and then generates the bill. This consumes lot of time and energy of both the shopper as well as cashier.

II. PROPOSED WORK

A. Overview

Till now we have been using the traditional way of shopping for a long time but it still has some flaws that can be improved with technology. The customers will be able to save much time while shopping without carrying the products and the weight and the efforts for pushing a cart is eliminated from the shopping experience. The user will get the bill as he/she is shopping so that the user can know how much he/she is spending and be helped staying in budget. Another flaw the traditional system has is the barcodes on the products. Barcodes are very inefficient as the store less data and are sometimes difficult to scan using the laser reader. Barcodes causes the most inefficiency during the billing which causes more time consumption. So adding a QR code in front of products which stores a lot more data and is very easy to read for the scanner from any angle helps in saving a lot of time and effort.

I. INTRODUCTION

B. DATABASE

The shop's database is designed using MYSQL workbench. It provides an interface with the help of which any database can be easily designed. The shop's database consists of six table-

- Inventory system- It provides information about the availability of the items, their unique id, product id etc.
- Item table- It provides detailed information of each item from its manufacturing date, price, weight, etc.
- Shopper details- The customer's information will be stored in this table including his address and phone number that will be used at the time of online payment.
- Shopper session- This table will have information about the time when customer has log in after scanning the shop's barcode till the successful payment completion.
- Store details- This table will have detail information about the shop's name, its branch and unique id

(barcode) that will be retrieved at the time of scanning of the shop's barcode.

- Final order table- This table maintains customer information about his purchases, total cost, session id and all those information that is required to generate a final bill.

III. LITERATURE SURVEY

After the careful study of the articles and references the following key findings were found :

1. A customer 3.97(mean), either on online or offline shopping Takes :
 ‘Quality’, ‘Genuineness’ of products and, ‘Customer Services’ into consideration . The satisfaction of customers was found out to be more while doing online shopping, around 0.733(mean) [4]. In another study using same parameters this mean value was found to increase for level of satisfaction for online shopping 2.895(mean). And the Quality factor was seen to decrease to 1.747(mean) [5].
2. The security issues while dealing online were found to be high in two different studies which indicates the flaws that can lead to loss in business and leak of private data of customers such as personal and bank details. [4][5].
3. The previous system were found to be using Barcode system which store less than 20 linear character. Can't put description in those 20 character.[7]
4. Automated analysis needs to tackle implicit and explicit flows to provide the user with guarantees about the security of their private data. Since permissions in android are very

IV. PROBLEM DEFINITION

Phase-1:

After identifying the gap in [1] the use of barcode in the project is substituted by the QR code which will be added to the products. ThisQR code scanner will be implemented by using the ZXing library.

Phase-2:

After identifying the gap in [2] further the security of the users will be maintained through form activity in android and then authenticated at the backend of the project on laravel, a framework of php. Here the authentication of the user will be done using JWT (JSON Web TOKEN) authentication provided by laravel.

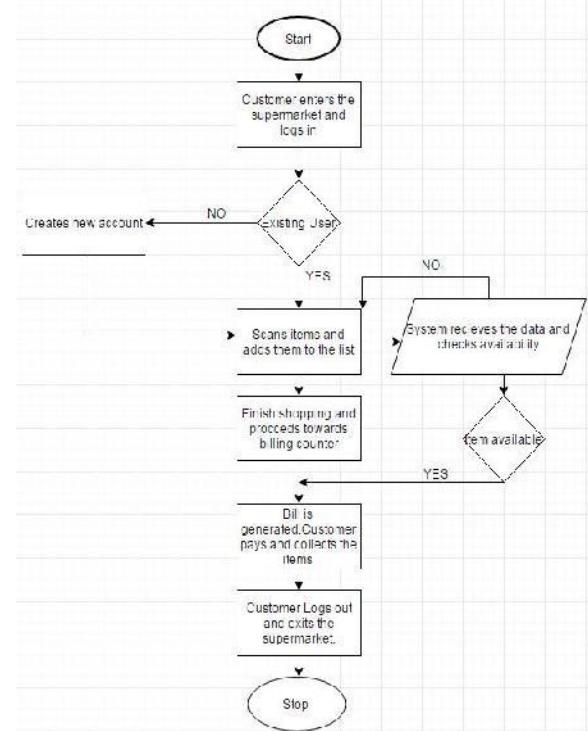
Phase-3:

After identifying the gap in [3] the project requires development of basic forms necessary for cart free shopping & prior to shopping inputs from end user will be collected to predict the pattern of their shopping. On basis of this customers will be guided with the products.

Phase-4:

In the proposed system the Genuineness & quality of the product will be exhibited by providing reviews and ratings by critics and other users.

V. METHODOLOGY USED



VI.TECHNOLOGIES USED

I. Using android studio and the available ZXing library that used to implement a Scanner activity for android application we will develop a dedicated QR code scanner which will be used by the user to scan the QR code of the products and add products to the final list for completion of shopping procedure.

II. The backend of the application is handled using the laravel framework of php where the registration, authentication and verification can be done using the JWT Authentication package provided by laravel. Also handling of users and their data along with database table creation and management is handled using laravel.

III. Adding and comparing the products in the database will be done using basic laravel comparing system of laravel. A detailed list of the users shopping pattern will be maintained and using predicting algorithms and data mining we will provide the users suggestions and the ratings for the products to buy. Also notifications will be sent regarding the sales and discount offers to the users.

IV. A basic questionnaire will be given to the users to take note of their experience using the application will be provided and the quality and the goal of the project is verified from the customers and will be mapped to the scope of the project and the research gaps of the existing system.

VII. FEASIBILITY

A. Operational Feasibility

Downloading of the application and installing it doesn't require any cost. Moreover, it also reduces the cost of additional carts required for shopping. As the software and hardware required to building the project is low and hence it is economically feasible.

B. Economic Feasibility

Downloading of the application and installing it doesn't require any cost. Moreover, it also reduces the cost of additional carts required for shopping. As the software and hardware required to building the project is low and hence it is economically feasible.

C. Social Feasibility

It will be helpful for the customer as it is less time consuming compared to the traditional one and will be more helpful for senior citizens and the customers who are in rush.

VIII. OUTCOMES

Phase-1:

Implementation of QR code scanner will be done on Android using the ZXing library which supports decoding of QR code.

Phase-2:

A registration and login activity will be developed on Android and the user authentication will be done with API developed using laravel using the JWT authentication package, Integrating it with the QR code Scanner.

Phase-3:

Based on the list generated of the scanned products by the customer/s it will suggest products based on their shopping pattern.

Phase-4:

A feedback form will be generated which will assure the genuineness & quality of the product.

IX. IMPROVEMENT IN RESEARCH GAPS

Barcodes are one dimensional numeric codes, and they are capable of up to 20 characters whereas QR codes can hold up to 7100 characters of data, rather than the much lower number which barcodes hold. Due to the fact that the codes are horizontal and vertical, they store the same exact amount as the barcode can, but in only 1/10 of the space the barcode requires. So, when choosing barcodes versus QR codes, in the arena of data storage, QR codes are far greater at holding and keeping storage. As far as the speed, precision, and scanning position, when considering barcodes versus QR codes, many companies are again choosing QR codes as the choice option for their code types. Due to the three positions and the detection patterns on the QR codes (which are located in the three corners of the code), they are going to be read no matter where they are scanned. Any reader can locate those codes, and easily detect the patterns, in order to give a proper readout of what products are being scanned. Therefore, this makes for faster scanning, and overall quicker checkout for customers waiting in line. A usual sql database contains no security feature since it doesn't have unique hashed key in order to identify the user to whom the data is assigned. This project makes use of JWT through laravel framework which has inbuilt validators for all inputs. Once the user is logged in, each subsequent request will include the JWT, allowing the user to access routes, services, and resources that are permitted with that token. Single Sign On is a feature that widely uses JWT nowadays, because of its small overhead and its ability to be easily used across different domains. Hence it provides enhanced security and there is no chance of sql injection and man in middle attack. domains. Hence it provides enhanced security and there is no chance of sql injection and man in middle attack.

X. OBSERVATION

Table I:Observation

Sr.no	Parameters	Research Gaps	Results Achieved
1.	Data Character	Barcode (Less than 100 characters)	QR code (Alpha numeric max character 4296,bits/char-5.5)
2.	Security	Risk of sql injection, man-in middle attack.	Laravel JWT Authenticator used for prevention of attacks.

XI.RESULTS

As the demand for the online shopping is increasing the requirement of more secure, safe and reliable transaction is of utmost demand. Smart phones, that have become an important part of today's life, have reduced all the efforts that are required for shopping. With camera feature in it, the user can scan the QR code of the item to be purchased and then directly add it into the cart. There are two advantages of it: first no need to stand in the queue for a long time in malls just for scanning the item, second there will be no scope for the frauds that happen in online shopping. The items so far purchased by the customer will be maintained in the app that can be used by the customer in the next purchase. The transactions that will take place frequently with the shop's database will be made secured. This will ensure no modifications in the shop's database either by the customer or by any unauthorized user. After the successful purchase of the products to avoid the chances of fraud there will be a secure OTP obtained to ensure the successful checkout. This application will help to reduce the time taken by the customer and hence provide the assurance of the product quality to the customers

REFERENCES

- [1].Fram, E. H., & Grandy, D.B. (1995), "Internet buyers: Will the surfers become buyers?" Direct Marketing, Vol. 57, No. 10, pp. 63-65. 2.Dr. ParulDeshwal (2015), E-commerce in India: creating bricks behind clicks
- [2].Dr. Rita Sachdev (2015), Role of e-commerce in Indian Economy, International Human Research Journal
- [3].Dr.R.Shanthi& Dr. DestiKannaiah, 2015, Consumers perception on Online shopping, ISSN 2422-8451
- [4] Pritish Jain, 2017, E-Commerce in India: A Comparative study of online shopping with Store Shopping, ISSN 2349-3224
- [5].Aakarsh Rastogi, 2015, E-Commerce – A Comparison with Physical Markets and Reasons for Its Success, ISSN: 2319-7668
- [6].Adarsh Borkar,MadhuraAnsingkar,MonaliKhobragade, Pooja Nashikkar, Arti Raut, 2015 , Smart Shopping- An Android Based Shopping Application , ISSN: 2278-1323
- [7].Hannah Gommerstadt, Devon Long, 2012 , Android Application Security
- [8].www.flick2know.com/QRCodes
- [9].http://qrcodetracking.com/qr-code-capacity/

Dynamic Digital Parking System

Siddhesh Rao

*Department of Information Technology
Thakur College of Engineering &Technology
Mumbai, India*

Yash Shah

*Department of Information Technology
Thakur College of Engineering &Technology
Mumbai, India*

Nikita Sharan

*Department of Information Technology
Thakur College of Engineering &Technology
Mumbai, India*

Neha Kapadia

*Department of Information Technology
Thakur College of Engineering &Technology
Mumbai, India*

Abstract- Today there is a lot of confusion involved in allotment of parking in Public Spaces, Paid parking spaces etc. However, there is no guarantee of availability of parking in some cases when there is too much traffic or crowd. Moreover, there is a lot of manpower involved in making the current infrastructure work, which in turn is a load on the management for making everything work smoothly. The current system is least dependent on technology and does not use any efficient procedure to manage and maintain the parking space. This in turn has led to increased load on human labor which also makes the management of manpower difficult for the park management authority. Due to lack of technical implementation, the system also falls back on security. Our proposed system will feature better technical implementations such as Detection of available spaces in a single parking lot. This is achieved using high definition cameras, powerful cloud infrastructure and various API's. Also, the proposed system will use an interfacing of physical sensors along with combined digital system architecture which will make the installation and management of the parking lot and the budget management comparatively easier than the current manual system. Our proposed system uses physical sensors capable of detecting vehicular movement, unusual activity without the need of heavy processing on the cloud. We use cloud infrastructure only to store details of various parking spaces and to ease the current payment system and provide variety of digital payment processing options.

Keywords: *Internet of Things, Dynamic Digital Parking System, Smart Parking, Dynamic Sensing, Real Time Allocation.*

I. INTRODUCTION

Dynamic Digital Parking System enables the end user to directly locate the nearest available parking space through a dedicated application on his/her mobile phone. The application will check if the locality has any parking spaces available, if yes, the user will be

able to navigate to any one of the displayed parking lot or else he will be asked to move ahead following the GPS on the device till the application finds an empty parking lot. The system aims to ease the current load on the infrastructure by directly giving the ease of location access to the end user.

Also, The Application allows the end user to monitor his vehicle, check the remaining time of parking and even contact assistance in case of any emergency. The Hardware is fitted at intermediate distances in every parking lot which eases the load of space management in the parking. All data transmission is done wirelessly and managed at one central workstation.

II. BACKGROUND

Current Parking System Involves self-navigation by the driver to the parking space, buying a parking ticket and moving forward to park the vehicle. However, the current system needs a lot of manpower for parking management and security among other needs. The system also faces lack of vehicle security and monitoring. Based on our current research we have learnt that the current systems are developed using the following technical advances such as use of IOT for connection of personal devices, use of vehicle detection algorithms based on various models and use of physical sensors which can be used for vehicle security.

Also, we found out some key findings behind the reason to develop a smart digital parking infrastructure. Some findings are as follows:

Increased number of private cars and difficulty in finding a parking spot, lack of management, emphasis to reduce current problem of resource management and reduce waiting cost. Finally need for use of GPS, real-time statistics, smart payments were required. A smartphone application connected to central cloud data center providing all such services is under development.

III. MOTIVATION

There are currently no smart applications and features available at the moment for the management and parking of vehicles which make the management and allocation of

parking spaces more difficult. Smart parking systems enables individuals to save time and reduce traffic problems as well as improves the way people commute. The proposed system includes an effective cloud based smart parking system based on IoT. The system has each car park as a network and data includes GPS location, distance from a specific parking lot & the number of free slots available there, this data is sent to the data center serving as a cloud to calculate the parking amount based on the time. A smartphone application is developed to give the users access and search for the parking available. In this model, the infrastructure allows the end user to reserve a parking space available nearby which can be reached in a stipulated time. Also, the revenue is altered, and new cost models are introduced. A parking solution with Unique Navigational planners is proposed. Efforts on the pricing side also differ from Others work as there is a proposal for real-time dynamic pricing in parking and prove its efficiency.

IV. PROBLEM DEFINITION

Today there is a lot of confusion involved in allotment of parking in Public Spaces, Paid parking spaces etc. However, there is no guarantee of availability of parking in some cases when there is too much traffic or crowd. Moreover, there is a lot of manpower involved in making the current infrastructure work, which in turn is a load on the management for making everything work smoothly.

The development procedure for the proposed system will include development of a mobile application and configuration of server using Raspberry Pi. The Mobile application will be used for live location detection and navigation for the end user to the selected parking space and even for payment purposes for reservation of parking space. The Application will provide a unique QR Code which can be verified at the parking lot which in turn would navigate the user to his/her parking space using the parking lot's internal navigation and mapping system which will be shown on the mobile application on the user's device.

Also, The Application allows the end user to monitor his vehicle, check the remaining time of parking and even contact assistance in case of any emergency. The Hardware is fitted at intermediate distances in every parking lot which eases the load of space management in the parking. All data transmission is done wirelessly and managed at one central workstation.

V. METHODOLOGY

We are using the concept of control points along with marking parking spaces. This increases the overall system accuracy. Every marked Parking Lot is checked for differences using Open CV and image similarity matrix. The Difference in the parking lot (Available/Booked) is

pushed to the server only if there are major changes in the image as analyzed using image similarity.

Every Parking Lot is checked every 5 seconds and three times to check similarities and push changes to server.

The Web Portal is developed using Html & MYSQL. MYSQL is used for establishing connection to the Web Portal from Raspberry Pi to show parking lots and availability to the end user.

The Following flowchart represents the actual working mechanism of the infrastructure.

The application lets the user login and locate the nearest parking space available based on his/her current location.

The user selects and navigates to the nearest car park based on the selection.

The user will be provided with a unique barcode which will identify himself as an authorized user while entering the carpark.

The project will be completed in 5 phases as mentioned below:

Phase-1:

As per the market research, the number of vehicles (Both in terms of Public and Private Vehicles) are expected to rise steeply. This growth in the automobile industry directly translates into an increasing demand for parking spaces. However, there are not much parking spaces available and as per the current condition, even the available parking spaces are creating a congestion for parking of multiple vehicles due to manual management.

Secondly, there are also no smart applications and features available at the moment for the management and parking of vehicles which make the management of parking spaces more difficult. Smart parking systems enables individuals to save time and reduce traffic problems as well as improves the way people commute. Increasing number of automobile and lack of efficient parking management systems are some of the factors driving the demand for smart parking across the globe. These systems help in making the cities greener, cleaner, and technologically advanced by offering several opportunities for the reuse of commercial and municipal parking areas and garages. Moreover, cellphone-enabled automated solutions and low-cost sensors boosting the market for smart parking.

However, the lack of awareness about the smart parking advantages especially in under-developed regions is projected to limit the growth opportunities. The lack of funding initiatives is further projected to hinder the industry growth for smart parking.

Phase-2:

For the development of the system, Certain requirements need to be gathered. The first step in the systems engineering process is understanding the needs of the relevant stakeholders and ascertaining the functional requirements of the crowdsourcing approach. To that end, we have conducted an extensive stakeholder assessment. The goal of the stakeholder assessment is to identify the salient parking challenges in the city and to illicit the role of a technology solution to address these challenges.

Next, we present a study of the effectiveness of the information currently provided by ParkPGH.org. Next, we also present our study that establishes the effectiveness of price changes to decrease cruising for parking in the implementation locality.

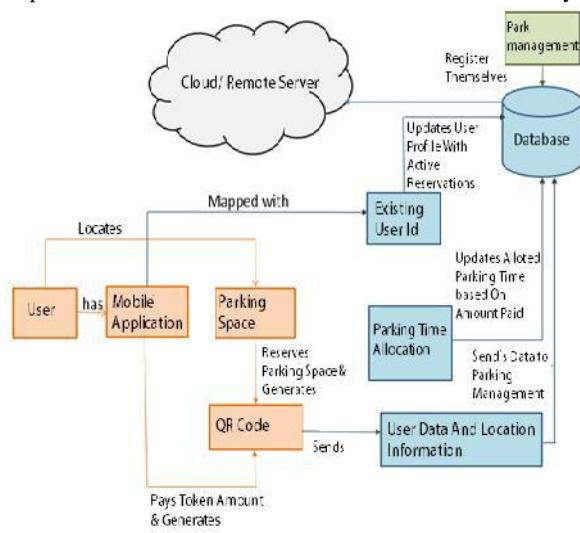


Fig 1: Workflow of Dynamic Digital Parking System using IoT

The goal of this study is to develop an inexpensive parking information and management system by leveraging Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. Our approach to leveraging V2V and V2I to develop a parking management system extends our previous work in parking management systems. We have developed a novel parking prediction scheme. We propose using data from V2V and/or V2I to infer the three key inputs used in our previous research. Each driver looking for parking has a belief about the current availability of parking. When the driver comes into contact with other drivers, updates his belief about the current availability of parking.

VI. RESULTS

We have achieved the following results from the implementation of our project so far. The results include screenshots and snapshots of our system from the web portal and the management dashboard installed on the

Phase-3:

After the market research and requirement gathering, we begin with the development of base model.

The development procedure will include development of a mobile application and configuration of server using Raspberry Pi. The Mobile application will be used for live location detection and navigation for the end user to the selected parking space and even for payment purposes for reservation of parking space.

The Application will provide a unique QR Code which can be verified at the parking lot which in turn would navigate the user to his/her parking space.

The server will be configured on the raspberry pi and will manage the processing of data entry from the user end. The Vehicle detection code will be done using an open source algorithm from OpenCV with Tkinter UI using Python2.

The Camera used for vehicle detection and surveillance will be a module of the raspberry pi programmed to detect vehicular objects and detect the vehicle in and out time.

Phase-4: Once the base model and prototype is ready, the system will be tested in real time at a limited parking location. The testing phase would be named BETA and will include testing in Following aspects:

1. Location detection efficiency.
2. Simulation of parking procedure through the mobile application
3. Testing of navigation to the selected parking space.
4. Testing of payment methods and testing server efficiency.

The testing will check these aspects and feedback will be recorded. These feedbacks will be used to perform improvements in the next phase and release the final product.

Phase-5: In the implementation phase of the project, we will include the bug fixes and changes that we find in the testing phase of the project. Once the implementations are done, the system will be provided as a live application to the end us.

Raspberry Pi.

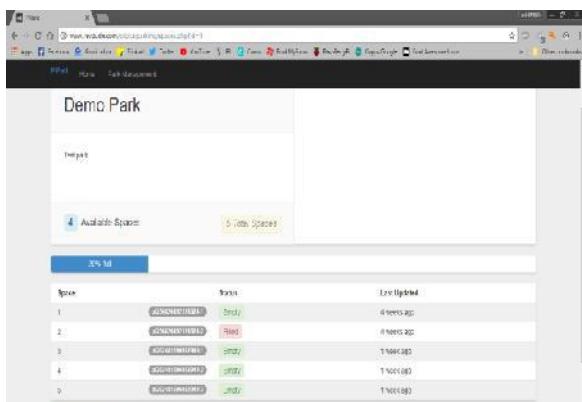


Fig 2: Extended view of a parking lot on web portal showing occupied and available spaces.

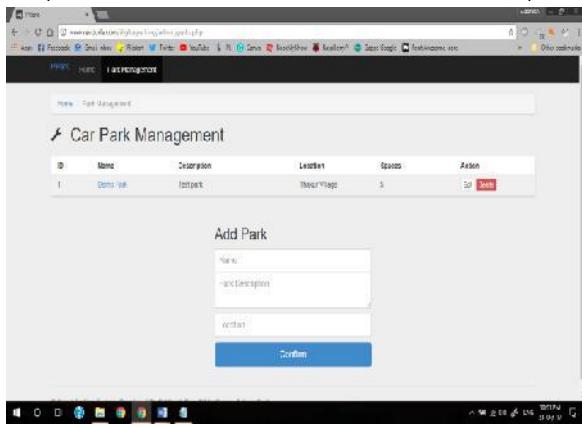


Fig 3: Parking Management Screen Allowing editing current parking lot, Adding New Space

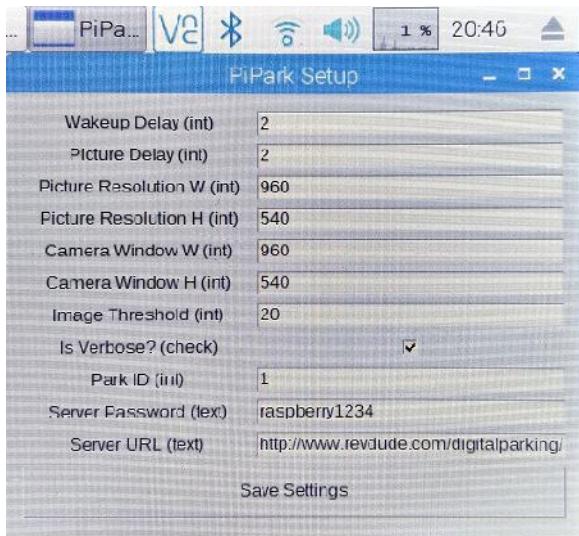


Fig 4: Web Portal configuration on Raspberry pi

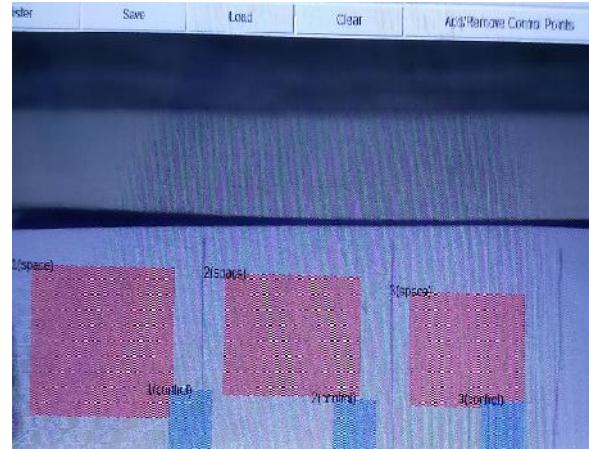


Fig 5: Demo marking of Parking Spaces and Control points on Raspberry Pi

VII. CONCLUSION

The developed system displays the available parking lots through the Web Portal.

The system detects the vehicles entering and leaving the parking lot.

Real Time Monitoring is done and in every 5 seconds the system updates on the server.

It also compares the current image with the previous image and if the parking space is empty it will update that lot as available on the central server.

While setting up the system at the proposed parking location an image of the location is captured, and control points are added to manage the accuracy of parking.

The project conclusion will be attained via two phases as mentioned below:

Software Development Phase:

1. Development of Mobile Application and Web Portal
2. Development of Centralized server for storage of parking localities and vehicle details
3. Development of cloud connection and payment portal to facilitate real time connection of mobile application and server for reservation of parking.

Hardware Configuration Phase:

1. Configuration of Raspberry Pi with Pi Camera Modules on Registered Parking Location's
2. Configuration of IR Sensor's and Ultrasonic Sensor's on parking Spaces and configuring its inputs on Parking management screen.

VIII. FUTURE SCOPE

The proposed system uses the built in GPS Module on the mobile phone to record the current location and allow the user to navigate to the parking space. The main idea behind implementing “Dynamic Digital Parking System” is to automate and simplify the parking allocation and payment process. Overall the proposed system ensures efficient parking allocation and navigation, simple payment modes and vehicle monitoring via mobile application also saving a lot of precious time and workforce.

The Scope of the project is as follows:

- The system can be a cost effective alternative if applied on a large scale.
- The system will reduce the manpower required to manage a parking lot, thereby reducing the overhead cost.
- The system can provide better security features to the vehicle owner through the mobile app like any suspicious movement alert over notification by using the inputs from physical sensors available on the lot.
- There is a scope to add direct connection to emergency services like Police or Rapid response team in case of any emergency.
- The System Can Store Vehicle Data like In Time and Out Time and Provide Help incase Security Situations
- Internal Mapping and Navigation System Designed as per different parking lots.

REFERENCES

- [1]. Thanh Nam Pham, Ming-Phong Tsai, Duc Binh Nguyen(2015)“A Cloud Based Smart Parking System based on Internet of Things Technologies”, Proceedings of the 2015 IEEE Access Volume 3 , Sept 2015. ISSN 2169-3536
- [2]. Abhirup Khanna,Rishi Anand“IOT based Smart Parking System”, International Conference on Internet of Things and Applications,2016,Pune. DOI:10.1109/IOTA.2016.7562735
- [3]. Amir O Koth,Yao-Chun Shen,Xu Zhu.“iParker-A New Smart Car Parking System Based on Dynamic Resource Allocation and Pricing”IEEE Intelligent Transportation System Society,ISSN:1568-0016
- [4]. Y. Geng, C. G. Cassandras, "A new ‘smart parking’ system based on optimal resource allocation and reservations", Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC), pp. 979-984, Oct. 2011.
- [5]. X. Zhao, K. Zhao, F. Hai, "An algorithm of parking planning for smart parking system", Proc. 11th World Congr. Intell. Control Autom. (WCICA), pp. 4965-4969, 2014.
- [6]. Y. Zheng, S. Rajasegaran, C. Leckie, "Parking availability prediction for sensor-enabled car parks in smart cities", Intelligent Sensors Sensor Networks and Information Processing (ISSNIP) 2015 IEEE Tenth International Conference on, IEEE, pp. 1-6, 2015, April.
- [7] Global Smart Parking Market Research Report - Forecast to 2022-<https://www.marketresearchfuture.com/reports/smart-parking-market-2323>
- [8] Development Of Smart Parking System-
<http://dspace.unimap.edu.my/dspace/bitstream/123456789/3108/4/Methodology.pdf>
- [9] UTC Report On Smart Parking- Hampshire-
http://utc.ices.cmu.edu/utc/CMU%20Reports%202013%202/UTC_FinalReport_Smart_Parking_Hampshire.pdf
- [10] ParkPGH.org- Realtime Parking for Pittsburg Garages

Vehicle Tracking/Parking System

Mrs. VandanaMunde
Information Technology
Thakur College of Engineering & Technology
Mumbai, India
vrsakharkar@gmail.com

Yash Mujumdar
Information Technology
Thakur College of Engineering & Technology
Mumbai, India
yash.mujumdar15@gmail.com

ShreyasSanil
Information Technology
Thakur College of Engineering & Technology
Mumbai, India
shreyas9610@gmail.com

Shweta Panchal
Information Technology
Thakur College of Engineering & Technology
Mumbai, India
pshweta1295@gmail.com

Abstract -

In recent times, the concept of smart cities has gained great popularity . Looking at the current scenario the problems such as, traffic congestion, limited car parking and tracking facilities and road safety are being addressed by IoT . In this paper, we present an IoT based System integrated reverse parking and vehicle tracking system. The proposed Reversed Parking and vehicle tracking system consists of an on-site deployment of an IoT module that is used to monitor and signalize the tracking path and the location of the vehicle. Proposed System is a tracking device focusing on vehicles that GSM Module and GPS module for the location of the vehicle with the help of longitude and altitude including obstacle detection with the help of the ultrasonic sensors. Proposed System will give information of location coordinate to mobile phone whenever there is a request for it through the Notification and the sensors will beep and alert the user or the driver. The integrated Global Positioning System and Global System for M obile communication (GPS-GSM) is used to track vehicle using application such as Google M aps or Google Location API.

Keywords - Tracking ; Parking Assistance; IoT; Ultrasonic Sensors, GSM Modem, GPS Modem, Google Location API.

I. INTRODUCTION

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. For matter will need to create these components, incorporating the applicable criteria that follow Vehicle tracking system is a system used to determine the location of a vehicle.using method like Global Positioning System (GPS) and operating system through satellites and device(Reference 1) By using the vehicle tracking system, it is easy to calculate and get the accurate location of the vehicle. Assisted driving technologies which use display system for driver (partial automation). A system and method of controlling an automotive vehicle includes determining a steering wheel angle, determining a steering wheel direction. Instead of completely relying on human intervention, use of device to take charge in critical situations. Reducing the time consumed to park in difficult situations through the help of this aid. Intelligent driver

assistance system provide the full assistance considering the traffic intensity & white lane detection. Reverse parking assistance provided by using Ultrasonic Sensor.

II. BACKGROUND:

Vehicle tracking system is a system used to determine the location of a vehicle using method like Global Positioning System (GPS) and operating system through satellites and device. By using the vehicle tracking system, it is easy to calculate and get the accurate location of the vehicle. Assisted driving technologies which use display system for driver (partial automation). A system and method of controlling an automotive vehicle includes determining a steering wheel angle, determining a steering wheel direction. Instead of completely relying on human intervention, use of device to take charge in critical situations. Reducing the time consumed to park in difficult situations through the help of this aid. Intelligent driver assistance system provides the full assistance considering the traffic intensity & white lane detection.Reverse parking assistance provided by using Ultrasonic Sensor.

III. MOTIVATION

The common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. Vehicle tracking system main aim is to give Security to all vehicles; as well as Locals can easily track/find the vehicle, where talking about reverse parking assistance is introduced after looking at the current scenario where it 's been difficult for locals to park in commercial and industrial not only the system is applicable for reverse parking but as well as for safety purpose on hill side areas.

IV. PROBLEM DEFINITION:

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. Vehicle tracking system main aim is to give Security to all vehicles. This is improved security systems for vehicles. After going through some research papers ‘Pham Hoang Oat, MichealDrieberg and Nguyen Chi Cuong Development of Vehicle Tracking System using GPS and GSM Modem’. And understanding the current scenario we have been able to develop a prototype, where with the association of the GSM Modem and GPS Modem it's been

easy to build a system which easily tracks the location of the vehicle with help of the network. And it has added value to it with the help of its features such as: Obstacle Detection that ensures safety of vehicle and smooth experience. Security through tracking system that used location tracking It can be used for trucks carrying valuable goods, to keep track of the status of delivery and location of the truck at all times You can also use it to keep tab on your driver. It reduces vehicle abuse and ultimately results in significant cost-savings for individuals, fleet owners and the like. You can locate your stolen vehicle easily using your mobile without any extra cost. Semi-Automatic Parking Experience that helps in parking our vehicles in difficult situation.

Phase-1: Market Research:

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. However, there are not much parking spaces available and as per the current condition, even the available parking spaces are creating a congestion for parking of multiple vehicles due to manual management.

Secondly, there are also no smart applications and features available at the moment for the management and parking of vehicles which make the management of parking spaces more difficult. Smart parking systems enable individuals to save time and reduce traffic problems as well as improves the way people commute. Increasing number of automobile and lack of efficient parking management systems are some of the factors driving the demand for smart parking across the globe. These systems help in making the cities greener, cleaner, and technologically advanced by offering several opportunities for the reuse of commercial and municipal parking areas and garages. Moreover, cell phone-enabled automated solutions and low-cost sensors boosting the market for smart parking and Tracking.

Phase-2: Requirement Gathering:

For the development of the system, certain requirements need to be gathered. The first step in the systems engineering process is understanding the needs of the relevant stakeholders and ascertaining the functional requirements of the consumer approach. To that end, conducted an extensive stakeholder assessment. The goal of the developer assessment is to identify the reverse parking challenges and tracking of the vehicle in the city and to illicit the role of a technology solution to address these challenges and understanding the hardware and software requirement.

Hardware:

- Ultrasonic Sensors
- GSM Modem
- GPS Modem
- Raspberry Pi b+
- Wires
- Mechanical Tools

Software:

- Python Idle

Phase-3: Development of Base Model

After the market research and requirement gathering, we begin with the development of base model.

The development procedure will include development of a mobile connectivity and configuration of server using Raspberry Pi. The Mobile device will be used for live location detection and navigation for the end user for tracking the vehicle and while reverse parking the vehicle selected parking space.

The Application will provide the detection of obstacle while reverse parking and tracking of vehicle in the public area and transportation and navigate the user towards the destination where the user car is parked.

The server will be configured on the raspberry pi and will manage the processing of obstacle detection and tracking the location entry from the user end. The Vehicle latitude and altitude will be shown on LCD Screen or the end user using device.

The Camera used for obstacle detection and GPS GSM Module for tracking the vehicle location will be a module of the raspberry pi programmed to detect objects and track the vehicle.

The Following is the flow of the proposed system:



Fig .1 Work flow of proposed model

Phase-4: Testing of Base Model/ Prototype

Once the base model and prototype is ready, the system will be tested in real time at a limited parking location. The testing phase would be named VTPS and will include testing in following aspects:

- I. Obstacle detection efficiency.
- II. Simulation of parking procedure through the Camera module
- III. Testing of navigation to the selected vehicle.
- IV. Testing of server efficiency.

The testing will check these aspects and feedback will be recorded. These feedbacks will be used to perform improvements in the next phase and release the final product.

Phase-5: Deployment of Project

In the implementation phase of the project, we will include the bug fixes and changes that we find in the testing phase of the project.

V. METHODOLOGY

This vehicle tracking system takes input from GPS and sends it through the GSM module to the desired mobile/ laptop using mobile communication. Vehicle Tracking System is one of the biggest technological advancements to track the activities of the vehicle. The security system uses Global Positioning System GPS, to find the location of the monitored or tracked vehicle and then uses satellite or radio systems to send to send the coordinates and the location data to the monitoring center. At monitoring center,

VI. RESULTS & CONCLUSION:

Vehicle tracking system main aim is to give Security to all vehicles. Accident alert system's main aim is rescuing people from potential accidents. This is an improved security system software is used to plot the Vehicle on a map. In this way the Vehicle owners are able to track their vehicle on a real-time basis. Due to real-time tracking facility, vehicle tracking systems are becoming increasingly popular among owners of expensive vehicles. The below block diagram explains the working of the system. Controlling of the other supporting devices those are GPS, GSM, ultrasonic sensors, mobile, Raspberry pi, LCD screen. Actually, this gives the practical working of a vehicle tracking parking system which can do reverse assisting for parking, tracking of moving vehicles in area. It consists of two sections, first which will be built on the prototype which is having GPS in it and as the car moves the location of the car goes on changing continuously, the GPS find the location in terms of two coordinates that are Longitude and latitude. These co-ordinates are communicated to another section by GSM device, as shown in the block diagram both the parts consist of GSM for communication. This GSM is connected to the computer through RS232 which should have internet connection in it. So, that as soon as the co-ordinates are received, it can be located.

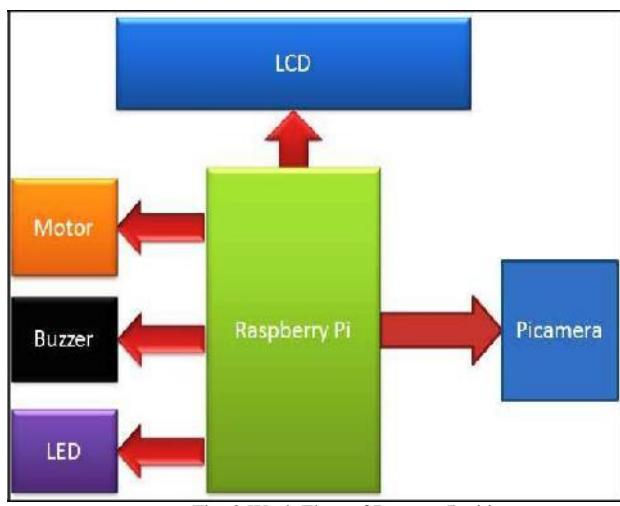


Fig 2 Work Flow of Reverse Parking

for vehicles as it offers GPS ability which is very useful in the current scenario where we have improved the management system with the help of Google Location API and Ultrasonic Sensors for Vehicle Tracking and Parking System. Vehicle Tracking & Parking Systems has the potential of creating many wonders in the security and safety of both, the vehicle in such a manner that it is not visible to anyone who is inside or outside of the vehicle.

This tracking system can store the whole data where the vehicle had gone, where it stopped, how much time it takes at every stop and can create whole data analysis. It is also used in buses and trains, to

estimate how far are they, how much time it takes for them to come to a particular stop.

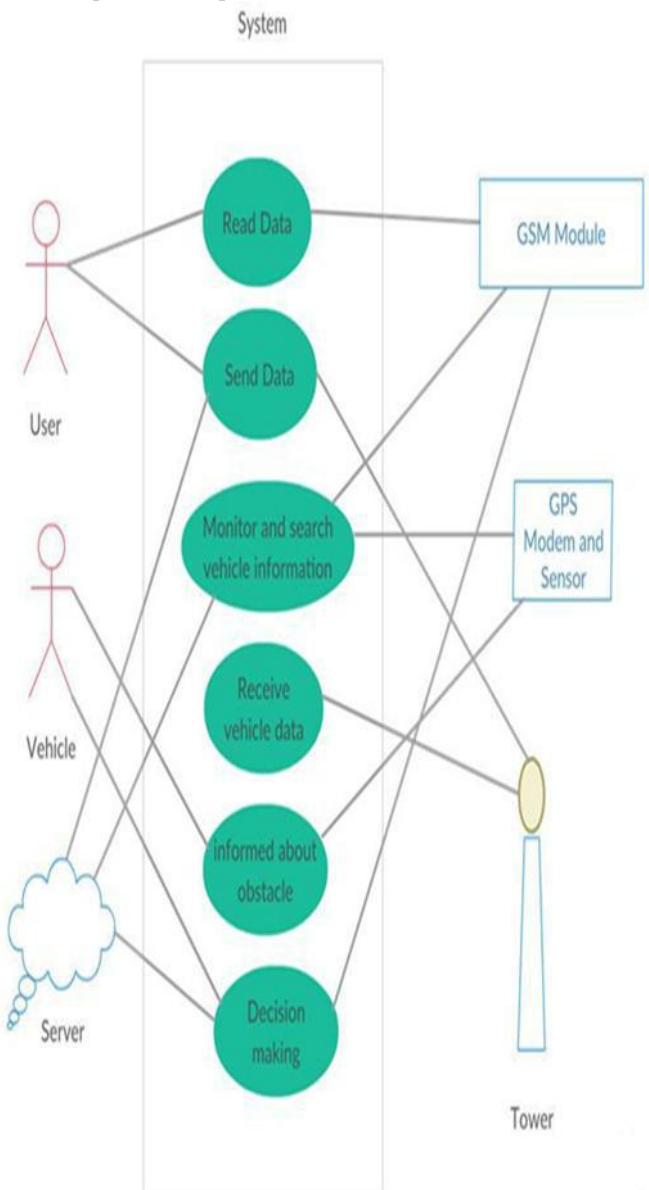


Fig 4 Use Case Diagram

By integrating all the data from various phases, it will be easy for the vehicle tracking and reverse parking to all the end user for their personal experiences and for business purposes as well as safety.

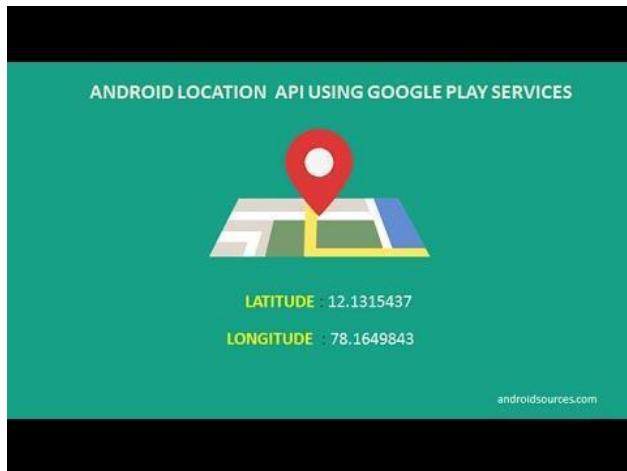


Fig 3 Proposed System Output

VII. FUTURE SCOPE:

More and more businesses are turning to GPS technology to manage vehicles, their employees, and their assets. Businesses use GPS Vehicle Tracking and reverse parking assistance to assess driver performance and to keep track and parking of users while they are travelling. As the devices become smaller and more accurate, it becomes a safe bet that even more businesses will begin to utilize this powerful tool. It is mainly benefit for the companies which are based on transport system. This tracking system can store the whole data where the vehicle had gone, where did it stop, how much time it takes at every stop and can create whole data analysis. It is also used in buses and trains, to estimate how far are they, how much time it takes for them to come to a particular stop.purposes while travelling to different location and parking their vehicle in the city.

REFERENCES

- [1]. Abhirup Khanna "IoT based Smart Parking System 978-1-5090-0044-9/ 16/\$31.00 ©2016 IEEE Pune, India
- [1]. Pham Hoang Oat, MichealDrieberg and Nguyen Chi Cuong Development of Vehicle Tracking System using GPS and GSM Modem 978 -1-4799-0285-9/13/\$31.00 ©2013 IEEE Malaysia
- [2]. FatinBalkisBinti A IzahriVehicle Tracking Device 978-1-5090-1636-5/ 16/\$31.00 ©2016 IEEE Pu lau Pinang, Malaysia

Fingerprint Recognition for Examination System

Ayush Datta

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
ayushdutta3@gmail.com*

Harshal Kale

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
hkael17@gmail.com*

Shubam Tare

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
shubamtare06@gmail.com*

Vijaykumar Yele

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
vijaypyele@gmail.com*

Abstract— In biometric system, the fingerprint recognition has been researched for the long period of time and it has shown the most promising future in the real world application. However, because of the complex distortions among the different impression of the same finger in real life, fingerprint recognition is still a challenging problem. Matching two fingerprints can be unsuccessful due to various reasons and also depends upon the method that is being used for matching. Fingerprints were missing from the database or the fingerprint images were corrupted due to which the system was not able to recognize the user and authenticate them. The fingerprint scanner was not able to scan the fingerprints of the user and create an image out of it, due to which the authorization of the user was found out to be difficult.

The research gap concluded from the literature survey had given us the result of the efficiency of the system as 80%. This result was based on the tests conducted on the systems similar to our project. The parameters used for measuring this result was the fingerprint image collected from the users and stored them in database and then using these images as a way to authenticate the users. The success rate of recognition, the number of images stored successfully in the database per user, the number of users authenticated against their fingerprint images stored in the database, these are all the parameters for measurement which resulted in depicting 80% efficiency of the system. According to the literature survey created by researching on the various papers related to the topic it was found that after considering the total number of users as 35 and two images per user we collected about 70 images of their fingerprints for biometric authentication. After determining all the results from the literature survey the correct recognition of about 90% in our project is expected. After studying the literature survey it was planned that the system was tested using the test databases and then verified the authorized user as per the test images stored in our database. This helped us in rectifying one of the technical challenges as the databases were able to store the images successfully and authenticate the test subjects. Further, in our project we will authenticate the users as per their thumb impression and then we will generate their mark sheets assigned against their database information stored.

unchanged throughout life. Fingerprints offer an infallible means of personal

identification. Other personal characteristics may change, but fingerprints do not. Fingerprints can be recorded on a standard fingerprint card or can be recorded digitally and transmitted electronically to the FBI for comparison. By comparing fingerprints at the scene of a crime with the fingerprint record of suspected persons, officials can establish absolute proof of the presence of identity of a person

II. EASE OF USE

A primary motivation for using biometrics is to easily and repeatedly recognize an individual so as to enable an automated action based on that recognition. The reasons for wanting to automatically recognize individuals can vary a great deal; they include reducing error rates and improving accuracy, reducing fraud and opportunities for circumvention, reducing costs, improving scalability, increasing physical safety, and improving convenience. Often some combination of these will apply. For example, almost all benefit and entitlement programs that have utilize. Biometrics has done so to reduce costs and fraud rates, but at the same time convenience may have been improved as well. Historically, personal identification numbers (PINs), passwords , names, social security numbers, and tokens (cards, keys, passports, and other physical objects) have been used to recognize an individual or to verify that a person is known to a system and may access its services or benefits. For example, access to an automatic teller machine (ATM) is generally controlled by

Keywords- *fingerprint, biometric, finger.*

I. INTRODUCTION

Fingerprint Identification is the method of identification using the impressions made by the minute ridge formations or patterns found on the fingertips. No two persons have exactly the same arrangement of ridge patterns, and the patterns of any one individual remain

requiring presentation of an ATM card and its corresponding PIN. Sometimes, however, recognition can lead to the denial of a benefit. This could happen if an individual tries to make a duplicate claim for a benefit or if an individual on a watch list tries to enter a controlled environment. But reflection shows that authorizing or restricting someone because he or she knows a password or possesses a token is just a proxy for verifying that person's presence. A password can be shared indiscriminately or a physical token can be given away or lost.

III. LITERATURE SURVEY

In the paper [1], the proposed solution refers to an automatic verification of a person based on their specific traits. Paper [2] refers to verification of a person based on their specific traits. In paper [3] it refers to Fingerprints is using Finger Geometry of a person/user. [4] It refers to be feature extracted from image and features measurement of Finger geometry are based on the measurements of length of finger, width of finger. So the research gap found out to be number of users is thirty five, images are seventy so the expected correct recognition is to ninety percent.

IV. PROBLEM DEFINITION

Phase-1:

Getting clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation.

Phase-2:

After analysis we will first study about it and do some research on it for our better understanding of the project and also getting a rough picture about what would be our problem definition for the particular project.

Phase-3:

Then we will construct the design of the project and according to that, will list down all the requirements needed for the construction for the prototype of our project.

Implementation: After acquiring the requirements we will proceed to the construction of the application by using MATLAB and other technologies.

Phase-4:

After the prototype is ready we will first connect the hardware with the assigned code and then we will check if it supports the mechanism or not. If not we will solve the issues regarding to it and will check again.

Phase-5:

After, complete integration and testing of project real time running and operation of the system will be done. Students are expected to validate against their previously stored data in the database to obtain their mark sheets and transcripts.

V. METHODOLOGY USED

The proposed solution, for our project is mainly used to verify fingerprint. If the person who wants to generate mark sheet and tries to validate himself then he has to enter his fingerprint. Then this fingerprint will be matched with the one stored in the database. If it matches mark sheet will be generated.

A. *Technical Feasibility*

Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

B. *Economic Feasibility*

As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

C. *Legal Feasibility*

Legal issues can affect a system's acceptance by users, its performance, or the decisions on whether to use it in the first place—so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any biometric system.

D. *Operational Feasibility*

It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

E. *Social Feasibility*

The acceptability of a biometric system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness

VI. OUTCOMES

Phase-1:

Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

Phase-2:

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done

Phase-3:

Proposing design architecture of Fingerprint recognition for examination system for implementing the surveyed case studies and synopsis.

Phase-4:

The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit/performance testing) and perform integration testing, system testing and stress testing.

Phase-5:

Realtime Deployment of project collecting data of 35 students. Recording 2 images per student. Obtaining 90% Accuracy in recognizing the input data with stored data.

VII. IMPROVEMENT IN RESEARCH GAPS

Referring to the literature survey performed by us and the research gaps observed we found out that the accuracy of most of the systems was around 80%, because it was observed that there was lack of testing of the system and hence the overall efficiency turned out to be low.

Another reason was that the data of the fingerprint was missing from the database; the images of the fingerprints were turned out to be corrupt or not recognized by the system. But, after performing the task of storing the images in the database we found out that the probability of missing images or corrupt images was low. Hence, improving the efficiency.

VIII. RESULTS

The expected outcome after completing the project by using arduino controller by taking fingerprint of 4 to 5 students in the database, By recognizing their fingerprint we will be able to generate the result on PC and send it further on their email id.

REFERENCES

- [1]. Anil Jain, Ruud Bolle and Sharath Pankanti,"Biometrics Personal Identification," in Networked Society.
- [2]. Aditya S., Amogh P., Rahul M. and Swapnil Z," Biometric Identification System: A Finger Geometry and Palm Print Based Approach , "International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 5, Issue 3, March 2016, pp: 2320 – 3765.
- [2]. Raffaele C., Dario M., Davide M., James L. W. and Anil K. Jain, 'Performance Evaluation of Fingerprint Verification Systems; IEEE transactions on pattern analysis and machine intelligence,' vol. 28, no. 1, January 2016.
- [3]. Arsalaan. F.R., Mehreen L., Balbir K., O. P. Aggarwal, Sajad H. and Neeraj Gupta,"Biometric Finger Print Identification- Is It a Reliable Tool or Not?," J Indian Acad Forensic Med., April-June 2013, Vol. 35, No. 2, pp: 0971-0973.
- [4]. Madhuri and Richa Mishra, Fingerprint Recognition using Robust Local Features , " International Journal of Advanced Research in Computer Science and Software Engineering, "Volume 2, Issue 6, June 2012, pp: 2277 128X.

Website and Application For Result Analysis

Arvin Gupta

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai India
arvindogupta@gmail.com*

Akhil Baliga

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai India
akhilbaliga@gmail.com*

Sandip Bankar

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
sandipbankar1@gmail.com*

Abstract – The Current exam-cell activities are mostly done on paper. Automated solutions using this system will make exam department activities more efficient by covering for the most important drawbacks of manual system, namely speed, precision and simplicity. A centralized system will ensure that the activities in the context of an examination can be managed effectively, while also making it more accessible and convenient for both students and staff. The final product would constitute a computerized module aimed at replicating offline exam cell process. The system is a new concept which came into existence because of the large amount of data being on paper and it made analysis of results a tedious task, apart from the unmanageable amount of data that is generated in an institution from various departments. The Automation system is like an intermediary between staff and students, thus easing the activities of each regarding examination. It is a system that will make the exam cell process much organized. It would require certain crucial data to be pre-existing, which definitely suggest a dependency on certain other systems, especially ones concerning data acquisition. Like in this case it is Automated Result analysis system. The Solution, however, will manage a great deal of minimal work. This keeps paperwork to its minimum, leading to ease of accountability, reducing confusions and increase in work rate and efficiency. The project will address firstly, access to various users including students, teacher, exam cell staff and admin. The customization based on requirement of College can be done.

Keywords- Credit based system, Automation, Data Acquisition.

I. INTRODUCTION

Examination cell, referred to as the Exam cell is an integral part of college management system. As we all know examination in an engineering student's life is the foremost and an important issue.[1] Hence management of this system requires a huge effort. The basic challenge of an institution is to centralize, track and resolve various student issues before and after the examination, manage the various operations and information in a flawless manner and deliver quality result oriented education. Student User Student will be conducting following activities:-

On a day to day basis, following are the basic problems faced by the institution:

- Operation is centred in administrative offices
- Procedures are heavily bureaucratic
- Information needs to be updated
- Information is inconsistent and replicated
- Result Generation
- Statistical Analysis of result

This is computer based system for result analysis .College got the result from university in PDF form .This system generate the report in document form which we can easily edit and update the result which makes the dynamic changes. The PDF format is very difficult to modify and requires the use of the complex software. To simply put, system is provide the result analysis report for the credit based system. It is very much essential to implement result analysis system in higher education System [3].

The main Goal is to analysis of student result as per the credit base system calculate and analysis grade wise

II. CREDIT BASED RESULT ANALYSIS SYSTEMS

Being at the beginning stage of the credit based result analysis Systems, computerization of the student result analysis is tedious task. The main reasons for using computer base result analysis system are to calculate the individual result, college result and to find out the first three topper ranker. In other words, Credit Based Result Analysis Systems increase effectiveness and efficiency by saving time[3].

This systems support not only Result analysis but also in adaptable to changes, and are helpful to cope with the demands for change. That is update and edit information about student result and own information of student.[3]

1. Student Activities:

Student User

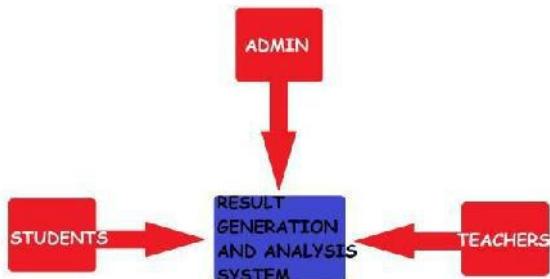


Figure 1: Student User

III. MOTIVATION

Considering todays scenario we decided to develop this system that will reduce the work and increase the productivity. User just need to provide data and our system will automatically develop the result[2].

IV. PROBLEM DEFINITION

For analyzing the marks obtained by students in an educational institution. We are tasked to build up student mark analyzing system. This is done to replace the manual entering and processing of marks which are error prone and tedious. This system also maintains information about student. The system will have a Windows based desktop interface and mobile application to allow the faculty to enter marks obtained by the students, update them and generate various reports. For security reasons, the administrator and faculty only can update the marks and other information. First the user needs to login to the system for accessing it. The system will retain information on all the students and the institution. The system analyses the marks and generate the result reports. The marks and information about the students are stored in a database and the system works with the database[4].

Phase 1(Information Gathering) :

Study different papers related to result analysis and getting more and more information about it. To know about the Result

Analysis and types of algorithms which is required in creating Result Analysis.

Phase 2(Analysis& Design) :

Develop the Graphical User Interface and analyze the internal structure and working model of the result analysis and how it will react to the user query. Making a list of different features that we can implement in our result analysis.

Phase 3(Implementation & Testing):

The design and implementation of this maintain student information. It replace the current paper records. User access grade wise marks.

Cumulative Grade Performance Index (CGPA):

An up to date of an overall performance of student from the time he/she enrolled in the university is obtained by calculating a number called cumulative grade performance index. It is calculated in CGPA is responsible to reflect final pass or fail status of student.

$$\text{CGPA} = \frac{C_1 g_1 + C_2 g_2 + \dots + C_n g_n}{C_1 + C_2 + \dots + C_n} [2]$$

We improve the way result analysis providing rich functionality. Our wraparound services provide value-added information and support for data, security, storage and implementation.

User able to directly access all aspects of a student's academic progress.

Phase 4(Deployment) :

Once the website is built, it will be deployed to hosting environment which is our college website. The hosted environment must be stable and needs its own monitoring and develops support.

V. METHODOLOGY

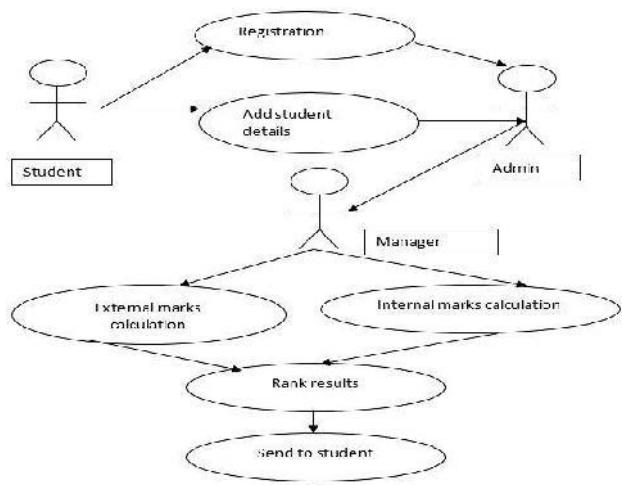


Figure 2: Use Case Diagram

VI. TECHNOLOGY USED

This system makes use of the following software is made using very basic programming languages and includes Microsoft ASP .Net(Framework 4.5): Front end.

SQL-Server: as database language HTML: at front end

ASP.NET is a set of technologies in the Microsoft .NET Framework for building Web applications and XML Web services. ASP.NET pages execute on the server and generate markup such as HTML, WML, or XML that is sent to a desktop or mobile browser. ASP.NET pages use a compiled, event-driven programming model that improves performance and enables the separation of application logic and user interface. ASP.NET pages and ASP.NET XML Web services files contain server-side logic (as opposed to client-side logic)

VII. SCREENSHOT



Figure 3. Login page

Figure 4 :Registration of User

Search Student Marks By PES Number					
Marks Details					
Subject Name	Internal Marks	External Marks	Total Marks	Grade	Pass/Fail
Principle of programming language	90	90	90	A	Pass
Algebra	40	40	40	C	Fail
Data Structure Design	80	80	80	B	Pass
Design and analysis of algorithm	80	80	80	B	Pass
Network Programming	80	80	80	B	Pass
Object oriented programming	80	80	80	B	Pass
Algorithm Design	80	80	80	B	Pass
Data Mining and Data warehousing	80	80	80	B	Pass
Project	90	90	90	A	Pass
Programming with C/C++	80	80	80	B	Pass
Artificial Intelligence	20	20	20	C	Fail
Software Testing & Application Testing	80	80	80	B	Pass
Mobile Computing	80	80	80	B	Pass
Cloud Computing	80	80	80	B	Pass
Big Data	80	80	80	B	Pass
Blockchain	80	80	80	B	Pass
Business Analytics	80	80	80	B	Pass
Machine Learning and Data mining	80	80	80	B	Pass

Figure5:Individual Mark

Student Information		Marks Details		
PES No-0000000000 Student Full Name-Vivek Vaishnavi Gender-Female E-mail-id-vivek@gmail.com		Subject Name	Internal Marks	External Marks
			Total Marks	
(C2-001) Principle of programming language	90	90	90	A
(C2-002) Advance networking	80	80	80	B
(C2-003) Object oriented programming	80	80	80	B
(C2-004) Design and analysis of algorithm	80	80	80	B
(C2-005) Discrete mathematics	80	80	80	B
(C2-006) Digital image processing	80	80	80	B
(C2-007) Advance operating system	80	80	80	B
(C2-008) Data mining and data warehousing	80	80	80	B
(C2-009) Software Engineering	80	80	80	B
(C2-010) Programming with C/C++	80	80	80	B
(C2-011) Advanced Java programming	80	80	80	B
(C2-012) Advance Design and Analysis of algorithms	80	80	80	B
(C2-013) Software Metrics & Project Management	80	80	80	B
(C2-014) Mobile Computing	80	80	80	B
(C2-015) Big Data	80	80	80	B
(C2-016) Project	80	80	80	B
(C2-017) Web Technologies	80	80	80	B
(C2-018) Database and its views (Relational)	80	80	80	B
(C2-019) Business Intelligence	80	80	80	B
(C2-020) Industrial Training, Non-Technical project	80	80	80	B

Figure 6: Individual student marks, percentage, student information details and grade weather student is pass or fail

VIII. RESULTS AND CONCLUSION

The main objectives of the project will be to develop an algorithm that will be used to get the data from the admin and generate the result as per student. To develop database were all related data will be stored and good interface.

The interface developed has two parts, one for simple user and one for the administrator.

A database is developed which stores information about marks, student details, student performance. A usable system will be designed, developed and deployed to the college web server.

REFRENCES

- [1].International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-2, Issue-4, April 2015
- [2].International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 1, January 2015.
- [3].Abel U. O. & Abu Mallam, (2013). "Data Analysis and Result Computation (DARC) Algorithm for Tertiary Institutions", IOSR Journal of Computer Engineering IOSR-JCE, Vol.14, No.3.63-69. Available: www.iosrjournals.org/iosr-jce/papers/Vol14-issue3/K01436369.pdf?id=7439
- [4].Ukem, E et al. (2012). "A software application for university student's results processing", Journal of Theoretical and Applied Information Technology, Vol. 35 No.1. Available: www.jatit.org, (July 2, 2012).
- [5].Zinged, Laird Darning, G. Blake Mike, and Masumi Nakamura," Programming Android", Second Edition 2009
- [6].Android SDK Android Developers <http://developer.android.com/sdk/index.html>.

Log Stacking and Threat Hunting

Pranjal Aswani

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
aswani.pranjal@gmail.com*

Abhishek Naik

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
abhishek.naik1996@gmail.com*

Amaan Cheval

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
amaan.cheval@gmail.com*

Hetal Amrutia

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
hetal.mecs@gmail.com*

Abstract

This project is aimed, as the name suggests, at analyzing logs. Logs are the trails that the systems we use leave when they are working. By looking at the trail that an application leaves behind, we can tell what the application was up to, where it went and what other applications it triggered. We can also tell if someone was controlling that application and trying to misuse it. This can lead us to find out any bad or malicious entities in our system. By applying stacking mechanisms and some analysis, we try to simplify this whole process for the user. This application is able to clean the logs using a delimiter or a grok pattern and then index those logs into Elasticsearch so as to aggregate them later. We have looked at the various aggregation functionalities that Elasticsearch and our algorithms provide and tried to improve on Splunk framework. We have also tried to increase the indexing capacity of Elasticsearch by using bulk indexing techniques and multithreading for better performance. We intended to extract specific parameters from several formats of logs such as system logs, network logs, application logs, windows logs, ssh logs and such from 134 core query parameters identified. And we intend to detect signs of compromise or attacks based on 11 or 15 parameters respectively, which based on research account for 90 percentile of all detected compromises. By using clusters of multiple homogeneous nodes and scheduling techniques we intend to improve the indexing rates by 15 percent as compared to what would have been achieved through a regular relational database write. For user friendliness and accessibility we have enhanced the 134 core query commands and added 515 commonly used GROK patterns which are named collections of regular expressions. The number of relevant parameters for indication of compromise will be used by machine learning algorithms in future work.

Keywords Log analysis, Threat Hunting, Log stacking.

I. INTRODUCTION

“In computer log management and intelligence, log analysis (or system and network log analysis) is an art and science seeking to

make sense out of computer-generated records (also called log or audit trail records). The process of creating such records is called data logging.”

Log analysis is a booming field and a lot of companies regularly analyze their logs to find out if something is going wrong with their systems and if there is smooth functioning or not. Log analysis enables us to keep the systems safe and intact when they are breached. Generally when we hunt for threats we use the available logs that the system generates. They can be apache logs (for apache servers), access logs, ssh or ssl logs, BRO logs and alike. Logs contain information about the requests sent to the system and what resources were accessed or were tried to be accessed by the user and were denied. Logs can tell us a lot about how the system is functioning and if someone tried to infiltrate into the system. Security is a crucial part of an organization. By logging, even if the organization's security gets breached, we can find out by WHOM, HOW and WHEN the system was breached. Information security is a vitally important aspect of all organizations and can singularly determine the success or failure of a project based on a single breach of trust or user-data alone; not to mention the variety of legal and compliance requirements. According to several compliance laws, an organisation is legally accountable for reporting a breach it has incurred within a specific deadline to users and to the federal authorities.

II. EASE OF USE

Our project allows for SOC teams to set up automatic log analysis systems to better visualize and detect possible anomalies indicated by behaviors exhibited by systems, networks, applications, and anything on the periphery that might generate logs. We aim to also allow for a less hands-on approach to threat-hunting by incorporating machine learning algorithms which will automatically analyze logs and highlight or alert suspicious events.

III. LITERATURE SURVEY

We intend to use Elastic search as our primary storage medium. Elastic search provides fast indexing if the cluster of nodes is properly set up but querying it is a bit complex. It is capable of indexing 1 Million queries per second on a 150 node cluster [1]. Although we did not scale to that level, we used threading and parallelism to achieve fast indexing. For processing the data, we intend to group and reduce the data set to identify the outliers and

make manual analysis of an entire environment feasible. Number of relevant parameters: 15 [2]. Bots and malware try to connect to the master node so as to get payloads and updates from them. We look at certain same combinations of busier-than-usual external destinations they connect to and certain fields that they manipulate so as to identify them inside the network [3]. Splunk is another log analysis tool kit but it consumes a lot of time and is not for the common users to quickly detect anomalies in heaps of data. It takes time to ingest data and convert it. Number of commands: 134 core commands are present. 10% of queries were aggregations.[4]. The main idea of data stacking is to find anomalies using LEAST FREQUENCY COUNT. That means that the entities which are in minority will most probably be an anomaly and malicious in nature [5].

To detect threats and attacks on the system by taking the logs generated by the system, cleaning the unnecessary data with the help of GROK patterns or delimiters. Then saving the data into Elastic search using schedulers and threading techniques for fast insertion and retrieval. Then analyzing logs and stacking and aggregating the data according to different fields so as to see the relations and spot the anomalies using least frequency count methodology using pandas library and flask, aiming to improve indexing rate by 15% and take lesser relevant parameters into account too.

IV. PROBLEM DEFINITION

The problem is how do we make detecting anomalies in a bit and complex environment easier for the monitoring team. How will threat hunting become more feasible and quick.

PHASE 1:

Research: Which storage software should we use to properly store the data? Elastic search? They should be easy to query, should have an API which is easily accessible and the data remains secure. Along that we have to decide how much control should the user have? With Elastic search, we will have to use threading and scheduling to increase the indexing capabilities. Will this be oriented towards the technical user or the manager without a technical background should also be able to use this software.

PHASE 2:

Experimentation: Experimenting with different types of softwares and modules so as to discover which ones can we use other than the basic ones. Are they feasible to use? Should we consider some other options?

PHASE 3:

Proof of Concept development: Here we will try and the very basic structure of the application which has the core functionality. This should be pretty simple once Phases 1 and 2 are done successfully. This should also help us decide which features we can add to the application which makes it stand out amongst the other applications.

PHASE 4:

Enhancing the PoC: Once phase 3 is complete. We will have a pretty good idea in the way this application is going to turn out. We should be able to then clean the code, structure it better and optimize it so as to make the application load and run better.

PHASE 5:

Testing and GUI modification: This will be the final stage of our project. It is during this stage that we will test the application to see if it is working the way we want it to and try to stress test it to see if it breaks and fix it if does.

We will also make the GUI more attractive and easy to use for any type of user.

V. METHODOLOGY

The backend of the application is python which uses the Flask framework to create the REST endpoints which the front-end relies on for the user friendly GUI. The application requires an Elasticsearch instance, either local or remote, which it can connect to dynamically through configuration parameters provided by the client. The port of the local Elasticsearch instance is configured to be the default 9200. It can be changed in the configuration file. Python functions have been implemented using the elastic search and Elasticsearch-dsl libraries which provide us with the ability to connect to Elasticsearch, query it, bulk index the logs and perform other necessary functions. Apart from these libraries, Pandas was used to create data frames and for security, we used Flask's CSRF to prevent the highly prevalent OWASP top 10 cross-site request forgery attacks on our APIs. The front-end is designed based on real-life examples of customer usage patterns for similar analysis applications. The template used was SBAadmin which was configured with respect to Flask to blend in. jQuery was also used to simplify upload and download of files across all major browsers for compatibility, as this application will be hosted on the cloud also. Users currently need to understand GROK patterns, but in the future a tooltip or help guide will also be provided inline with the web application

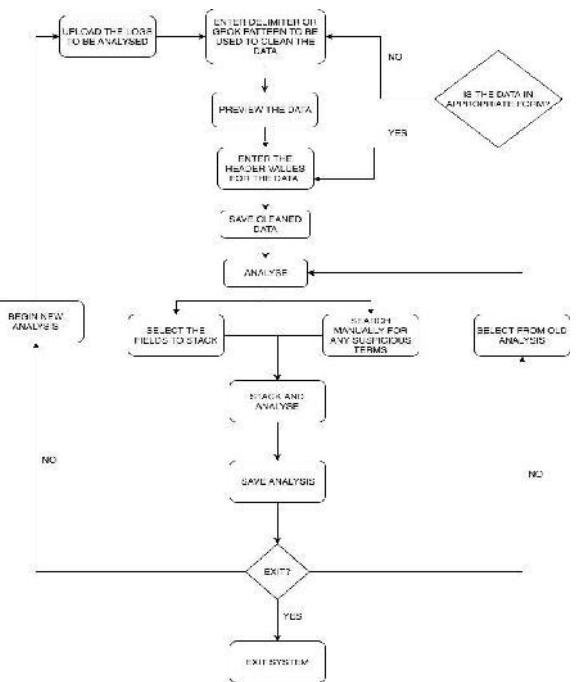


Fig 1: Flowchart

parameters from several logs such as system logs, network logs, application logs, windows

logs, ssh logs and such from 134 core query parameters identified. And we intend to detect signs of compromise or attacks based on 11 or 15 parameters respectively, which based on research account for 90 percentile of all detected compromises. We intend to improve the indexing rates by 15 percent as compared to what would have been achieved through a regular relational database write. For user friendliness and accessibility we have enhanced the 134 core query commands and added 515 commonly used GROK patterns which are named collections of regular expressions. Our proof of concept already supports pattern matching and indexing speeds of 335 requests per second. The number of relevant parameters for indication of compromise will be used by machine learning algorithms in future work.

I. OUTCOMES

Phase-1:

We selected Elastic search as our indexing server due to its high throughput, its versatility, and its rigid schema, which allows for uniform APIs to be used and for better aggregation capabilities for future feature additions.

Phase-2:

We experimented with programming languages and frameworks to determine the ideal technology stack to allow for both ease of development and feature-growth, and high-performance application support. We settled on using Python, Flask, Elasticsearch-DSL, and PyGROK. For future additions we might use Celery for asynchronous multithreading support.

Phase-3:

We built a proof of concept REST API using Python and the Flask package, where the API supports uploading logs, providing or deleting GROK patterns, and stacking the data to find patterns within the aggregated and pattern-matched data.

Phase-4:

The performance of the project will be improved by rigorous benchmarking and optimization of code to allow for concurrent programming models to be used. We will also examine the possibility of implementing machine learning algorithms (such as logistic regression, DGA analysis, anomaly detection, beaconing detection) in order to automatically detect or highlight anomalous behavior which may indicate malicious / compromised systems.

Phase-5:

As part of the closure, we will polish the UI and ensure that all requirements have been met for this project to be a successful and competitive product in the market.

VII. RESULTS

Table I: Results

Sr no	Parameters	Research Gaps	Results Achieved
1.	Rate of pattern matching and indexing	Regular Expression and relational DB (200 requests / second)	GROK patterns and ElasticSearch (335 requests / second)
2.	Number of core commands to query logs	134 patterns	515 GROK patterns

The use of a relational database is unnecessary for the purpose of threat hunting logs; as such, we use Elastic Search instances to index data after using GROK patterns to filter data. This allows for faster ingestion of logs and therefore allows for higher throughput and performant detection.

We also provide 515 preconfigured GROK patterns for more efficient user friendly pattern matching and querying of specific fields within data logs. 134 patterns form the basis of all log analysis required; we therefore build upon that research to implement more human-friendly abstractions with 515 patterns.

VIII. CONCLUSION

As threat hunting becomes a part of the security administrator's toolkit, applications that allow for proactive detection of security incidents such as through canary files, honeypots, baseline detection, and SOC monitoring, our project addresses a key aspect of the threat hunting landscape. The project makes it easier for both casual end-users to analyze logs for patterns in order to visualize baselines, and for automated tools to be programmed by teams for security critical applications. The application allows one to detect compromises before any malicious activity may even have occurred simply by detecting all anomalies; this proves very useful for corporations as it enables them to prevent breaches in the bud and possibly even catch attackers in the act, possibly redirecting them to honeypots or simply adding additional logging to actually find the perpetrator. This application will help to reduce the time taken by the customer and hence provide the assurance of the product quality to the customers.

IX. ACKNOWLEDGMENT

We sincerely thank our guide Ms. Hetal Amrutia for her guidance and constant support and also for always being there for us. We also thank our guide for giving us new ideas to develop our project well. We also thank our project coordinators for arranging the necessary facilities to carry out the project work. We thank Dr. Rajesh Bansode (HOD - IT), Dr. R.R Sedamkar (Dean Academic), Dr. B. K. Mishra (Principal) and the college management for their support.

REFERENCES

- [1] <https://benchmarks.elastic.co/index.html>
- [2] <http://people.ischool.berkeley.edu/~hearst/papers/1isa2014-final-version.pdf>
- [3] <http://whatthedude.com/research/2015/Benchmarking%20Elasticsearch%201%20Million%20Writes%20per%20Sec.pdf>
- [4] <http://whatthedude.com/research/2012/in-depth-data-stacking.pdf>
- [5] <https://www.fireeye.com/blog/threat-research/2012/11/indepth-data-stacking.html>

Look Based Media Player

Atulkumar Mishra

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
atulkumar.v.mishra@gmail.com*

Abhijit Chaudhary

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
abhijitchaudhary95@gmail.com*

Rahul Vishwakarma

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
vishwakarma.rahal05@gmail.com*

Vandana Munde

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India
vrsakharkar@gmail.com*

Abstract: In this project we are developing an advanced media player which plays and pauses the video by detecting the user's face looking at screen or not. System monitors whether the user is looking at the screen or not using a web camera. If yes then it doesn't interrupt the video and allows it to play. In case if the user is not looking at the screen or say the system couldn't detect the user's face then it immediately stops the video. This system can also be used as a prototype for developing other more powerful systems. Moreover this system also has multiple other features such as controlling other aspects of a media player like volume up and volume down, play, pause and stop the video using hand gestures. We are trying to integrate both the systems together and make it one system as a whole.

Keywords-Media Player;Hand Gestures Haar Cascade;Sensors.

I. INTRODUCTION

Usually when users are watching a video and someone calls, the user has to look somewhere else or go away from the PC for some time so the user misses some part of the video. Later the user needs to drag back the video from where the user saw it. Well here is a solution to this problem. A look based media player that pauses itself when the user is not looking at it. The player starts running again as soon as the user looks at it again. This is done using the camera or web camera on top of the computer. As long as the camera detects the user's face looking at it, the media is played. The player pauses as soon as the user's face is not completely seen. Even this system can be used as a prototype for developing other systems which can be used for security purposes.

This system also provides the feature of controlling other functions of media players such as volume up, volume down, forward and backward using hand gestures.

II. AIM AND GOAL OF THE SYSTEM

The goal of our project is to create an advanced media player based on face detection and hand gestures recognition. We have set the following objectives for our media player to achieve the goal:

1. The user interface of the media player should be efficient and user friendly.
2. The media player should be accurate in terms of result.
3. The media player should pause the video as soon as the user's face is not detected.
4. The hand gestures should be captured accurately and actions associated to them should be performed perfectly.
5. We are trying to make this project with an aim for 10.

III. CONCEPT AND EVALUATION

A. Existing System:

Mostly existing systems use eye recognition. Due to which results aren't accurate. One of the previous systems tried to induce both face detection and hand gesture recognition

system in a single media player but the results weren't published.

B. Proposed System:

In this project we are using face detection and hand gestures recognition for controlling media player. Face

detection is used for pausing and playing. Various hand gestures are used for controlling other functions of media player.

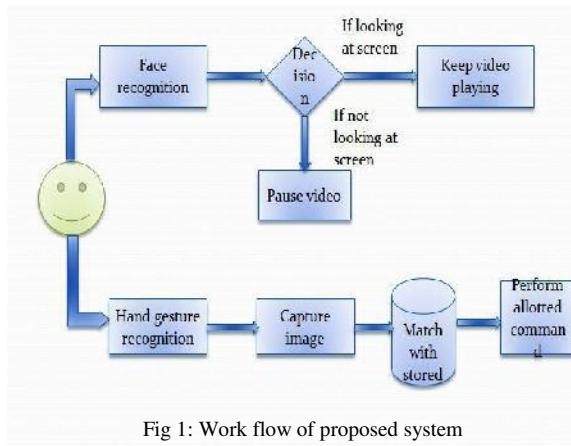


Fig 1: Work flow of proposed system

IV. IMPLEMENTATION METHOD

The proposed system uses Haar Cascade Classifier algorithm [1] HAAR CASCADE CLASSIFIERS:

In the Viola-Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier. The key advantage of a Haar-like feature over most other features is its calculation speed. Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in first real-time face detector. In the detection phase of the Viola-Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola-Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature).

Open CV's algorithm is currently using the following Haar-like features which are the input to the basic classifiers:

1. Feature = $w_1 \times \text{RecSum}(r1) + w_2 \times \text{RecSum}(r2)$
2. Weights can be positive or negative .

3. Weights are directly proportional to the area.
4. Calculated at every point and scale.

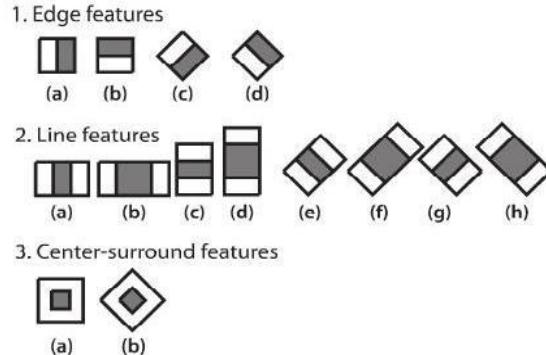


Fig 2: Haar Cascade Classifier

[2] CONTOUR EXTRACTION:

Contour tracing is a technique that is applied to digital images in order to extract their boundary. A digital image is a group of pixels on a square tessellation each having a certain value. This method will be dealing with bi-level images i.e. each pixel can have one of 2 possible values namely: 1, in which case we'll consider it a "black" pixel and it will be part of the pattern, OR 0, in which case we'll consider it a "white" pixel and it will be part of the background. Contour tracing is a technique that is applied to digital images in order to extract their boundary.

A digital image is a group of pixels on a square tessellation each having a certain value. This method will be dealing with bi-level images i.e. each pixel can have one of 2 possible values namely: 1, in which case we'll consider it a "black" pixel and it will be part of the pattern, OR 0, in which case we'll consider it a "white" pixel and it will be part of the background.

Well, the contour pixels are generally a small subset of the total number of pixels representing a pattern. Therefore, the amount of computation is greatly reduced when we run feature extracting algorithms on the contour instead of on the whole pattern. Since the contour shares a lot of features with the original pattern, the feature extraction process becomes much more efficient when performed on the contour rather than on the original pattern.

INPUT	TP	TN	FP	FN
Female	10	0	0	10
Male	10	0	0	10
Kids	10	1	0	10
Others	10	0	0	10
Avg.	10	0.25	0	10

V. RESULT

1. Accuracy: - The degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard for image processing.
2. Processing Time:- The period during which one or more inputs are transformed into a finished output

Table I: Accuracy For Face detection

AVG. PROCESSING TIME:

Table II: Average Processing Time

INPUT	AVG. PROCESSING TIME
Female	6s
Male	6s
Kids	5s
Others	6s
Total Average	5s

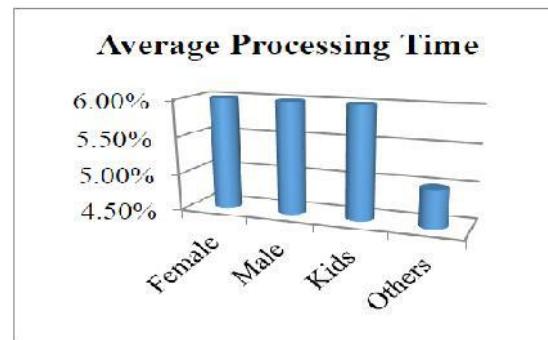


Fig 3: Contour Extraction

ACCURACY for Face detection :

Total Accuracy: - $(TP+TN)/ (TN+TP+FN+FP)$

$$=(10+0.25)/ (10+0.25+0+10)$$

$$=98\%$$

ACCURACY for Hand Gesture)

Table III: Accuracy for Hand Gesture

INPUT	TP	TN	FP	FN
Volume UP	9	10	0	0
Volume Down	10	9	1	0
Play	9	9	0	0
Pause	10	9	0	0
Avg.	9.5	9.25	1	0

:

$$\text{Total Accuracy: } - \frac{(TP+TN)}{(TN+TP+FN+FP)}$$

$$= \frac{(9.5+9.25)}{(9.5+9.25+1+0)}$$

$$= 94\%$$

BAR GRAPH:

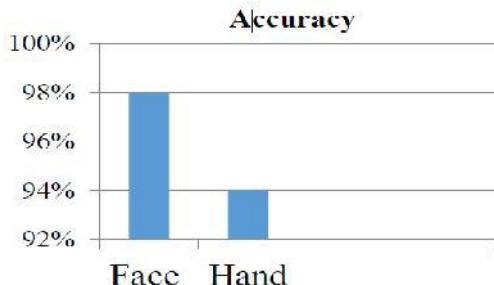


Fig 4: bar graph:

VI. CONCLUSION

The main concern of this project is to help the user get best experience of using a media player. We have tried to achieve this goal by automating the media player in a wide extent. We are doing this by using face recognition and hand gestures for controlling varied features of the media player such as pausing and starting the video again when the user isn't looking at the screen(for which face recognition is used), and controlling functions as forwarding , backwarding, volume up and volume down(for which hand gestures are used).

REFERENCES

- [1] Sidharth Rautarayet Anupam Agrawal, "A Vision Based Hand Gesture Interface for Controlling VLC Media Player", Journ. International Journal Of Computer Applications(0975-8887), vol. 10-No.7, November 2010.
- [2] Harsha Jadhav,Sabiha Pathan,Neha Rokadeet Uma Annamalai "Controlling Multimedia Applications Using Hand Gesture Recognition", Journ, International Research Journal Of Engineering and Technology (IRJET), vol. 02, PP 1200-1203, August-2014.
- [3] Jyoti Ranjet Kanwal Garg, "Emotion Detection Using Facial Expression", Journ. International Journal Of Advanced Research & Software Engineering, Vol. 4, PP-465467, April 2014
- [4] Xiaoming Liu et Tsuhan Chen "Video-Based Face Recognition using Adaptive Hidden Markov Model", Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, U.S.A.
- [5] D Jadhavet Prof L.M.R.J Lobo "Hand Gesture Recognition System To Control Slide Show Navigation", Journ. International Journal Of Applications or Innovation in Engineering & Management (IJAIEM)", vol. 3, PP-283-286, January 2014.
- [6] N. Krishna Chaitanya et R. Janardan Rao "Controlling OF Windows Media Player Using Hand Recognition System", Journ. The International Journal Of Engineering And Science (IJES), vol. 3, PP 01-04, 2014.
- [7] Viola and Jones, "Rapid object detection using a boosted cascade of simple features", Computer Vision and Pattern Recognition, 2001.

RFID Based Electronic Payment System

Shreya Mistry

Department of Information and Technology

Thakur College of Engineering and Technology

Mumbai, India

shreyamistry72@gmail.com

Nishita Desai

Department of Information and Technology

Thakur College of Engineering and Technology

Mumbai, India

nishita.tcet@gmail.com

Tanay Sanghavi

Department of Information and Technology

Thakur College of Engineering and Technology

Mumbai India

tanaysanghavi@gmail.com

Abstract-The project aims at creating a cashless payment system which will function properly in the college campus. It will handle all the transactions in all the parts of the college campus. This project will require proper working software as well as hardware. Connecting the software with the hardware is a challenging task. The RFID reader will have to accurately read the RFID card and update all the transactions and information in the system. However, it is essential to have accurate and proper transaction updates in the software. Existing systems have used algorithms like AES, which are light-weighted. This project will be implemented using encryption algorithm present in SQL and also by writing a short algorithm of our own. This will ensure that the data stored inside the database will have proper encryption. It will also ensure that the data cannot be accessed or interpreted by any third person trying to get into the system.

The project requires the students to carry individual RFID cards. On the campus, at the time of any kind of transaction, the card will have to be placed on the RFID reader, which will be set up at various points in the college. The RFID card will have certain amount stored in it. Each time the card is placed over the reader, the reader will make changes in the data of the software. Once the changes are made inside the software, the transaction will be completed. Accordingly, the amount will either be deducted or added into the individual account of the student.

The RFID reader and card provide efficiency and ensure that all the transactions are carried out in a hassle-free manner. The system aims at providing a complete cash-free environment all over the college campus. Thus, helping in better management, keeping a record and conducting complete transactions. There by, helping in increasing the overall efficiency of all transactions made inside the college campus.

Keyword: *RFID;AES*

danger of a student losing his wallet. However, these issues can be greatly managed by means of electronic payment. Electronic transactions are transactions which are made without the use of paper. Our project aims at providing an individual, unique card to each and every student. This card can be used to make payments in various areas of the college campus, using RFID technology. This can greatly help in Overcoming the problems faced by using hard cash.

II. PROPOSED WORK

This project presented a survey on RFID technology. RFID technology has a big potential to become ubiquitous in the near future. Today it is already successfully used in supply chain management to track pallets of items. Tracking allows better coordination and control in the production cycle. Now the industry is pushing towards item-level tagging to increase the control even further. However, that also creates concerns, most common privacy concern, but also other security related issues. The paper presented possible scenarios how privacy can be compromised by RFID tags but also several solutions to protect against it. Since RFID technology becomes more and more common, attacks against the system itself start to appear. This paper listed the most common, starting from common sniffing and eavesdropping over denial of service to new RFID viruses.

The project concludes by looking at some social implications that RFID causes. Although not technically relevant, it provides a good outside perspective.

I. INTRODUCTION

There have always been a lot of issues in transactions involving hard cash. In the college campus itself, there are numerous situations where the student needs to pay certain amount of money. The need to carry exact change is essential. Apart from this, there is always a

III.LITERATURE SURVEY

Table I: LITERATURE SURVEY

Reference Number	Year	Author Name	Topic of the paper reviewed	Key results and discussion	Research gaps
[1]	2014	Ahmed Alzahrani	Security analysis of RFID based devices in educative environments	Security provided: 42.75% Expected security 51.75% RFID tag power: - - 10dB i.e. 100microWatts	Security provided: 42.75% Expected security 61% RFID tag power: - less than 50microWatts
[2]	2013	ChristophJ echlitschek	A Survey paper on Radio Frequency Identification(RFID) trends	High frequency RFID chips. Authentication: Light weighted algorithm AES	Low frequency RFID chips. Authentication: Heavy weighted algorithm in SQL
3]	2015	B Muniraja, JRajanikant h	In-Time Billing Process for Canteen Management System	Restricted to canteen. Recharge Limit Rs200-2000	Covers entire college campus. Recharge Limit Rs1000
[4]	2016	PrachiChaudhari, SayaliRane, Nikita Ahire , HarshaliShinde , Prof.A.G.Patil	Review on- RFID based Campus Management System	Makes use of web service. UHF/HF cards used. Frequency range of HF/UHF cards 125 or 134KHz	Web service not used. LF cards used. Frequency range of LF cards 13.56 MHz
[5]	2012	Bichlein Hoang Ashley Caudill	RFID	Active RFID cards Proximity: 100 metes Frequency range 455MHz,2.45GHz, or 5.8GHz Cost Rs200 per card.	Passive RFID cards Proximity: 10 cm Frequency range 128KHz, 13.6MHz, 915MHz or 2.45GHz Cost Rs75 per card.

III. PROBLEM DEFINITION

Phase1:

Planning: The entire flow will be decided.

Requirement Gathering: The project will be planned and requirements will be enlisted

Designing: The front end will be ready, the design of all the phase will be done.

Phase2:

Coding: The code will be written and entire software code will be targeted to complete.

Encryption: DES algorithm will be used.

Phase 3:

Coding: The backend database will be created and connection will be established between front end and backend

Hardware Gathering: The hardware required will be procured

Phase 4:

Hardware-Software Compatibility: the hardware will be set up according to the software. Connection between them will be established

Testing: The connection will be tested and changes will be made accordingly

Phase 5:

The entire project will be tested and further required changes will be done .

IV. METHODOLOGY USED

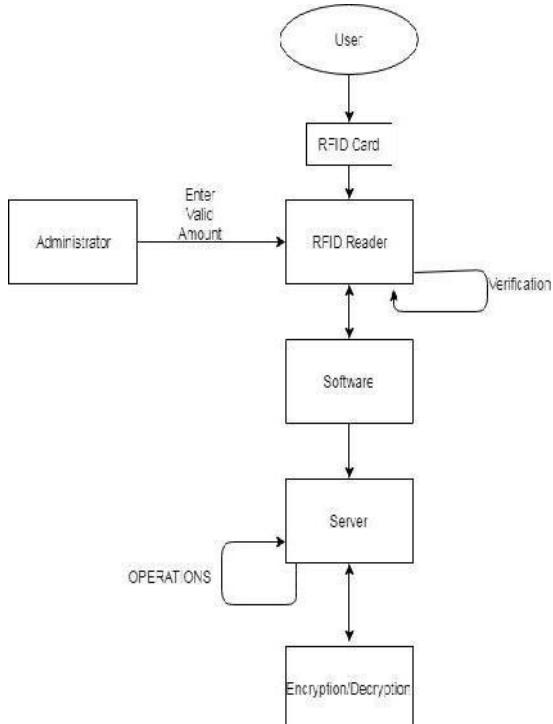


Fig 1:Methodology

V. FEASIBILITY

RFID based Electronic Payment System:

Operational Feasibility:

There is no hardcoded input like mathematical formula and complex functions. The RFID reader will read the card which has been placed on it and will authenticate the student. If the student is authenticated, the amount entered will be deducted from his account. It will in all make transactions quicker and hassle free.

Economic Feasibility:

The program can be executed on computer of normal configuration of at least 4 GB ram. The project can be executed independently of other programs or modules and there is no need to install any extra dependent modules or programs to run this project without facing overhead or load on System.

Technical Feasibility:

All the complex algorithms and techniques are hidden from user and all input/outputs can be executed without much technical knowledge. All the algorithms and techniques are optimized to yield best possible results.

Legal Feasibility:

All the programs are developed using either open source software or freely available framework without much licensing issue. All the Framework used to develop allows us to use application developed to use and integrate anywhere free of cost.

Social Feasibility:

This Project can be helpful for implementing the ‘cashless’ system in the college campus. This Project will be useful to the students to perform hassle-free transactions on the campus of the college.

VI. OUTCOMES

Phase1:

Planning: The entire flow will be decided.

Requirement Gathering: The project will be planned and requirements will be enlisted

Designing: The front end will be ready, the design of all the phase will be done.

Phase2:

Coding: The code will be written and entire software code will be targeted to complete

Encryption: DES algorithm will be used .

Phase3:

Coding: The backend database will be created and connection will be established between front end and backend

Hardware Gathering: The hardware required will BE

procured

Phase 4:

Hardware-Software Compatibility: the hardware will be set up according to the software. Connection between them will be established

Testing: The connection will be tested and changes will be made accordingly

Phase5:

The entire project will be tested and further required changes will be done.

IX. REFERENCES

- 1.Life Science Journal 2014;11(1) Department of Computer Science, Faculty of Computing and Information Technology,
2. King Abdulaziz University, Jeddah 21589, Saudi Arabia.
- 3.International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 200- 203 (2015)
4. International Journal of Modern Trends in Engineering and Research
5. RFID: Technology and Applications by Sridhar Iyer
6. www.impinj.com www.zebra.com
- 7.barcodesinc.com

VII. IMPROVEMENT IN RESEARCH GAPS

The real challenge in maintaining such payment systems is maintaining security in them. The research papers use AES algorithm, where as we use DES algorithm. Therefore, the security results will be different than the other projects.

Along with the algorithm, we will also be implementing a small encryption algorithm of our own in order to increase the overall security of the system.

The previous system could achieve security up to 51.75%, whereas, our system aims at achieving 61% security

Table II: Research Gap

Research Gaps	Improvements in Research Gaps
Security provided: 42.75% Expected security 51.75% RFID tag power: - - 10dB i.e. 100microWatts	Security provided: 42.75% Expected security 61% RFID tag power: - less than 50microWatts
High frequency RFID chips. Authentication: Light weighted algorithm AES	Low frequency RFID chips. Authentication: Heavy weighted algorithm in SQL
Restricted to canteen. Recharge Limit Rs200-2000	Covers entire college campus. Recharge Limit Rs1000
Makes use of web service. UHF/HF cards used. Frequency range of HF/UHF cards 125 or 134KHz	Web service not used. LF cards used. Frequency range of LF cards 13.56 MHz
Active RFID cards Proximity: 100 metres Frequency range 455MHz,2.45GHz, or 5.8GHz Cost Rs200 per card.	Passive RFID cards Proximity: 10 cm Frequency range 128KHz, 13.6MHz, 915MHz or 2.45GHz Cost Rs75 per card.

Result Analysis and Prediction

Hetal Amrutia

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Siddhant Acharya

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Ujjwal Sharma

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Sudhesh Singhani

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Abstract - Website for result analysis provides a way of solving the problems faced by Institutes by saving time and effort of analyzing the data. The objective of the result analytics system is to make analysis easier to understand, monitor, track and resolve, and to provide the institutes with an effective tool to identify and target problem areas, monitor records, handling performance and make improvements in the manual process. Result Analytics is a management technique for assessing, analyzing and responding to institute data. This software is used to record data and respond to changes taking place in them timely, as well as provide an outlook of business opportunity to Institutes. Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events.

Keywords: Analysis; Machine Learning, Predictive and Result, etc

I. INTRODUCTION

Website for result analysis provides a way of solving the problems faced by Institutes by saving time and effort of analysing the data. The objective of the result analytics system is to make analysis easier to understand, monitor, track and resolve, and to provide the institutes with an effective tool to identify and target problem areas, monitor records, handling performance and make improvements in the manual process. Result Analytics is a management technique for assessing, analysing and responding to hospital data. This software is used to record data and respond to changes taking place in them timely, as well as provide an outlook of business opportunity to Institutes.

II. OVERVIEW

This software is primarily aimed for schools and colleges to evaluate the performances of both students and teachers. The system can generate performance reports such as progress reports, merit lists, etc. for both students and teachers. In the manual student result processing system, all the database commands have to be typed by the users. This procedure is very time consuming, and is limited to a single system. If someone wants to get information about a particular subject score, teacher's Performance or any other information, they have to contact the administrator handling the system.

Project Description:

It is a website for analyzing the results of students from the given database .It not only analyzes the overall result but also provides an overview of the results of each subject separately along with the pictorial representation in the form of index graph .By this the students and their parents can map the progress of their kids ,This website not only provides student centric analysis but also teacher centric results i.e it helps examine the progress graph of each student under given teacher which in turn helps evaluate the performance of the teacher as well

Lastly it also provides the option of feedback wherein the students can give anonymous feedback to a teacher.

In the current system all the activities are done manually. It is very time consuming and costly. Our Student Result Analysis System deals with the various activities related to the students. During analysis, data collected on the various files, decision points and transactions handled by the present system. The commonly used tools in the system are Data Flow Diagram, interviews, etc.

III. PROBLEM DEFINITION

A lot of teaching organizations today face a problem today of evaluating a student based on his/her performance in the test .This website and mobile application eliminates this issue by providing instant analysis of a student's performance based on their marks which helps the organization evaluate the student better

Many organization do possess such soft wares that are used to evaluate the student but fail to evaluate the teachers .This website bridges the gaps of all the previous proposed software like teacher centric analysis , student feedback and the quality of evaluation.

Phase I: (Planning, Analysis, Design, Coding, Implementation)

Perform literature survey and identify all the research gaps, learn about the functioning of the institutes by observing the teachers and students and the work done by the staff .

Prepare the snowflake schema design for integration of the table. Integration of data from all the tables of the database into a data warehouse.

Create code for changing the loading of data from SQL tables from full load to incremental load.

Code to create a software that fills all the research gaps such as teacher centric analysis of data, student feedback, and improving the quality of result.

Phase II: (Testing, Deployment)

Create various cases of test case data to check if the integrated data warehouse functions as desired by the client.

Conduct beta testing for identifying any further errors and the identify and improvements and that can be performed.

After testing and approval, deploy the proposed system to the hospital

FEATURES OF THE PROJECT

IV. METHODOLOGY

The Proposed methodology is illustrated in Figure. It contains following steps:

Data Extraction

Data Preprocessing

Data Integration and Transformation

Feature Selection

Classification

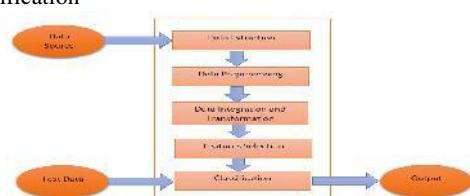


Fig 1:Methodology

Random Forest:

In bagging (Bootstrap Aggregation), numerous replicates of the original dataset are created to reduce the variance in prediction. Random Forest use multiple decision trees which are built on separate sets of examples drawn from the dataset. In each tree, we can use a subset of all the features we have. By using more decision trees and averaging the result, the variance of the model can be greatly lowered.

Decision Tree:

decision tree is commonly used in classification in machine learning, it is a tree-like graph or model to make decisions and get consequent results. Decision tree implicitly perform variable screening or feature selection and the best feature of using it is easy to interpret and explain!

Naïve Bayes:

Naive Bayes classifier is a probabilistic classifier based on Bayes' theorem. It depends on the assumption that some features are conditionally independent. It is super simple and can converge quicker than some other models, so it has a chance to achieve fast and easy process.

V. FEASIBILITY STUDY

Technical Feasibility – The application will operate on Web Browser. The system will be implemented using python, and web technology as required. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms which supports python.

Economic Feasibility– Online website is used to analyse the results of students and at the same time to predictive analyses using machine learning .Due to this mapping and improving the results of students will be become easy and improve the overall result of institute and attract more people thus making this website economically feasible .

Legal Feasibility-The project is developing under legal license of windows, python and other software's required. The data will not be used for some illegal purpose and will be used for making the news relevant to user as well as for publisher for making under certain constraint. Data will not be misused or misguided under any circumstance.

Operational Feasibility -An overview of the process portrayed that the proposed Machine Learning for

Result Prediction be beneficial for news publisher to monitor the growth and improvement in the future result. Project Administration

Communications Plan:- The communication among the team members as well as the client and developers is very important. Proper communication is a must for effective development of the project. The Project manager is at the apex of the communication tree. He should effectively

communicate with the Project administrator who in turn communicates with the programmers as well as the testers. If effective communication fails, the end result would be a total chaos.

Scope Management Plan: -The scope of the software project for University is described above. The Project Administrator takes appropriate care that the project doesn't go out of the scope so that the timeframe does not get disturbed as it would cause issues. It is an important to see that nothing outside of the scope be touched upon.

Quality Management Plan: -Quality management can be considered to have four main components: quality planning, quality control, quality assurance and quality improvement. Quality management is focused not only on product/service quality, but also the means to achieve it. Again it is up to the Project Administrator to maintain high quality so that it maintains the customer satisfaction.

Change Management Plan: -Changes in any software project are an integral part of it. The changes are caused due to the change in requirements of the University system. Whenever any change is arrived the project administrator understands the C.R. first and will set priority to the same. Next he would assign the change request to any of the developer who after developing the same would give it to the tester. Just in case the changes are too many then project manager would coordinate with the University members.

VI. USE CASE DIAGRAM

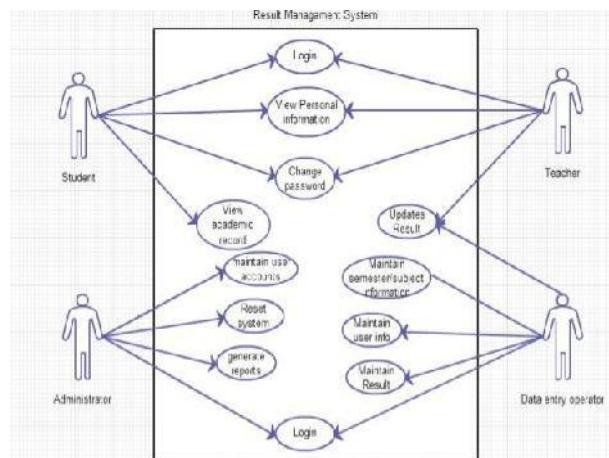


Fig 2: Use Case

DFD DIAGRAM

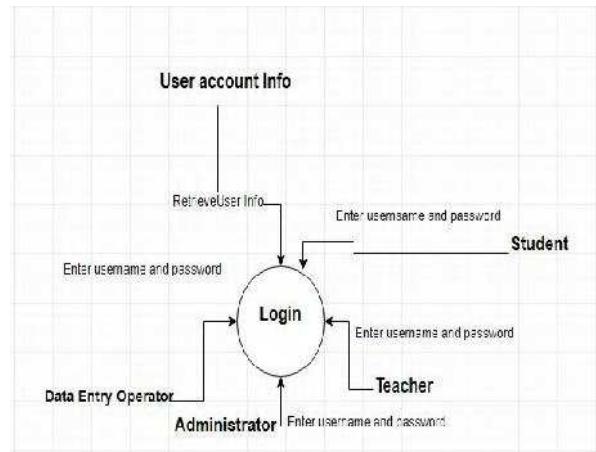


Fig 3: DFD LEVEL 0

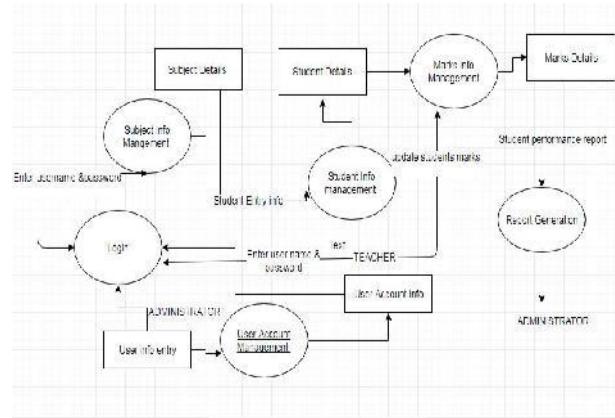


Fig 4: DFD Level 1

After implementing the following algorithms like decision tree, Knn, random forest, naïve bayes, it is being observed that random forest achieves better accuracy.

a. Outputs

Performance analysis and results:

Accuracy:

TABLE I: ACCURACY

Algorithm	Decision Tree	KNN	Random	Naïve Bayes
Accuracy (Original)	57.9%	57%	66.4%	61.2%
Accuracy (Cross Validation)	55.5%	54.9%	64.4%	60.8%

TABLE II: SPEED

Algorithm	Decision Tree	KNN	Random	Naïve Bayes
Time elapsed (Original)	1.2s	3.23s	6.33s	0.07s
Time elapsed (Cross Validation)	8.44s	10.83s	37.71s	0.48s

From two tables above, it is observed Random Forest model achieves the highest accuracy (66.4%), but it consuming the most time (6.63s) among 4 models. Naive Bayes model achieves the least time elapsed(0.07s) and a properly medium accuracy. The other two models have bad accuracy. The kfold cross validation has proved that the accuracy should be less than that gained from the original method. But considering its big time consuming, the difference can be ignored relatively under the purpose of comparing these 4 models.

VII. RESULT

Through this we hope to achieve faster analysis of result for any institute which will in turn improve the overall evaluation process of the institute as well help the students and the teachers improve their result on the basis of the evaluation.

Phase I:

1. Facts and Dimension Tables in the Snowflake schema should be properly linked and there should be no confusion in their respective orders.
2. Integration of data from all the tables into one big data warehouse without data duplication or redundancy and in the specified format.
3. The loading of data from SQL tables takes place incrementally.
4. The software so developed will be able to bridge all the gaps in the previous proposed solutions .

Phase II:

1. The integrated data warehouse functions as desired by the client.
 2. The proposed system will be tested by the user giving us hands on feedback from the user .
 3. The proposed system meets all the requirements of the institute and serves them efficiently.
- From the above discussion it is clear that website for result analysis and result prediction is very useful and helps the organization to improve the result of the student as well as of the organization.

VIII. FUTURE SCOPE

The IT department using our software would be able to access precise data of all the related results of any particular institution and use it intelligently to help all the institutions in maintaining the database efficiently. This software would be useful for any institutions having related branches as it does the work of connecting the data objects of many different databases together into a centralized database and keep record of all the data objects that are growing with time and provide many business opportunities.

Dashboard can be used by any organization so as to keep track of their data time to time. In today's time data is said to be the biggest asset for any company or organization. Thus, it becomes of immense importance to organize the storage of data and keep a track of it. With proper and timely tracking of this data a business organization can find ample of growth opportunities and expand their business.

IX. CONCLUSION

Our main aim through this project is to reduce manual work, to provide efficient way of handling data. Our Project reduces the work of staff for analyzing the result. Result Analysis generates graph of students marks and time table generation is done automatically. In future, our project will be done as a mobile application.

This website will also provide the prediction of future results based on the previous results. Also, it has additional features to improve individual performance by adding remarks too.

X. ACKNOWLEDGEMENT

We sincerely thank to our guide Mrs.HetalAmrutia our HOD Dr. Rajesh S. Bansode, our Dean Dr. Kamal Shah and our principle Dr. B. K. Mishra for his/her guidance and support for carrying out our project work

REFERENCES

- [1]. C. Romero S. Ventura P. G. Espejo and C. Hervás "Data mining algorithms to classify students ", EDM 2008 pp. 8-17.
- [2].Miss.GhadgeVrushali S "Result analysis and schedule generation." , EDM 2016 pp.9-10.
- [3].Helena Borožová, Jan Rydval "Analysis of students results in distance –studies centres.", ERIES Journal vol. 7 no. 3-4

An Approach to Maintain Attendance Using FTP Tools

Saurav Jangir

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Aakash Tripathi

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Mahendrasingh Parmar

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Purvi Sankhe

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

I. INTRODUCTION

Abstract - Attendance management is a difficult and essential part of every organization especially for those organizations that are not using finger print readers or automatic scanners to manage attendance. There are lots of attendance management systems available in the market but most of them do not provide the flexibility to modify and change details at random situations, which can sometimes be the need of time. This work is useful for easy user interface. The planning is to utilize the powerful database management, data retrieval and data manipulation, Ftp, Android Application. The System will provide more ease for managing the data than manually maintaining in the documents. This work is useful for saving valuable time and reduces the huge paper work. Also, Digital Empowerment to the Students in keeping view of Digital India. The main objective of this System is to check and prepare report or attendance for any organizations such as for hospitals, schools, colleges, shopping mall etc. With the help of attendance management system, it will eliminate the manual calculation of attending days for each student, its admin will easily able to analyze the result based on their requirement such as number of absentees per day, number of student required to maintain their attendance etc. To remove the process of taking attendance in register sheet, attendance management system uses the concept of Android application to record the attendance through router. There are custom search capabilities to aid in finding student information and working on student records. This can make the system easier to navigate and to use maximizing the effectiveness of time and other resources.

Keywords- component; Android, Router, Attendance management, Smartphone, and FTP etc.

“An approach to maintain attendance using FTP tools” is software developed for maintaining the attendance of the student on the daily basis in the collage. Here the staffs, who are handling the subjects, will be responsible to mark the attendance of the students.

Each staff will be given with a separate user-name and password based on the subject they handle. An accurate report based on the student attendance is generated here. This system will also help in evaluating attendance eligibility criteria of a student. Report of the student’s attendance on weekly and monthly basis is generated. Student attendance management system deals with the maintenance of the student’s attendance details. It generates the attendance of the student on basis of presence in class. It is maintained on the daily basis of their attendance. The staffs will be provided with the separate user-name & password to make the student’s status. The staffs handling the subjects are responsible to make the attendance for all students. Only if the student present on that period, the attendance will be calculated.

The student’s attendance reports based on weekly and consolidate will be generated

This software is primarily aimed for schools and colleges to gather the attendance using router based system and evaluate the effective attendance of student. The system can generate Attendance reports such as monthly attendance, Subject-Wise attendance using database.

Automated Time and Attendance marking system can help schools and higher education in many ways. There is no doubt that an attendance management system will help save time and money by eliminating a great deal of manual processes involved in attendance and leave entry and calculating hours attended. With automatic class attendance system, teachers can more accurately and quickly track student’s time on the classroom.

A.)Manual attendance system [2]

It is the conventional method of taking attendance by calling names or signing on paper but it is inefficient due to more chances of malfunctioning and more paper work as well.

B.)RFID with Object Counter [2]

Radio Frequency Identification (RFID) based attendance system is one of the solutions to address this problem, but that is time consuming and unsafe. Anyone can carry others card to mark proxy attendance.

C.)Bluetooth Based Attendance System [2]

In this, attendance is being taken using instructor's mobile phone. Application software is installed in instructor's mobile telephone, enables it to query student's mobile via Bluetooth. It transfers student's mobile Media Access Control (MAC) addresses to the instructor's mobile phone and presence of the student can be confirmed. The problem of this proposed system is student's phone is required for attendance. In case of absent student if his mobile is given to his friend and if kept it in coverage area then also his presence would be marked. All the above systems are time consuming and unsafe. In the proposed project Android based attendance system is designed which is less time consuming, safe and easy to implement.

PROPOSED SYSTEM

In the proposed system, we will use a router/modem as a medium and through this medium we will establish a connection between student and teacher using FTP (SSH protocol) and fetch the students present in the class. The system will broadcast an android application which only the authenticated students will be able to install in their devices. After installation, the students need to give their attendance in the application and this attendance will be recorded in the database. The proposed system has a special feature which will allow the application to automatically uninstall itself from the student device. This uninstallation will be triggered only when the students submit their attendance by clicking on the submit button of the application.

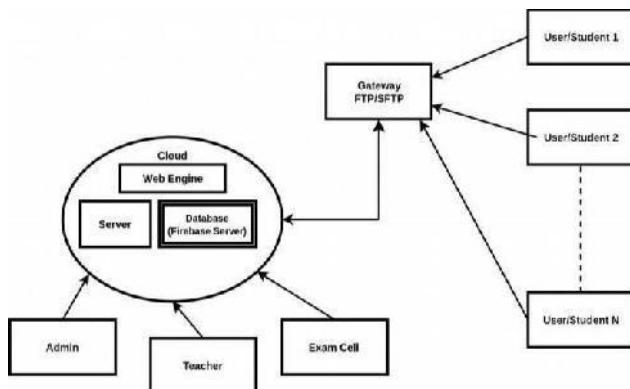
II. METHODOLOGY

Fig 1:Block Diagram

Step-1: Setting up Router or portable modem (WiFi).

1. Installing Router in the classroom. Portable WiFi device can also be used (cost effective).
2. Configure router/WiFi settings such that each corner of the classroom is covered by the signal.

Step-2: Installing and Configuring the FTP tool (One time process).

1. Install OpenSSH (Free) for Linux based system.
2. For windows Based system install FileZilla (Free) or WinSCP (Free) tool.
3. Configure the FTP tools making sure that the system is secured by setting up the credentials.

Step-3: Connect to the Router/Portable WiFi device

1. Connect to the connection provided.
2. Connect to the FTP connection by logging in (System ID, Password provided by the teacher).
3. After logging in, installing the Android application available for the student (the student will not be able to read/write other files and folders after logging into the system as the files/folders will be locked for them).

Step-4: Report your attendance

1. After installation, the student need to provide the following information:

Class-Division
Date
Roll No.

2. After required fields are filled the student need to submit his record by clicking the SUBMIT button.

3. As soon as the student clicks the SUBMIT button the

application will automatically ask the Student to Uninstall the application. If the student clicks 'Yes', then only his response will be recorded otherwise his response will not be recorded.

HARDWARE REQUIREMENTS

- 1) Personal desktop/Laptop -Central server with processing engine.
- 2) The minimum memory size required is 1GB.
- 3) Smartphone – Offers more advanced computing ability and connectivity.

III. GUI DESIGN/RESULTS (SCREEN SHOTS)

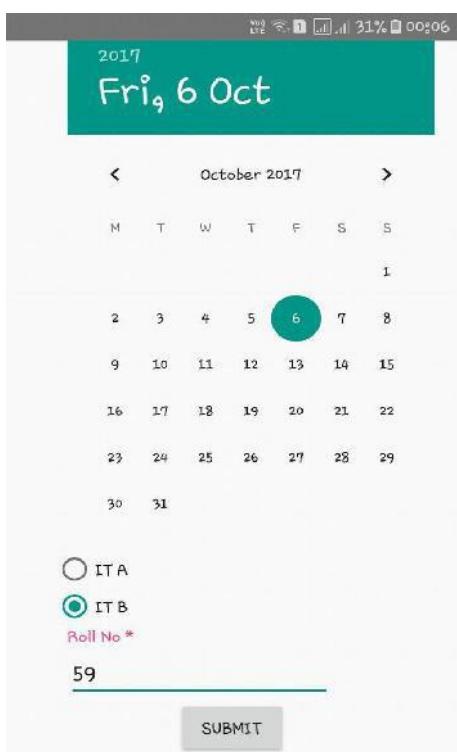


Fig.1:Application Homepage

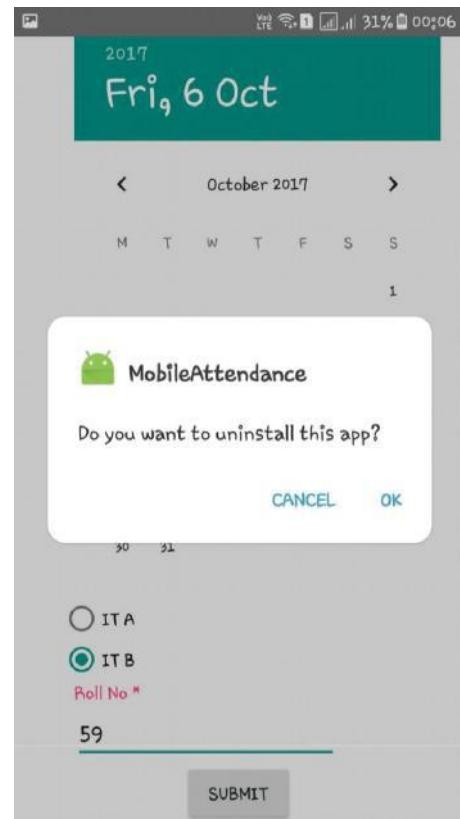


Fig.3:Application for FTP connection with the system.

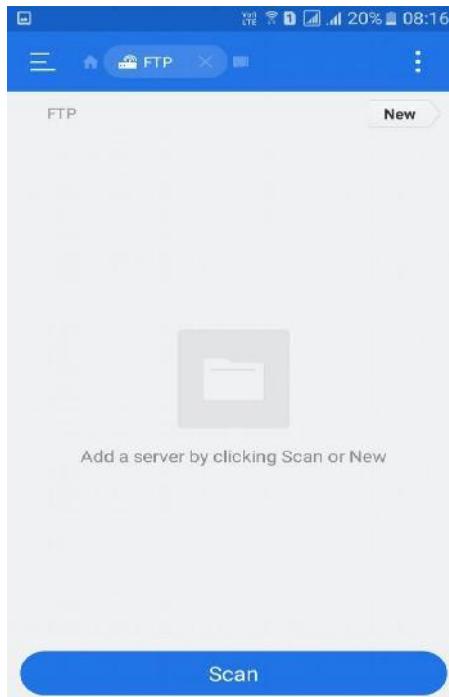


Fig.2:Attendance Recording

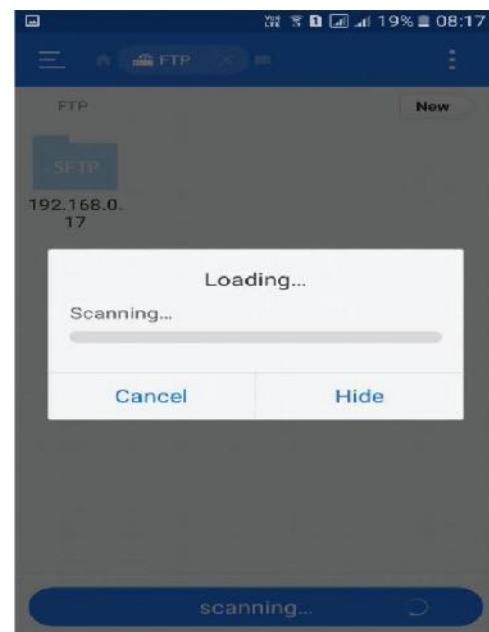




Fig.4.Search for FTP connection established

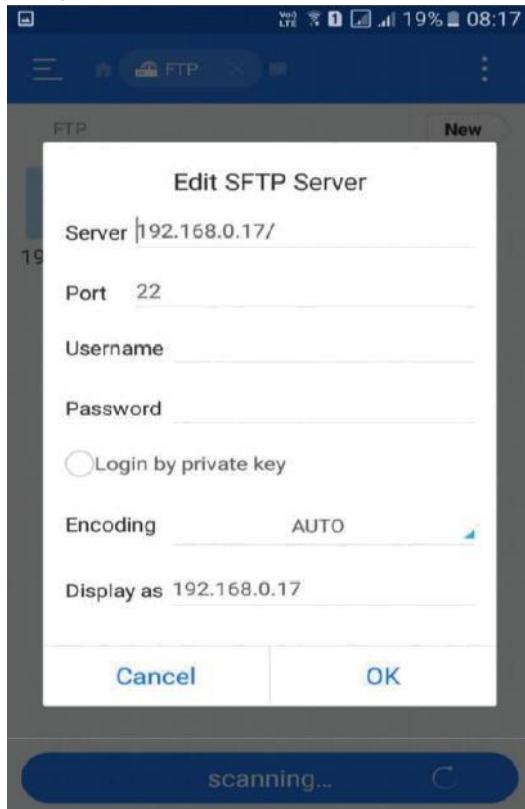
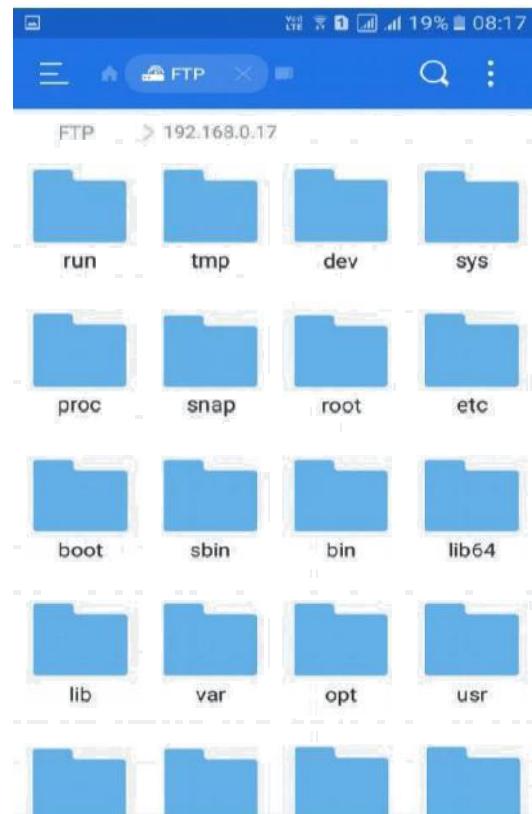
Fig.5: Successful scanning
(Admin detected by the user).

Fig.6:Authentication Successful

IV. CONCLUSION

The Attendance Management System is developed using Android Studio, Computer Network, and Router. It fully meets the objectives of the system for which it has been developed. The system will reach a steady state where all bugs would have been eliminated as per the planning and Deployment date. The system will be operated at a high level of efficiency and all the lecturers and users associated with the system understands it. From the tests performed on the new system/design it will improve on the overall performance on the management of student's attendance, thus the system solves the problem that it was intended to solve. Furthermore, computerization of attendance system is not just a matter of technological innovation and development; it is a process which involves individual, organization and society in general. The computerization of the attendance system has brought about faster access and easier method of viewing, recording, editing, printing (documentation) of student attendance record

V. FUTURE SCOPE

The project has a vast scope in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. The following are the future scope for the project:

- Discontinue of student eliminate potential attendance.
- Bar code Reader based attendance system.

VI. ACKNOWLEDGMENT

We have taken up the implementation of this System. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them. We are highly indebted of Mrs. Purvi Sankhe for her guidance and constant supervision as well as for providing necessary information regarding the project and for her support in completing the project. We would also thank the college for providing the necessary facilities that was used by us during our project. We would like to express our deepest gratitude to our HOD Dr. Rajesh Bansode who managed to help us despite his busy schedule and took out time to provide guidance and inputs which were helpful in building this project. We would also like to thank our parents who gave us the moral support which motivated us to work even harder. Special thanks go to our classmates who have extended their help in accomplishing this undertaking.

VII. REFERENCES

- [1] V. Somasundaram*, M. Kannan and V. Sriram Indian Journal of Science and Technology, Vol 9(35), DOI: 10.17485/ijst/2016/v9i35/101807, September 2016.
- [2] Rakhi Joshi1, V. V. Shete2, S. B. Somani3 International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015.
- [3] <http://www.w3schools.com/>

Mathematical Morphology Based Algorithm for Image Noise Reduction.

Divya Sharma

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Arpit Vakharia

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Juhi Vadia

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Bijith Marakarkandy

*Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India*

Abstract –

There are various types of noises present in an image. Noise such as Gaussian noise (Principal sources of Gaussian noise in digital images arise during acquisition e.g. sensor noise caused by poor illumination and/or high temperature, and/or transmission e.g. electronic circuit noise), Salt and Pepper noise (An image containing salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions), Speckle noise (Speckle is a granular noise that inherently exists in and degrades the quality of the active radar, synthetic aperture radar, medical ultrasound images). These noises in the images are removed using various technologies and algorithms such as kFill algorithm, Median algorithm, iterative noise removal method and other various pseudo morphologies. Although, these algorithms are used for noise reduction in images they do not provide an image with its detailed and thin line features. Many signals or images contain information at different scales or levels of details. Analyzing information at the same scale will not be effective.

Certain images contain noise which distorts the image and hence becomes difficult to obtain a clear image. This algorithm works to reduce these noises from the image using Mathematical morphology concept to obtain a smooth, noise free image.

I. INTRODUCTION

In photography and computing, grayscale or greyscale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black

at the weakest intensity to white at the strongest. Morphological openings and closings are useful for the smoothing of gray-scale images. However, their use for

image noise reduction is limited by their tendency to remove important, thin features from an image along with the noise. This algorithm works to reduce these noises from the image using Mathematical morphology concept to obtain a smooth, noise free image.

This paper discusses the algorithm used to reduce the noise from the images and their applications in image processing

II. Related theory

Certain images contain noise which distorts the image and hence becomes difficult to obtain a clear image. This algorithm works to reduce these noises from the image

using Mathematical morphology concept to obtain a smooth, noise free image.

Image noise is random (not present in the object imaged) variation of brightness or color information in images, and is usually an aspect of electronic noise. It can be produced by the sensor and circuitry of a scanner or digital camera. Image noise can also originate in film grain and in the unavoidable shot noise of an ideal photon detector. Image noise is an undesirable by-product of image capture that adds spurious and extraneous information.

The original meaning of "noise" was and remains "unwanted signal"; unwanted electrical fluctuations in signals received by AM radio caused audible acoustic noise ("static"). By analogy unwanted electrical fluctuations themselves came to be known as "noise". Image noise is, of course, inaudible.



Figure 1: Noise Reduction

III. PROBLEM DEFINITION

In photography and computing, grayscale or greyscale digit al image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. Morphological openings and closings are useful for the smoothing of gray-scale images. However, their use for image noise reduction is limited by their tendency to remove important, thin features from an image along with the noise. Hence new analysis of a new morphological image cleaning algorithm (MIC) that preserves thin features while removing noise. MIC is useful for gray-scale images corrupted by dense, low-amplitude, random, or patterned noise. Such noise is typical of scanned or still-video images.

Phase 1: Analysis and Planning

We have taken reference of 4 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers are secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and are found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

Phase 2: Designing and implementation

MIC is useful for gray-scale images corrupted by dense, low-amplitude, random, or patterned noise. Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighbourhood of pixels. Some operations test whether the element "fits" within the neighbourhood, while others test whether it "hits" or intersects the neighbourhood.

MATLAB tool will be used to process the algorithm for image noise reduction.

Phase 3: Testing and Deployment

To check if the image filtering process works properly and try to compare it with the original image to make sure that the new image produced is noise free.

IV. FEATURES OF THE PROJECT

Morphological openings and closings are useful for the smoothing of gray-scale images. However, their use for image noise reduction is limited by their tendency to remove important, thin features from an image along with the noise. New analysis of a new morphological image cleaning algorithm (MIC) that preserves thin features while removing noise. MIC is useful for gray-scale images corrupted by dense, low-amplitude, random, or patterned noise. Such noise is typical of scanned or still-video images.

A useful aspect of MIC is that it enhances JPEG compression in the grayscale images. MIC is applicable to a range of imagery beyond faces. Edges, thin lines, and small features are made sharp and clear unlike other filtering algorithms like median filtration. Also, areas between these features are smoothly varying.

V. METHODOLOGY

The Proposed methodology is illustrated in Figure. It contains following steps:-

- Data Preprocessing
- Data Integration
- Implementation

VI. FEASIBILITY STUDY

- Technical Feasibility – The algorithm will operate and will be implemented on MATLAB. Proven and tested grayscale images to deliver the proposed solution. It works on all platforms which supports MATLAB.

Legal Feasibility-The project is developing under legal license of windows, MATLAB and other software's required. The data will not be used for some illegal purpose and will be used for improving the quality of the image relevant to user. Data and image will not be misused or misguided under any circumstance. The proposed system will not violate any laws, rules or regulations as it is a research based project.

Operational Feasibility - With the proposed algorithm we will receive an image with sharp details which is noise free.

Social Feasibility- The proposed system is being developed by people with technical skills and knowledge about MIC Algorithm.

Combine the features identified in D with S while leaving the noise behind.

Conclusion:

The result is an image that is smoothly varying and whose edges and other 1D features are as sharp as original yet has smooth regions between them.

Data Pre Processing

Load the image. Identify/Calculate the openings and closing of gray-sacle images.

Data IntegrationPreserve and store thin features of corrupted images by dense, low-amplitude, random/patterned noise.

Identify edges, thin lines and small features that will make image “clean” and “sharp”.

Also, area between these features should be smooth.

Implementation

Identify the type of noise. The type of noise could be- Salt and pepper noise, Gaussian noise etc.

Consider image I. let S be the result of smoothing I. Also S is now noise free.

The difference image in the image is D such that, $D = I - S$ D contains all the noise in I.

S cannot contain any features with non-zero support. Hence D contains features and noise.

VII. BLOCK DIAGRAM

Many techniques for noise reduction replace each pixel with some function of the pixel's neighborhood. Because 1- D features and 2-D noise usually have common frequency components, they are not separable in the frequency domain. Linear filters tend either to amplify the noise along with the 1-D features or smooth out the noise and blur the 1-D features. both goals. Morphological filters are, perhaps, the most well-known nonlinear filters for image enhancement.

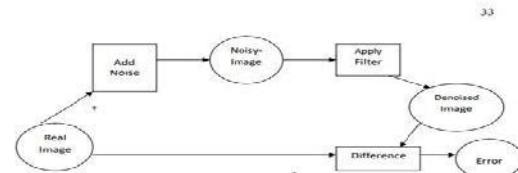


Fig 2: Flow Diagram

VIII. FLOW CHART

Below flow chart shows the flow of the working of the proposed algorithm.

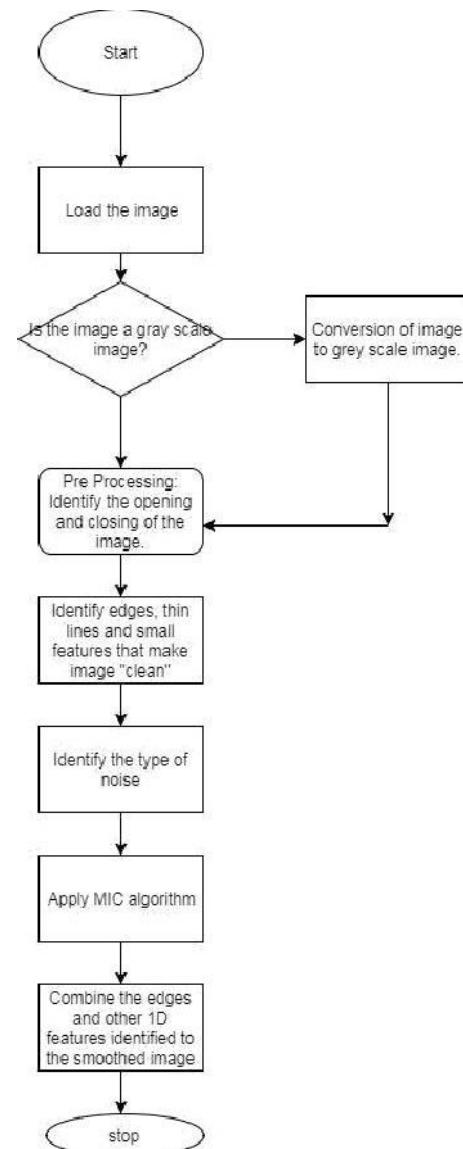


Fig 3: Flow Chart

IX. RESULTS

It is necessary to evaluate the performance of the system in presence of noise to measure the efficiency and reliability of the system. The Fig. 3.1 depicts the evaluation of the noise reduction filter performance. The noise will be added artificially with the original image for getting noisy image and then the filtering technique is utilized to remove the unwanted information. The filtered image is compared with the real image for calculating the difference to find the errors in the filtering process.

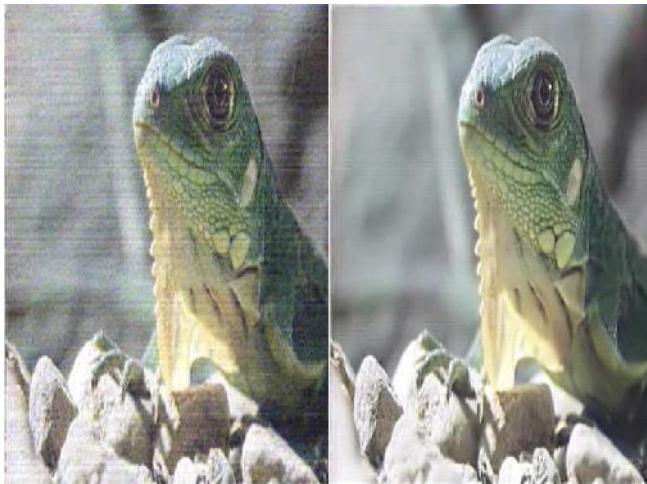


Fig 4: Result

X. FUTURE SCOPE

- In this research work, comprehensive experiments on restoration and enhancement process on images of 1D and 2D images has been conducted. Several aspects related to remove noise, have been addressed in detail. This work can be further enhanced for 3D images.
- The noise removal scheme has been implemented for stationary images. This can be extended to noise removal in case of non-stationary images for dynamic denoising.

XI. CONCLUSION

Thus we have used a new algorithm (MIC) for image noise reduction based on morphological size distributions. MIC smooths the image in a number of size bands, subtracts these bands out of the image to create residual images, segments the residuals into features and noise, and adds the features back to the smoothed image. This way it also retains the thin features of the images.

XII. ACKNOWLEDGMENT

We sincerely thank to our guide Dr. Bijith Marakarkandy, our HOD Dr. Rajesh S. Bansode, our Dean Dr. Kamal Shah

and our principle Dr. B. K. Mishra for his/her guidance and support for carrying out our project work.

REFERENCES

- [1]Richard Alan Peters, II, Member, IEEE: "A new algorithm for image noise reduction using mathematical morphology".(1995)
- [2]K. Chinnasaran, Y. Rangsanseri, P. Thitinajshima:"Removing salt and pepper noise in text or graphics images".(1998).
- [3] Anna Fabijańska, Dominik Sankowski :"Image Noise Removal – The New Approach".(2007)
- [4] RADU-MIHAI COLIBAN.,MIHAI IVANOVICI,:"IMPROVEDPROBABILISTIC PSEUDO-MORPHOLOGY FORNOISE REDUCTION IN COLOUR IMAGES".(2016).

Tracking System For Mentally Challenged People

Chirag Rai

Department of Information Technology

Thakur College of Engineering and Technology

Mumbai, India

Namdeo Badhe

Department of Information Technology

Thakur College of Engineering and Technology

Mumbai, India

Aditya Rao Nuty

Department of Information Technology

Thakur College of Engineering and Technology

Mumbai, India

Ayush Kedia

Department of Information Technology

Thakur College of Engineering and Technology

Mumbai, India

Abstract —The treatment for Mentally challenged patients can be very troublesome and difficult for the doctors and the patient's family as they have to be attended to constantly which is quite tiresome for the patient's family and doctors. Also, the patients may be unable to express any type of discomfort that they may be feeling, therefore, we create a device and supporting application to help locate the patient and monitor the patient's heart rate. The device will monitor the patient's heart rate constantly so if there is any drastic variation in the patient's heart rate a list of emergency contacts can be notified. The device is connected to an android application that will upload all the patients data on to a cloud so that the doctors and patient's family can easily track the patient.

Keywords – analysis; GPS; cloud; database; monitoring; application; smart band.

I. INTRODUCTION

Census 2001 has revealed that over 21 million people in India as suffering from one or the other kind of disability. This is equivalent to 2.1% of the population. Among the total disabled in the country, 12.6 million are males and 9.3 million are females. Although the number of disabled is more in rural and urban areas. Such proportion of the disabled by sex in rural and urban areas. Such proportion has been reported between 57-58 percent for males and 42-43 percent females. The disability rate (number of disabled per 100,000 populations) for the country as whole works out to 2130. This is 2,369 in the case of males and 1,874 in the case of females.

Tracking system for mentally challenged people is an application which is primarily aimed to help mentally challenged people by continuously monitoring their location and notifying the emergency contacts in case some unusual behavior is seen. The proposed system is connected to a cloud and all the database of the patients are stored on the cloud and whenever required can be retrieved. Also the patient's heart rate will be monitored and recorded so that any other heart indication which is different from that record will be checked upon for an emergency.

To locate the patient, we use a portable GPS device in the smart band and a heart rate monitor to monitor the patient heart rate. In case of any emergency, the emergency contacts are notified. All the patient's database is stored on the cloud. The smart band is connected to a smart phone using android programming which uploads all the data to the cloud.

II. BACKGROUND

About 15 percent of the world's population -- some 785 million people -- has a significant physical or mental disability, including about 5 percent of children, according to a new report prepared jointly by the World Health Organization and the World Bank. Disability is further complicated simply by definition, as was evident in two surveys done in Ireland in the past decade. In 2005, the Central Statistics Office conducted a rigorous survey of the population that included two questions about physical functioning. It found that 9.3 percent of the population was disabled. Later that year, a separate survey interviewed households more extensively and added chronic pain and breathing problems to the menu of disabling symptoms. It came up with a national disability rate of 18.5 percent.

The **Internet of things (IoT)** is the network of physical devices, vehicles, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data. Each thing is uniquely identifiable through its embedded computing system but is able to interoperate within the existing Internet infrastructure. The IoT allows objects to be sensed or controlled remotely across existing network infrastructure,^[5] creating opportunities for more direct integration of the physical world into computer-based systems. **Cloud computing** is an information technology (IT) paradigm, a model for enabling ubiquitous access to shared pools of configurable resources (such as computer networks, servers, storage, applications and services),^{[1][2]} which can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing allows users and enterprises with various computing capabilities to store and process data either in a privately-owned cloud, or on a third-party server located in a data center - thus making data-accessing mechanisms more efficient and reliable. Android apps can be written using Kotlin, Java, and C++ languages. The Android SDK tools compile your code along with any data and resource files into an APK, an Android package, which is an archive file with an.apk suffix. One APK file contains all the contents of an Android app and is the file that Android-powered devices use to install the app

III. MOTIVATION

Problems regarding monitoring of every single patient and his or her activities in the hospital are a major part for better functioning of the hospital as well as for the family of the patient. So the family of the patient face many problems such as if the patient's shows up with some unusual activities or is in some danger so for overcoming this problem we came with our project in which we will make a track band which will track all the patients activities and heart rate of the patient so that even if any unusual activity is observed so the emergency contacts such as patients family, hospital or the ambulance can be notified.

The main purpose of our project is to analyse the data objects which will be monitored timely and provide growth opportunities to the patient. Through the survey we found that there is need to develop such bands which will be used by the hospitals effectively to provide stability to the mentally ill patients and their families. And this motivated us to develop a smart band which will fulfil the needs of the patient as well as the family of the patient.

IV. SCOPE

The patient may have a smart band on his wrist that may continuously monitor patient's location and heart rate and may update it to the cloud. The patient data is stored on the cloud and the doctors may access all the patients' data through cloud. In case of any unusual behaviour the

emergency contacts are notified about the patient so that his/her parents or guardian can look after the patient or the hospital can have check over it.

V. PROBLEM DEFINITION(PHASE WISE)

Phase 1: Analysis and Planning

We have taken reference of 3 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers re secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and are found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

Phase 2: Designing and Implementation:

To locate the patient, we use a portable GPS device in the smart band and a heart rate monitor to monitor the patient heart rate. In case of any emergency, the emergency contacts are notified. All the patient's database is stored on the cloud.

The smart band is connected to a smart phone using android programming which uploads all the data to the cloud.

Phase 3: Testing:

To check if the GPS and heart rate monitor are working properly and sending all the data accurately to the smart phone.

To check if all the data is uploaded correctly to the cloud.

Phase 4: Deployment:

If new key findings are found during testing phase we will add up to our project and deploy the same product to the proposed market.

Methodology

1. A device is to be created which will track the patient's location and his heart rate and notify the emergency contacts if any danger is seen.
2. An application is to be created that is linked to the device and provides all the functionalities of the device to the users.
3. The application has to be linked to a cloud service sos as to make the data easily available to all.



Figure 1: Overview of project

VI. FEASIBILITY STUDY

Technical Feasibility: The system works on the app model and is the system implemented using Android and C/C++ for Arduino, as programming and scripting languages. Proven and tested mature technologies to deliver the proposed solution.

Economic Feasibility: The cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud. Cloud computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud. Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

Legal Feasibility- The project is developed under legal license of Android, Arduino and other software's and databases. The data will not be used for some illegal purpose and will be encrypted using algorithms.

Operational Feasibility- An overview of the process portrayed that the proposed Tracing System and Android application would be beneficial by providing a unique concept to users hence being the only player in the market.

VII. FLOW DIAGRAM

Figure 1 is a flowchart diagram. Flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic

representation illustrates a solution model to a given problem.

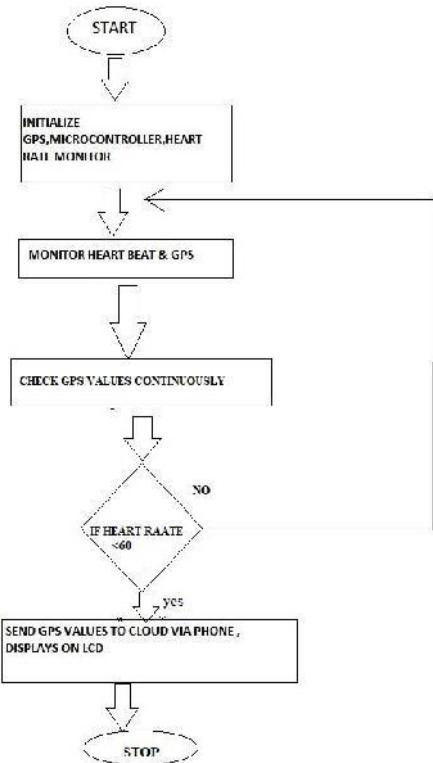


Figure 2: Flow Chart Diagram for registration and sign in.

VIII. RESULTS

Phase 1: Analysis and Planning:

Collected all the required information and tools regarding the building of the smart band which will help to track the patients via different set of activities.

Phase 2: Designing and Implementation:

Designed the basic GUI of the application which contains all the basic functionalities provided to the users, thereby simplifying the process of tracking the patients.

Phase 3: Testing:

With the developed applications different test cases were performed to find the efficiency of the product.

Phase 4: Deployment:

After performing different test cases the product is ready for deployment and now can be used by the patient's family to track their daily activities.

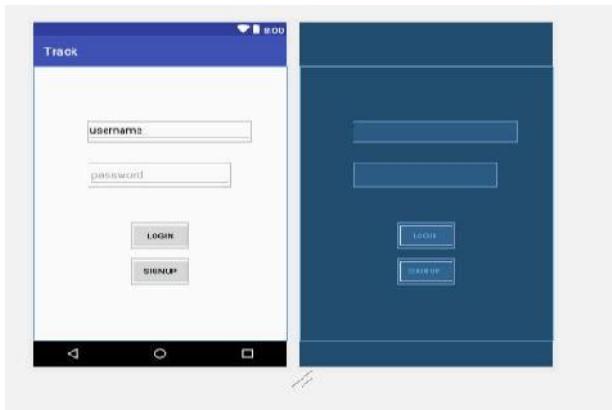


Figure 3: Login page of the application.

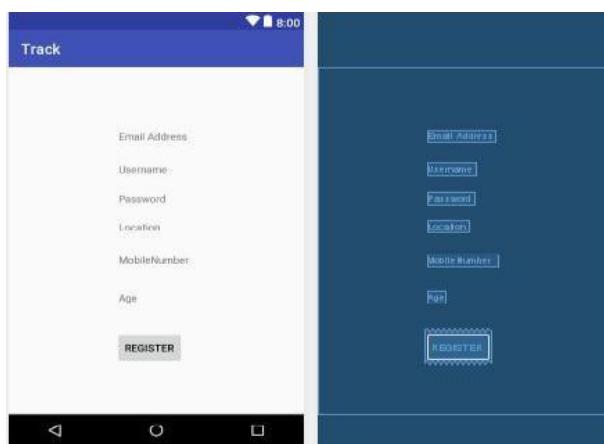


Figure 4: Sign Up page of the application.

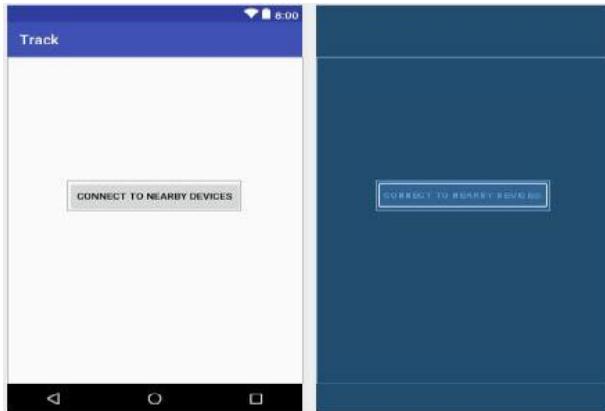


Figure 5 :patients location address page

IX. CONCLUSION

The tracing system has been a work in progress in terms of the backend connectivity and the user interface. The final version of the app is expected to support the entire feature set along with the smart band being able to efficiently track the patient. The tracking system is expected to strive based on the lack of any competition in the market currently.

REFERENCES

Research Papers:

- [1]. KaYiu Sham, Philip Lim-Kong Wong "Fitness monitoring device having an electronic pedometer and a wireless heart rate monitor", Apr 6, 1999, In US 5891042 A
- [2]. Erin-Ee-Lin Lau, Wan-Young Chung "Enhanced RSSI-based Real-Time user location tracking system for indoor and outdoor environments [online] May 2007, 04:109/ICITET17/ICITET17002.pdf
- [3] Tom OuYang, Farah FerdousIshita, Peter Le Ashton Novak-Louie, GiftyQuansah“– Intelligent wearable wristband for personal safety.”February 12,2016, In ENSC 440.

Web References:

- 1) <http://www.ijitet.com/ICITET17/ICITET17002.pdf>
- 2) <https://www.google.co.in/patents/US5891042>
- 3) <http://ieeexplore.ieee.org/document/4420422/>

Text Books:

- 1) Moving to the cloud- SITARAM
- 2) Arduino: 101 Beginners Guide-ERIK SAVASGARD

Anti-Theft Detection and Tracking System for Car Using IoT

Harish Baheti

Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India

Abhishek Patahak

Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India

Akshata Desai

Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India

Zahir Alam

Department of Information Technology
Thakur College of Engineering & Technology
Mumbai, India

Abstract-Despite the various technologies that have been introduced in recent years to deter car thefts and tracking it, it was reported that as many as cars were stolen yearly in the world. According to National Crime Information Center (NCIC), in 2006, 1,192,809 motor vehicles were reported stolen, the losses were \$7.9 billion. Several security and tracking systems are designed to assist corporations with large number of vehicles and several usage purposes. A fleet management system can minimize the cost and effort of employees to finish road assignments within a minimal time. An efficient automotive security system is implemented for anti-theft and accident detection, using an embedded system consisting of a Global Positioning System (GPS) and GSM Module. All the existing technology support tracking the vehicle location and status. The GPS/IoT based system is one of the most important systems, which integrate both GPS technology and upcoming IoT or Internet of Things terminology. The system in the event of theft will send predefined message to the owner of vehicle. The user can start tracking the position of targeted vehicle on Google Earth on a dedicated vehicle tracker server. Using GPS locator, the target's current location is determined and sent, along with various parameters received by vehicle's data port, via GSM module on the mobile phone of the user. According to the literature survey the project is prepared by performing research on various papers related to our topic, it was found that the accuracy of GSM module was 5-10 meters. After determining all the results from the literature survey, the project was developed across expecting an accuracy of 1-5 meters while tracking the car in case of theft. The project will use MAX 232 IC which is dual transmitter and dual receiver. After studying the literature survey, the planning for project scope was done and it was decided to use Raspberry pi 3 instead of Arduino board. In case of accident, by using GPS and GSM module, the location can be send to the hospital and police station. This model can be used in private bus and motorcycle. The use of this system can also be extended to the car dispatching management field which has a vast potential for future development.

I. INTRODUCTION

In the last few decades, India has progressed at such an enormous rate that many companies have strongly established themselves here. These companies bring a huge amount of workforce with them. Arranging transportation to such a huge mass is a cumbersome task involving many intricacies. Generally, this transport is arranged through the local transport vendors on a yearly contract basis, with increasingly happening mishaps such as burglary, accidents etc. posing a bigger concern the development of satellite communication technology is easy to identify the vehicle locations. Vehicle tracking systems have brought this technology to the day-to-day life of the common person. Today GPS used in cars, ambulances, fleets and police vehicles are common sights on the roads of developed countries.

All the existing technology support tracking the vehicle location and status. The GPS/IoT based system is one of the most important systems, which integrate both GPS technology and upcoming IoT or Internet of Things terminology. With the help of GPS, the vehicle can be tracked on a continuous basis which in conjunction with law enforcement authorities or hospitals can be used to track the vehicle if it gets stolen or if the vehicle gets involved in a serious mishap. The IoT term is a relatively new and upcoming term. According to Wikipedia, The Internet of Things (IoT) is the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data.

II. SCOPE

The future direction of this work includes the consideration of the characteristics of the owner and the car should respond accordingly. Cameras can be added to identify the person responsible. Similarly, with the help of high intensity vibration sensors we can integrate accident detection in the proposed anti-theft detection system. In case of accident by using GPS and GSM module we can send the location to the hospital and police station. We can use this model in private bus and motorcycle. We can extend the use of the system to the car dispatching management field which has a vast potential for future development.

III. MOTIVATION

A primary motivation of developing this project is increasing number of car thefts. According to the NCRB data, 8,171 motor vehicles were stolen in Delhi in 2012. It is a serious problem in automobile field. This issue is a motivation to design this system. A vehicle is stolen every 13 minutes in New Delhi marking a sharp 44% rise over the same period last year. Only around 4% of these cars are recovered. An estimated 53 passenger and light-commercial vehicles are stolen each week in Australia for conversion into scrap metal. Many car owners assume that because the cars they drive are neither new nor flashy, that they needn't be concerned with potential car theft. The reality is that not every car that is stolen is being considered for resale; many stolen vehicles are taken simply because they can be dismantled into valuable scrap metal. Owners of older or less-expensive vehicles who fail to take the same precautions as owners of new vehicles put themselves in danger of losing a car, and never being able to recover it.

IV. PROBLEM DEFINITION (PHASE WISE)

A novel method of vehicle tracking and theft detection using GPS and GSM technology. This project aims at developing a cost effective anti-theft detection for car which could cater the needs of the society. It uses the Driver IC to connect Raspberry Pi 3 to the car. It uses MEMS ADXL-335 as accelerometer sensor which is converted from analog to digital using MCP 3208.

A. PHASE 1: (Planning, Analysis, Design, Coding)

Planning – Conceptualizing and designing the architecture for proposed solution. Choosing whether to use Raspberry Pi 3, Arduino, or Orange Pi systems.

Analysis – Critical study, analysis and review of feasibility for proposed solution.

Choosing the right type of components such as Raspberry Pi 3, GSM and GPS Modules & MEMS ADXL 335.

Design – Designing the flow chart for the proposed system.

Coding – Creating various modules of an IT based solution product for analyzing the theft detection of the car

B. PHASE 2: (Integration, Testing, Deployment)

Integration – Integrating of various proposed modules such as rash driving, notifying the user about the theft of the car.

Testing – Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.

Deployment – Give the completed prototype to the authorized person for evaluation purpose and beta testing.

V. METHODOLOGY

The proposed solution, “Anti-theft Detection and Tracking System for car using IoT” uses a MEMS ADXL 335 Accelerometer to detect the behavior of the car. It uses GPS Module to locate the real time location of the car and also uses GSM Module to notify the user about the same. It uses relay to connect the GSM and GPS Modules to MAX 232. MAX 232 is connected to the Raspberry Pi3. It also uses MCP 3208 Analog to Digital converter. So, any kind of theft can be detected and the car can be tracked using the proposed model. Relay is used in the system to connect GPS and GSM modules to the MAX232 integrated circuit which is further connected to the Raspberry Pi 3.

VI. FEASIBILITY STUDY

Technical Feasibility - Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.

Economic Feasibility - As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.

Legal Feasibility - Legal issues can affect a system's

acceptance by Bottom users, its performance, or the decisions on whether to use it in the first place so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any log analysing system.

Operational Feasibility - It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.

Social Feasibility - The acceptability of a log analysis system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness.

VII. BLOCK DIAGRAM

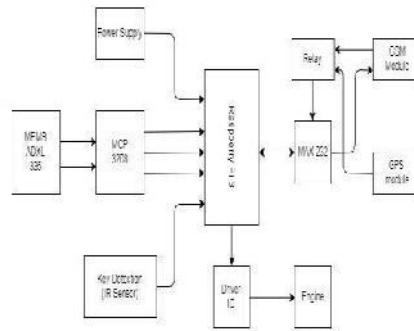


Fig. 1. Block diagram

The components required for the proposed system and their connectivity is as shown in Figure. It consists of 3 main components Raspberry pi 3, GSM module and GPS module.

The figure is for anti-theft detection and tracking system for car using IoT. It has a microprocessor which is connected to GSM and GPS modules through relay. MAX232 acts as a dual transmitter and dual receiver. MEMS ADXL335 is accelerometer sensor and MCP3208 is analog to digital converter. Power supply is needed for Raspberry Pi 3. Key detection, Driver IC and Engine are also connected to demonstrate the working of the proposed system.

VIII. LITERATURE SURVEY

In the paper [1], Getting proper geographical location with the 3 axis of the car. Send location of car to law enforcement authority in case of theft. Paper [2] refers to use of microcontroller Atmega 162 based Arduino.Used GPS and GSM module for tracking and communicating with the owner of the car.. In paper [3] it refers to alerting the owner of the car regarding the theft. It makes use of 32-bit microcontroller LPC2148.Paper [4] It refers to designing the database to store and managed received vehicle's location information. So the research gap found out to be to lock the door of the car during the theft to capture the thief with the help of Raspberry pi 3. Accident detection and generating message which will inform family members about the location of accident.

IX. EXPECTED RESULT

a. Outputs:

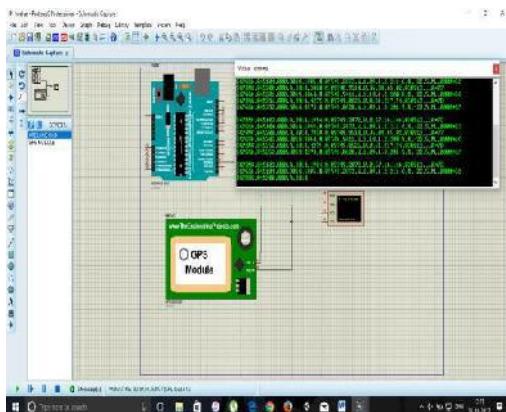


Fig2. simulation result with Proteus 8

The figure shows the simulation which was done on Proteus 8 Professional software. It is a third-party application which acts as a software to create simulations before working on actual project. In this result we have used a microcontroller and GSM module to show the location of the system.

b. Outcomes:

Phase 1.1 Planning

- Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done

- Using Raspberry Pi3 and GPS & GSM Modules as a communication tool.

Phase 1.2 Analysis

- Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.
- Choosing GPS & GSM Modules as a communication tool and RaspberryPi3 over Arduino and Orange Pi Board.

Phase 1.3 Design

- Proposing detection system for theft of car and tracking the real time location of the car.

Phase 1.4 Coding

- Various modules will be created like MEMS ADXL 335
- Analog to digital converter
- Getting the geographical location of the car

Phase 2.1 Integration

- Constructing the various modules of the proposed solution to integrate them into a prototype.

Phase 2.2 Testing

- The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stress testing.

Phase 2.3 Deployment

The benefits of Analysis on after deploying the proposed prototype include:

- Proper and more efficient way of tracking the car in metropolitan cities.
- Easier to get the location of the car in case of theft or rash driving.
- Easy to communicate with the authorized owner in case of theft.

c. Percentage improvement in results

In GSM module, 66.67% of improvement is seen in the result as the accuracy has been increased from 5-10 meters to 1-5 meters.

X. FUTURE SCOPE

The future direction of this work includes the consideration of the characteristics of the owner and the car should respond accordingly. Cameras can be added to identify the person responsible. Similarly, with the help of high intensity vibration sensors we can integrate accident detection in the proposed anti-theft detection system. In case of accident by using GPS and GSM module we can send the location to the hospital and police

station. We can use this model in private bus and motorcycle. We can extend the use of the system to the car dispatching management field which has a vast potential for future development.

XI. CONCLUSION

- Although no vehicle can be fully secured, this project aims to bridge the gap between flexibility and security. This proposed system provides a reasonable idea to achieve and improve security and preventive measures in the events of thefts and accidents.
- In the proposed model, a low-cost vehicle tracking and theft detection system is presented. The model includes a transmitting module which combines GSM and GPS modules to retrieve location of the car and send it to the authorized person.
- The owner can simply use mobile phone to locate and monitor the car in real time, which provides the technical foundation for the development of wireless network vehicle anti-theft system.
- The owner can use the proposed model to contact nearest hospital and police station with the help of GSM and GPS module in case of accident.
- The proposed system can be used in companies where product of high security quotient is to be delivered to the user.
- For future use we can extend the use of the system to the car dispatching management field.
- The project also tries to improve upon existing security systems which provides a reasonable security but have certain shortcomings.
- There are areas in which the proposed system can be improved upon and further improve the robustness and capabilities of the system.

REFERENCES

- [1] I.Karande; G.Deshpande; S.Kumbhar; A.V.Deshmukh, "Intelligent Anti-Theft Tracking and Accident Detection System for Automobiles Based on Internet of Things", "International Journal of Innovative Research in Computer and Communication Engineering, India Online ISSN: 2319-7064" on, Vol. 4.Issue 3,March 2016.
- [2] Vigneshwaran.K; Sumithra.S; Janani.R, "An intelligent system based on GSM and GPS", "International Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering" on, Vol. 4, Issue 5, May 2015.
- [3] Tarapiah, S.; Atalla, S.; Alsaid , B., "Smart on-board transportation management system Geo-Casting featured," Computer Applications and Information Systems (WCCAIS), 2014 World Congress" on, vol. 1, Issue. 6, Jan. 2014.
- [4] D.Santhoshi Rani ;K. Radhika Reddy ,” Raspberry pi based vehicle tracking and security System for Real Time Applications”, “International Journal of Computing Science and Mobile Computing, IJCSMC ”, Vol. 5, Issue. 7, July 2016, pg.387 – 393.

Secure Cloud storage of Biometric Templates with Secure Network Coding

Jewel Correia

*Department of Information Technology
Thakur College Of Engineering and Technology
Mumbai, India
jewelcorreai@gmail.com*

Deven Shah

*Department of Information Technology
Thakur College Of Engineering and Technology
Mumbai, India
sir.deven@gmail.com*

Vikas Kaul

*Department of Information Technology
Thakur College Of Engineering and Technology
Mumbai, India
mvjvikas@gmail.com*

S K Narayankhedkar

*Department of Information Technology
MGMCOT
Navi Mumbai.India
skniitb@yahoo.com*

Abstract— Biometric designs have attracted attention in practical technological schemes with high requirements in terms of accuracy, security and privacy. Cloud service providers offer storage outsourcing facility to their clients. However, with Cloud platforms user's data is moved into remotely located storages such that users lose control over their data. This unique feature of the Cloud is facing many security and privacy challenges which need to be clearly understood and resolved. One of the important concerns that needs to be addressed is to provide the proof of data integrity, i.e., correctness of the user's data stored in the Cloud storage. The data in Clouds is physically not accessible to the users. Therefore, a mechanism is required where users can check if the integrity of their valuable data is maintained or compromised. For this purpose some methods are proposed like mirroring, check summing. In a secure cloud storage (SCS) protocol, the integrity of the client's data is maintained. In this work, we construct a publicly verifiable secure cloud storage protocol based on a secure network coding where the client can update the outsourced data as needed. In this paper, we address the problem of proving data integrity in Cloud computing by proposing a scheme through which users are able to check the integrity of their data stored in Clouds.

Keywords-Cloud Storage auditing; network coding security; user authentication; Data Integrity; AES 256 Encryption

I. INTRODUCTION

Cloud computing is the latest paradigm that involves delivering hosted services over the internet based on pay as you go approach. Cloud computing is a technology that uses the internet and central remote servers to maintain data and applications. It is a marketing term for technologies that provide computation, software, data access and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing. Cloud computing allows consumers and businesses to use

applications without installation and access their personal files at any computer with internet access. Cloud computing is broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world.

Cloud computing environments are likely to suffer from a number of known vulnerabilities, enabling attackers either to obtain computing services for free (attack against cloud providers), steal information from cloud users (attack against cloud customers data), or penetrate the infrastructure remaining in client premises through cloud connections (attack against cloud customer infrastructures). Big IT giants like Google, Amazon, and salesforce.com are providing computing facility like storage, computation and application by pay as per usage through Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) cloud service models. Since cloud computing supports distributed service oriented architecture, multi users and multi-domain administrative infrastructure, it's more prone to security threats and vulnerabilities. Security issues are of more concern to cloud service providers who are actually hosting the services. With the sudden technology-outburst in the recent past, security is one of the biggest rising concern in the real world as well as online systems. Human identification and authentication is an important criterion for the surveillance systems in real world as well as online security systems. Biometrics is the most efficient answer for all the problems and it has been widely accepted.

II. LITERATURE SURVEY

The primary goal of this research is to explore how biometrics can be deployed in the cloud and how the enrolled biometric templates can be shared between the API calls and Services, taking into account the challenges faced when using biometrics for distributed applications. These challenges include the security of biometric templates, the privacy of users, and the trust for remote biometric operations. In paper[1] the author addresses the issues faced by traditional online signature recognition system and how they are resolved by implementing the online signature recognition system on cloud. The

proposed architecture has the Public cloud based architecture and successfully implemented on Microsoft Azure cloud which aims to devise a highly scalable, pluggable and faster online signature recognition system. This paper describes in detail the process of signatures acquisition through the digitizing tablet, storing signature in a blob storage, sending messages to and from the service bus queue and lastly feature vector extraction on the cloud and is capable of operating on enormous amounts of data, which, in turn, induces the need for sufficient storage capacity and significant processing power with a more cost effective approach.

In paper [2] describes two types of implementation, public cloud and hybrid cloud based architecture.

In this research the public cloud based architecture is discussed in detail and successfully implemented on Microsoft Azure cloud which aims to devise a highly scalable, pluggable and faster online signature recognition system. This architecture gives 90% improvement in execution speed as compared to existing implementation.

This report describes in detail the process of signatures acquisition through the digitizing tablet, storing signature in a blob storage, sending messages to and from the service bus queue and lastly feature vector extraction on the cloud. The successive geometric center of depth-2 is used for feature vector extraction. The feature vector is modified by adding the Timestamp information. The modified feature vector along with Classifier SaaS gives higher CCR.

In paper [3] Raspberry Pi is implemented as a remote authentication node. The AES is implemented in MATLAB and the biometric traits are encrypted using Round Structure and the dynamic S-box. Using AES in Round structure with more no of rounds, runtime is increased but complexity of network is also increased.

Increasing complexity will make the system attack resistant and secure data from attackers. So this system can be used in the application where time is not the constraint. The time taken by the AES in Round structure with one round is nearly same as traditional AES hence it can be used in the applications where speed is required with complexity. This is a low-cost system which will ensure high security and high scalability.

In paper [4] it has revealed a relationship between these two different areas, i.e., secure cloud storage and secure network coding. The main result is that it can construct a publicly verifiable secure cloud storage protocol given any publicly verifiable secure linear network coding protocol. In contrast, secure cloud storage protocols are currently designed in a rather ad hoc way and there are only few successful protocols, the design has first publicly verifiable secure cloud storage protocol which is secure in the standard model, i.e., without modeling a hash function is a random function when disagreeing for the security of the protocol. Moreover, the paper did structure to support advanced functionalities, in particular, user anonymity, and third-party public auditing.

In paper [5] The authors describe how computationally intensive biometric recognition can be performed on a mobile device by offloading the actual acknowledgment process to the cloud. The authors have anticipated a systematic approach for dividing a recognition operation and a bulk enrollment operation into multiple tasks, which can be executed in analogous on a set of servers in the cloud, and shown how the results of each task can be combined and post-processed for individual recognition.

In paper [7] the use of an additional key to increase the level of difficulty to crack the algorithm. The Key space of AES is limited, which is solved by making use of any additional key called Chaos Key.

It not only increases the key space, but also provides security against statistical attacks. The Encryption & Decryption time depends on the complexity of the equation chosen for Chaos.

In paper [8] a conceptual understanding of the AES is explained

III. PROBLEM DEFINITION

In order to protect the privacy and security of biometric data, secure connections, e.g. Secure Socket Layer (SSL), might be used to encrypt the biometric data while in transfer to a remote server. However, using a secure connection alone does not fully resolve the problem. Firstly, the security and privacy of biometric data are not only at risk while in transfer. Biometric data can also be compromised or identified while stored or processed at the server. In addition, some privacy-concerned users want to take advantage of the usable and strong authentication capabilities of biometrics, but they do not want to reveal their identities to the remote server. For protecting biometric data that is stored on a server, standard encryption methods are not a secure and practical solution.

Hence in the proposed system end-to-end encryption based on AES-256 will be used as secure network coding as shown in Figure 1. Just simple encryption technique is not secured to the project, so here the project will be using AES with Chaos in round structure encryption process and SHA-256/MD5 for data integrity check. The biometric templates will be secured on cloud using this mechanism. The encrypted files and hash file of encrypted file is stored on cloud. Also the directory alongwith the single hash file is stored on cloud. The comparision of hash files verifies the data integrity.

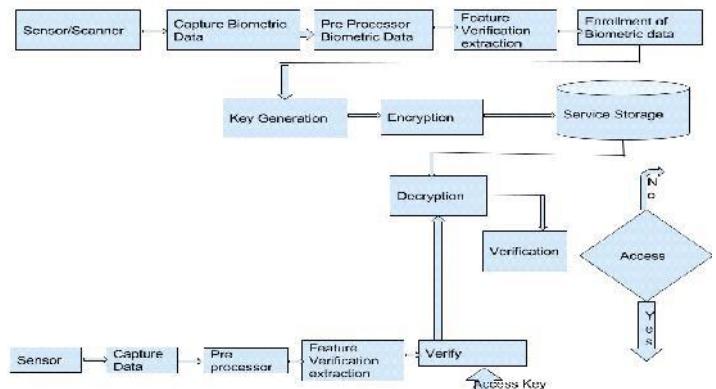


Figure 1: Proposed System

A. AES

The Advanced Encryption Standard (AES) is an encryption algorithm for securing sensitive information by U.S. Government agencies. The Rijndael algorithm proposed as an accepted standard is discussed in the sections that follow. Figure 1 elucidates the main steps comprising the AES algorithms [6]. The four major functions that comprise the AES are Add Round Key, Substitute bytes, Shift Rows and Mix Columns. Rijndael is a substitution linear transformation network with 10, 12 or 14 rounds, depending on the key size.

AES has a fixed block size of 128 bits and a key size of 128, 192, or 256 bits. Rijndael specified with block and key sizes that

it may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits as shown in Figure 2.

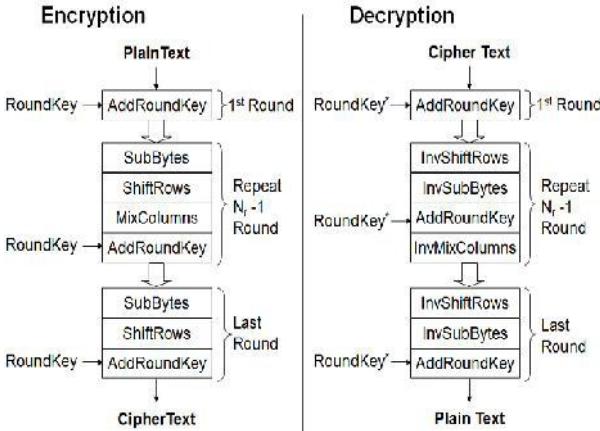


Fig 2: Major Functions of AES

B. Chaos

The chaos is one kind of nonlinear movement form. It is produced by a definite system, and it relies on the initial condition, and it is unpredictable. The chaos system has several characteristics [9]: stochastic, sensitive to initial condition, long-term unpredictability and so on. Chaos theory studies the behavior of dynamical systems that are highly sensitive to initial conditions. Small differences in initial conditions yield widely diverging outcomes for such dynamical systems, rendering long-term prediction impossible in general. This happens even though these systems are deterministic, meaning that their future behavior is fully determined by their initial conditions, with no random elements involved. In other words, the deterministic nature of these systems does not make them predictable. This behavior is known as deterministic chaos, or simply chaos.

C. Improvement in AES using Chaos

The AES used within the Fiestel Network can be improved by this concept. Some scholars used two two-dimensional chaotic systems to get two sequences, and then they do “XOR” with the two sequences. The results of this operation are the keys of the whole system [20]. We also used two chaos systems which can generate two keys as shown in Figure 3. One can be used as the encryption key, the other one as the controller key which can control the times of row-shift. The characteristics of chaos sequence make the space of key infinite.

In order to test easily, the improved algorithm uses two simple chaotic systems [9]. The key is the initial value of each chaotic sequence. Its expression is:

$$\begin{cases} y_{i+1} = (y_i + \frac{kH}{2\pi} \sin \frac{x_i 2\pi}{H}) \bmod H \\ x_{i+1} = (x_i + y_{i+1}) \bmod H \end{cases}$$

where $x \in [0,1], y \in [0,1], k > 0$.
[19] shows methods to get chaotic maps according to various block cipher encryption based on Fiestel network.

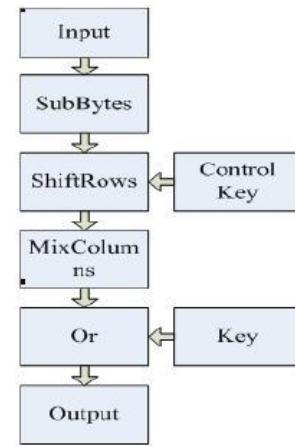


Fig 3 : Structure of enhancement in AES

Step1: Reading 256 bit data as input, dividing it into the two parts of equal size (128 bit).

Step2: Perform fiestel function operation with the right part using round key. XOR the result with the left part.

Step3: The result of above step will be an input to AES rounds.

Step4: Inputs two chaos system's starting values to produce sequence which needs in the encryption process. The first sequence uses for doing the 'XOR' operation with the input date which had been transformed, the second sequence each time produces four keys. The first five of each key is the order numbers of each line shift.

Step5: After each round of AES, two chaos sequences produce the new encryption key and the control key;

Step6: The output we get after all AES round should be considered as next left part of the Fiestel network.

Step 7: Continue the same process until we complete all the rounds we fixed for Fiestel network.

Figure 4 on next page shows the detail structure of hybrid AES in which AES rounds are also shown with chaotic sequenced keys as shown in Figure 4.

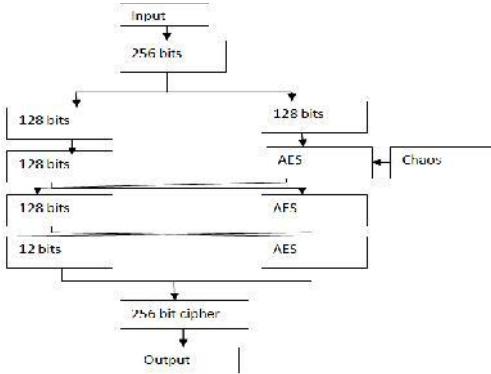


Fig 4: AES with Chaos

D. PERFORMANCE ANALYSIS OF IMPROVED AES

1. Key space analysis

Shneier a famous scholar put forward: a good key space must be large, so the powerful attack is impossible [15]. Chaos system meets it very well. The ergodicity of chaotic system make the space of key distribution uniformity and the key space unlimited, so it can better against some shortcuts. This new method has six initial values. If the type of initial value selection is double, the key space of the chaotic sequences can be 10^{15} , and the key space of the whole encryption system can be 10^{90} . It is more than the original AES whose key space is 2^{128} .

2. Key sensitivity analysis

Shneier pointed that a good encryption system must be extremely sensitive to the key of encryption, a small change of the key leads to great changes of the results. In each encryption process, the chaotic sequences operate with plaintext so we can say cipher text is dependent on plaintext.

3. Statistical analysis

According to the Shneier's theory, any encrypted system can be broken by statistical analysis method [15]. We use chaotic sequence [14] to increase the nonlinear parts of encryption process. So the plaintexts are high diffusing, the analysis method also become very difficult.

IV. RESULTS

The algorithms have been implemented in NETBEANS8.2 IDE and data uploaded on AMAZON AWS S3 s shown in Figure 6. The following parameters have been used:

1. Encryption Time: The time required by the algorithm for processing completely a particular length of data is called the simulation time. It depends on the processor speed, complexity

of the algorithm etc. The smallest value of simulation time is desired.

2. CPU Usage: The amount of CPU memory utilization during the execution of algorithms.

3. Throughput: It depicts the number of bits encrypted per unit time. The Formula is as follows:

$$\text{Throughput} = \frac{\text{Total no of bits}}{\text{Total Encryption Time}}$$

4. Avalanche Effect: A desirable property of any encryption algorithm is that a small change in either the plaintext or the key should produce a significant change in the cipher text. In, particular a change in one bit of the plaintext or one bit of the key should produce a change in many bits of the cipher texts.

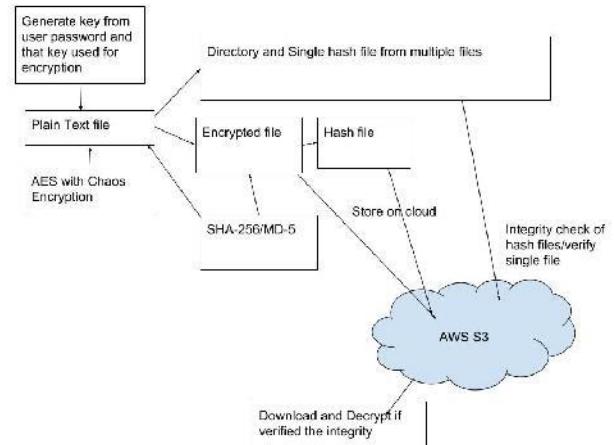


Fig 5:Module Design

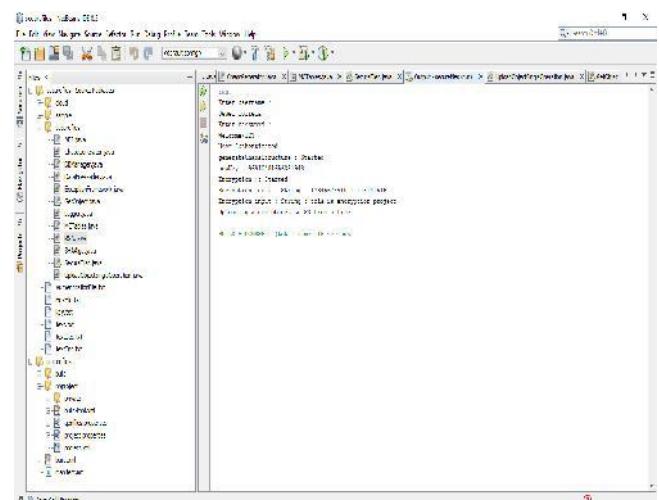


Fig 6: Netbeans IDE 8.2 Working of java code

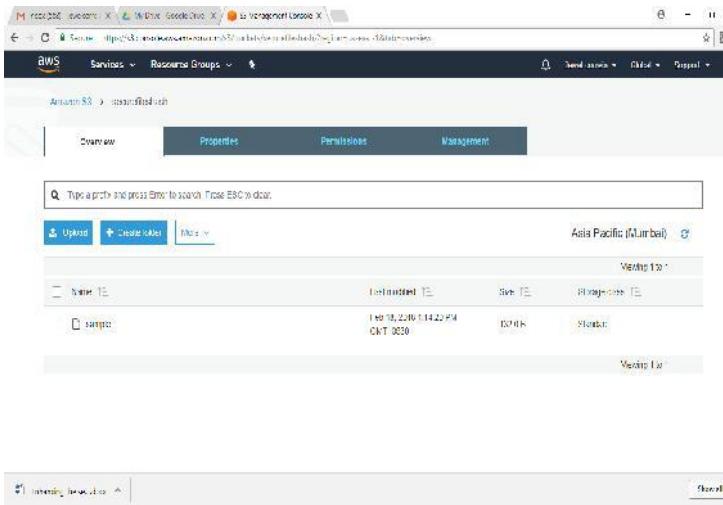


Fig 7: Amazon AWS S3 storage ;file stored through netbeans to cloud

V. CONCLUSION

The proposed system shows how biometrics can leverage the cloud for storage and performance capabilities. The encryption algorithm will prove the security and efficiency of templates. This report describes in detail the process of signatures acquisition through the digitizing tablet, uploading the signature to the AWS cloud and securing the templates using the enhanced encryption algorithm. In this project first of all the module will be developed to secure the cloud storage which meets the secure network coding. The user anonymity will be provided. And the data integrity will be maintained. This all will be done through encryption technique which is enhanced AES-256 with Chaos in round structure. Thus this module will be later implemented on biometric templates to secure them while data sharing in cloud.

REFERENCES

- [1] V.A. Bharadi and Joel Philip, "Signature Verification SaaS Implementation on Microsoft Azure Cloud," in 2016 Proceedings of International Conference on Communication, Computing and Virtualization (ICCV) 2016
- [2] V. A. Bharadi and G. M. D'Silva, "Online Signature Recognition Using Software as a Service (SaaS) Model on Public Cloud," in 2015 IEEE International Conference on Computing Communication Control and Automation, 2015, pp. 65-72.
- [3] Dhvani K Shah, Vinayak Bharadi and Sameer Amrutia, "Implementing Enhanced AES for Cloud based Biometric SaaS on Raspberry Pi as a Remote Authentication Node", IJAIS Proceedings on International Conference and workshop on Communication, Computing and Virtualization, series icwccv2015, no.2, February 2015
- [4] Fei Chen, Tao Xiang, Yuanyuan Yang, and Sherman S. M. Chow, "Secure Cloud Storage Meets with Secure Network Coding", in 2015 IEEE Transactions on Computers
- [5] A. S. Bommagani, M.C. Valenti, A. Ross, "A Framework for Secure Cloud-empowered Mobile Biometrics", Proc. of IEEE Military Communications Conference (MILCOM), (Baltimore,MD),October2014
- [6] A. A. Pawle and V. P. Pawar, "Face Recognition System (FRS) on Cloud Computing for User Authentication," no. 4, pp. 189–192, 2013.
- [7] by Yuan Kun, Zhang Han, Li Zhaohui , "An improved AES algorithm based on chaos" in 2009 at International Conference on Multimedia Information Networking and Security.
- [8] William Stallings, "Cryptography and Network Security", 4th Edition
- [9] Shannon C E, Communication Theory of Secrecy Systems, bell Systems Technical Journal,1949:28:656-715
- [10] Feng Guodeng, Pei Ding Cryptography introduction .BeiJing Science Press 1999
- [11] Nicolas T. Courtois and Josef Pieprzyk,Cryptanalysis of a block cipher with over defined systems of equations
- [12] S. Murphy, M.J.B Robshaw, "Essential Algebraic Structure within the AES", Advances in Cryptology CRYPTO 2002, Vol. 2442 of Lecture Notes in Computer Science, Springer-Verlag, August 2002
- [13] H. Nover, "Algebraic Cryptanalysis of AES: Overview", University of Wisconsin, USA, 2005.
- [14] Bruce Schneier .Applied Cryptography :protocol Algorithms, and source code in C.John Wiley& Sons,Inc, 1996
- [15] L Kocarev.Chaos-based Cryptography: a Brief overview. IEEE Circuits and Systems Magazines, 2001:1(3):6-22
- [16] M.B. Vishnu, S.K. Tiong. Security Enhancement of Digital Motion Image Transmission Using Hybrid AES-DES Algorithm, APCC2008 OF IEICE ,2008
- [17] Yuan Kun, Zhang Han Li Zhaohui. An Improved AES algorithm based on chaos, International Conference on Multimedia Information Networking and Security 2009
- [18] Goce Jakimoski and Ljupčo Kocarev, Chaos and cryptography: Block Encryption Ciphers Based on Chaotic Maps.
- [19] Zhao Rui. Wang Qingsheng. Wen Huiping. Design of AES Algorithm Based on Two Dimensional Logistic and Chebyshev Chaotic Mapping 2008
- [20] Nadeem, A. and Javed, M.Y., "A Performance Comparison of Data Encryption Algorithms," IEEE Information and Communication Technologies, 2005.
- [21] Vikas Kaul, Neha Jain, "Security Enhancement of Data Transmission for Next Generation networks," IJCA 2012.

Chat-Bot for Banking Industry

Meghna Singh

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Abhileen Pandey

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Raj Singh

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Pranjali Kasture

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Abstract-A Chat-bot is a computer program that you can talk to through messaging apps, chat windows or increasingly by voice. The Chat-bot replies using the same application creating a back and forth conversation. Chat-bots can now allow 24/7 customer service on simple matters, initiating fund transfers, set up recurring payments, check bank statements and figure out a customer's spending habits among other things. The main purpose of this project is to analyze the basic queries that a customer has and asks on daily or routine basis and to provide the bot with those set of queries

Keywords: Chat-bot, fintech world.

I. INTRODUCTION

Nowadays we have come across huge number of online applications developed for fulfilling various needs of user. For complaining about any problem we have many website. The problem that the user face is that they need to inform their higher authority or contact admin department wait for until they are free. There no other means of getting collective information of the desired problem of it being resolved. If a problem isn't resolved the customer goes through the same process again and time is wasted. Now if the problem is not resolved then the complaint lodger recommends another set of management staff.

The introduction of IOT and the AI has worked like a charm for many corporations. The advancement in technology has served almost every industry in an unmatched and flawless manner. Their behavior and ability to build a firm command over analytics and consumer behavior not only make them stand ahead in the competition but also your organization.

Watson is a question answering (QA) computing system that IBM built to apply advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering.

The key difference between QA technology and document search is that document search takes a keyword query and returns a list of documents, ranked in order of relevance to the query (often based on popularity and page ranking), while QA technology takes a question expressed in natural language, seeks to understand it in much greater detail, and returns a precise answer to the question.

SQL server is a relational database management system contained in a C programming library. In contrast to many other database management systems, SQL server not a client-server database engine. Rather, it is embedded into the end program. SQL server is ACID-compliant and implements most of the SQL standard, using a dynamically and weakly typed SQL syntax that does not guarantee the domain integrity. So by using this software and languages, we are able to develop the portable efficiently and without errors.

The day is not far where these chat-bots would become full-fledged financial advisors which will not only help customers make informed decisions, but also help them improve their financial planning.

II. BACKGROUND

Financial institutions are notoriously slow at embracing new technologies, and are struggling to keep pace with changing customer expectations. With more competition coming from non-traditional players e.g. digitally native fin-tech start-ups, banks need to rethink their legacy products. Banks are recognizing that their systems and applications lack this "light feel" as well as the ease and convenience of use that their customers now demand. The more progressive banks are already using bots to extend their customer service to customer-preferred channels, such as Facebook Messenger, WhatsApp, etc. Others are looking at options for enabling transactions in messaging platforms. Banking applications are typically clunky, complex and frustrating for the users.

Banks need to be able to answer customers' questions within communication mediums their customers are already using, without them having to log into a website or an app each time they want to view recent transactions or check their balance. By using chat bots, banks can alleviate a lot of the complexity.

III. MOTIVATION & SCOPE

In today's world driving itself with the booming advancement in the field of technology and its merger with almost each and every industries and sectors, the finance sector has recently felt the urge to drive its industry towards automation for better and faster customer service and to be one of the best service providers in the fintech world. Chat-bot provides an easier medium for the banking customers to get their queries solved without wasting much time

The main purpose of our project is to analyse the basic queries that a customer has and asks on daily or routine basis and to provide the bot with those set of queries. Through the survey we found that there is need to develop such software which will be used by the banking industry and their customers effectively. And this motivated us to develop a system which will fulfil the needs of the fintech world.

The IT department using our system would be able to provide the banking industry, an easier and faster medium to solve the queries of their customers and to provide instant suggestions and advices using an intelligent bot system. This software would be useful for any banking company that is keen to improve the customer relationship with their bank and will be able to move their bank towards the field of automation.

IV. PROBLEM DEFINITION

We aim to provide the banking industry with an easier medium for their customers to interact in the easiest manner without downloading any special application, which is eventually going to be more swift for the customers to reach out for any kind of financial information like bank statement, credit card bill etc.

A. Phase 1

Analysis and Planning

We have taken reference of 4 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers are secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and is found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

B. Phase 2

Designing and implementation

The intelligent tree algorithm covers both **classification and regression**. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision-making. As the name goes, it uses a tree-like model of decisions.

IBM Watson platform will be used to develop the front end of the Chat-bot system for the customers of any desired bank.

C. Phase 3

Testing and Deployment

To check if the chat-bot system is able to provide the correct solution to customer's/user's query, and give proper advice/options for further actions and policy changes.

V. METHODOLOGY

A user will send a message from the registered mobile number to the bot system using a messaging application.

1. The Bot system will verify the registered mobile number and the necessary credentials provided by the end user in the Data Warehouse of the bank and its customers.
2. On successful verification from the data warehouse, the user can access the bot's services as much as it is wanted.

Figure 1 is a flowchart diagram. Flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem.

VI. FEASIBILITY STUDY

Technical Feasibility: The system works on the app model and is the system implemented using Java, C#, etc. as programming and scripting languages. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms.

Economic Feasibility: The cloud owner has a more than one clouds to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing application to manage a cloud environment which has more than one cloud. Cloud computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Cloud. Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

Legal Feasibility- The project is developed under legal license of Windows and other software's and databases. The data will not be used for some illegal purpose and will be encrypted using algorithms.

Operational Feasibility- An overview of the process portrayed that the proposed Chat bot would be beneficial by providing a unique concept to users hence being the only player in the market.

Social Feasibility- The Application allows users to chat with each other giving a sense of social connect before actually doing through the rent agreement. This offers another aspect of the feasibility.

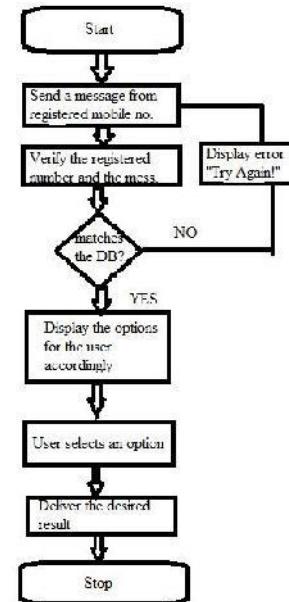


Fig. 1 Flow of Proposed System

VII. CONCLUSION

Application development is defined as a very useful and efficient development that fulfills the required needs as per banking industry. In the world of entertainment, Facebook Messenger chatbot games are very attractive and popular. In the News Industry, we've done an extensive research on chatbot usage. In this section we demonstrated example of convenient bank chatbots. It will save the time and money of the customer. In addition, bank chatbot can be your personal Customer Service. Chatbots successfully serve their clients. Chatbots are on the stage of development in Real Estate Industry also. In the nearest future, they will be doing a lot of agent's functions.

Most people prefer to engage with programs that are human-like, and this gives chatbot-style techniques a potentially useful role in interactive systems that need to elicit information from users, as long as that information is Relatively straightforward and falls into predictable categories. Thus, for example, online help systems can usefully employ chatbot techniques to identify the area of help that users require, potentially providing a "friendlier" interface than a more formal search or menu system. This sort of usage holds the prospect of moving chatbot technology from Weizenbaum's "shelf ... reserved for curios" to that marked "genuinely useful computational methods".

REFERENCES

- [1] Anirudh Khanna, Bishwajeet, "A study of today AI through chat bots", International journal for u and e service, July2015
- [2] Mr. George, Okeyo, "Framework of improving learning management system", International Journal of Computer Science, May2017
- [3] Mr. Aniket Dole, "Intelligent chat bot for baking application system", IJETTCS journal, October 2015
- [4] Mr. KinerB.Shah, "Approach towards building a banking assistant", International Journal of Computer Science, May2011
- [5] Computer History Museum (2006), "Internet History—1970's", Exhibits, Computer History Museum, archived from the original on 2008-02-21,retrieved 2008-03-05
- [6] Güzeldere, Güven; Franchi, Stefano (1995-07-24), "Constructions of the Mind", Stanford HumanitiesReview, SEHR, Stanford University, 2008-03-05
- [7] Mauldin, Michael (1994), "ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition", Proceedings of the Eleventh National Conference on Artificial Intelligence, AAAI Press, retrieved 2008-03-05 (abstract)
- [8] <https://www.chatbots.org/paper/>

RFID Based Attendance Management System

Shashank Shukla
Senior Software Engineer
Tech Mahindra
New Jersey, USA
kshashank343@gmail.com; ss6779@att.com

Abstract— Radio-frequency identification (RFID) is a technology that uses radio waves to transfer data from an electronic tag, called RFID tag or label, attached to an object, through a reader for the purpose of identifying and tracking the object. RFID technology is a matured technology that has been widely deployed by various organizations as a part of their automation systems. In this project, an RFID based system has been built in order to produce an attendance management system. An automated attendance management software will not only make the entire process simple, but will also provide a well-structured and analyzed report of the pattern of student attendance and time management, which can further help in allocating and using the human resources in an organization to the maximum possible benefit. This system consists of two main parts which include: the hardware and the software. The hardware consists of a motor unit and RFID reader. The RFID reader, which is a low-frequency reader (125 kHz), is connected to the host computer via a serial to USB converter cable. The Time-Attendance System GUI is developed using Visual Basic.Net. The Attendance Management System provides the functionalities of the overall system such as displaying live ID tags transactions, registering ID, deleting ID, recording attendance and other minor functions. This interface was installed in the host computer.

Keywords— *RFID, Automated, Tracking, Attendance*

I. INTRODUCTION

The existing conventional attendance system requires students to manually sign the attendance sheet every time they attend a class. As common as it seems, such system lacks automation, where a number of problems may arise. This includes the time unnecessarily consumed by the students to find and sign their name on the attendance sheet; some students may mistakenly or purposely sign another student's name. Also the attendance sheet may get misplaced [1].

Having a system that can automatically capture student's attendance by flashing their student card at the RFID reader can really save all the mentioned troubles. This is the main motive of our system and in addition having an online system accessible anywhere and anytime can greatly help the lecturers to keep track of their students' attendance. Looking at a bigger picture, deploying the system throughout the academic faculty will benefit the academic management as students' attendance to classes is one of the key factor in improving the quality of teaching and monitoring their students' performance. Besides, this system provides valuable online facilities for easy record

maintenance offered not only to lecturers but also to related academic management staffs especially for the purpose of students' progress monitoring [2-3].

II. LITERATURE SURVEY

In the process of system development, literature reviews conducted to understand the theory, methods and technologies associated with systems that have been developed. Background research on the organization and comparative studies of existing systems is also done to understand the system requirements before the system was developed [4]. Student Attendance Using RFID System is an automatic record of student attendance developed especially for universities. Figure 1 shows overall block diagram of the system [5].

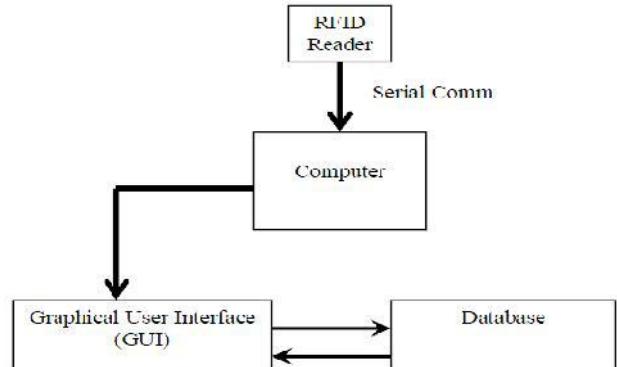


Figure 1: Block Diagram

It's generally said that the roots of radio frequency identification technology can be traced back to World War II. The Germans, Japanese, Americans and British were all using radar which had been discovered in 1935 by Scottish physicist Sir Robert Alexander Watson-Watt to warn of approaching planes while they were still miles away. The problem was there was no way to identify which planes belonged to the enemy and which were a country's own pilots returning from a mission. Radio Frequency Identification (RFID) research and discovery began in earnest in the 1970s. RFID is commonly used to transmit and receive information without

wires. RFID readers and tags communicate through a distance using radio waves. There are a lot of advantages in RFID system, included their price, size, memory capacity and their capability. Advances in radar and RF communications systems continued through the 1950s and 1960s. Electronic article surveillance tags, which are still used in packaging today, have a 1-bit tag. The bit is either on or off. If someone pays for the item, the bit is turned off, and a person can leave the store. But if the person doesn't pay and tries to walk out of the store, readers at the door detect the tag and sound an alarm [5].

The First RFID Patents Mario W. Cardullo claims to have received the first U.S. patent for an active RFID tag with rewritable memory on January 23, 1973 [5]. Later, companies developed a low-frequency (125 kHz) system, featuring smaller transponders. A transponder encapsulated in glass could be injected under the cows' skin. This system is still used in cows around the world today. Low frequency transponders were also put in cards and used to control the access to buildings [5].

Today, 13.56 MHz RFID systems are used for access control, payment systems (Mobile Speedpass) and contactless smart cards. They're also used as an anti-theft device in cars. A reader in the steering column reads the passive RFID tag in the plastic housing around the key. If it doesn't get the ID number it is programmed to look for, the car won't start [5].

In the early 1990s, IBM engineers developed and patented an ultra-high frequency (UHF) RFID system. UHF offered longer read range (up to 20 feet under good conditions) and faster data transfer. IBM did some early pilots with Wal-Mart, but never commercialized this technology. When it ran into financial trouble in the mid-1990s, IBM sold its patents to Intermec, a bar code systems provider. Intermec RFID systems have been installed in numerous different applications, from warehouse tracking to farming. But the technology was expensive at the time due to the low volume of sales and the lack of open, international standards [5].

III. DESIGN, IMPLEMENTATION & INSTALLATION

Since the system is based on RFID, we do have some hardware component in the application. To keep the project simple and more software based, we are not designing any hardware architecture to the software system. We are going for basic hardware used as an interfacing device to read the data from the card and store it on application's database.

The Figures 2 and 3 will give a clearer view of a basic RFID Scanner & Card:



Figure 2. RFID Scanner



Figure 3: RFID Card

The task of the scanner is to read a ten digit unique RFID tag and feed it to any display device or visible fields on the screen. So each card has a film with a 10 digit unique RFID and the number is printed on the card itself for recognition purpose, as shown in Figure 4.

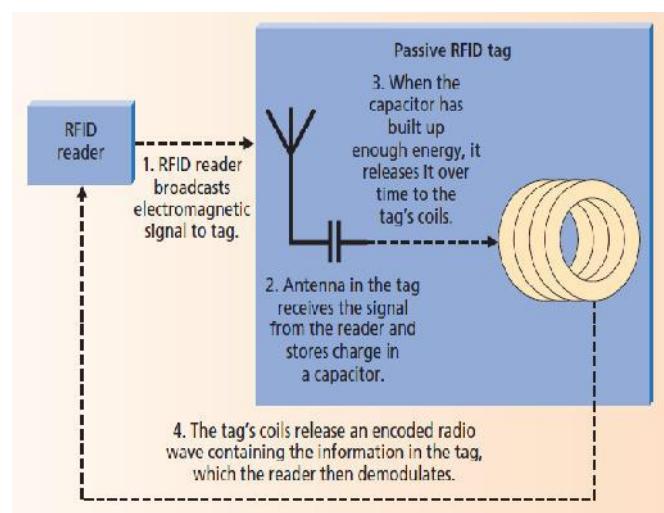


Figure 4: Simplified view of data transfer in low frequency passive RFID tags

A. Technical Specifications

System Requirements

Server:

Hardware

- Processor: Core 2 Duo Processor 2.4 GHz or Higher

- Memory: 2 GB RAM or Higher
- HDD Storage: 80 GB or Greater
- Software
- Microsoft Windows Server 2003 or 2008
- IIS (Internet Information Server)
- .NET Framework 2.0

Client:

 Hardware

- Processor: Pentium 4 Processor at 2.2 GHz or Higher
- Memory: Minimum 256 MB RAM
- HDD Storage: Minimum 10 GB Space
- Software
- Web Browser (Mozilla Firefox, Internet Explorer 6.0+, etc.)

B. Block diagram & Network diagram of the system

The application is a web based application. So when we deploy it will have a server where all the web pages reside. All the data is stored in the database. The Client accesses this data using internet. The database is accessed via Server and the application works in the browser. To access the application in the browser, the user must have a valid RFID card (which is analogous to the one shown in the diagram above) and he needs to swipe it in front of the RFID Scanner. The block diagram of system is shown in Figure 5. RFID Scanner being plug and play can be replaced very easily without affecting the current application and there would be no data losses as there is no hardware storage involved in terms of the RFID Scanner. Moreover the RFID Scanner and the cards are cost effective.

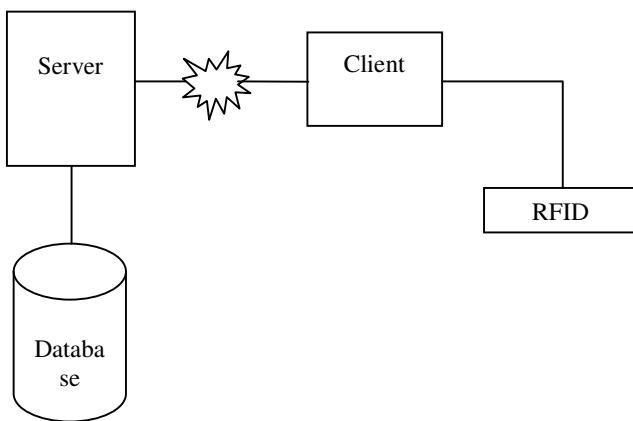


Figure 5: Block Diagram of System

So as far as the system implementation is concerned, we have to develop a web application in ASP.net with SQL Server as the backend. The RFID scanner acts as a tool to provide input in the form of attendance to the system. The advantage here is that the storage and intelligence is handled by the system itself; so even a basic scanner will work perfectly alright for the system.

The Block Diagram consists of 4 blocks. Each of them is explained below.

- RFID Scanner

RFID Scanner is a plug-play USB device that has only one task – If a valid task is swiped in front of the scanner, the RFID tag is read and its unique code is pasted on any text field on the active window of the screen.

- Database

Database of the application is MS SQL Server. So the database file is of type .mdf. Database would be MS SQL Server 2008 Management studio compatible. It'll be supplemented by Microsoft ASP.net Membership Provider.

- Client

On the client side the application is just loaded as a website. So on the client side, the only role is of the browser. The application runs in the browser with proper authentication and authorization.

- Server

Server is where all the web-pages reside. It consists of all the codes into two sets. One is the design side of the code and the other one is the actual coding. So page extensions are .aspx and .aspx.cs. The directory is divided accordingly for protection and security. Whenever, the client requests a data, the server serves it. If at all, the data is required, then it'll retrieve from the database and send it as http stream to the browser. At the browser level, the product turns into html for rendering the data.

C. Network Diagram of the system

Flow of information from one block to another is shown in Figure 6.



Figure 6: Network Diagram

IV. RESULTS

Homepage

New Student page

Log In page

Attendance page

The figure displays four screenshots of the 'COLLEGE ATTENDANCE CONTROL' system:

- Homepage:** Shows the main landing page with the title 'COLLEGE ATTENDANCE CONTROL' and a welcome message: 'WELCOME TO COLLEGE ATTENDANCE CONTROL SYSTEM'. It features a large image of a college building.
- New Student page:** Shows a form titled 'Add Student Details' with various input fields for student information such as Name, Address, Date of Birth, etc.
- Log In page:** Shows a login form with fields for 'Username' and 'Password'. There is also a checkbox for 'Keep me logged in' and a 'Log In' button.
- Attendance page:** Shows a search interface for 'Attendance' with dropdown menus for 'Select Department' and 'Select Subject'.

Check attendance Page

This screenshot shows the 'Check Attendance' page of the College Attendance Control system. It features a search form with fields for 'Enter Roll No.' and 'OR Scan RFID Card.', and dropdown menus for 'Select Department', 'Select Device', 'Select Device', and 'Select Date'. Below the form, a message says 'No records found.' A table displays student information with columns for 'Roll No.', 'Name', and 'Percentage'. One row is highlighted with a yellow background.

Student Profile Page

This screenshot shows the 'Your Profile' section of the student profile page. It includes a thumbnail photo, personal information like 'Name: Shashank Shukla', 'Department: Information Technology', and 'Branch: Game'. It also lists academic details such as 'CGPA: 9.000', 'Attendance: 100%', and 'Grade: A+', along with other course and grade information.

My Attendance Page

This screenshot shows the 'My Attendance' page. It displays a table of attendance records with columns for 'Date', 'Present', 'Absent', and 'Total'. The data shows attendance for various dates from October 2012 to November 2012, with counts ranging from 40 to 45.

Search by RFID Page

This screenshot shows the 'SEARCH USING RFID' page. It has a single input field labeled 'Enter Roll No.' and a 'Search' button below it. The page footer includes copyright information for 2012-2013 College Attendance Control System.

Successful Mail & SMS delivery

This screenshot shows a desktop environment with several windows open. In the foreground, a small window says 'Successful Mail & SMS delivery'. Behind it, a Google search results page shows a link to 'http://www.sendgrid.com'. Another window in the background displays a list of messages or logs.

Student Homepage

This screenshot shows the 'Welcome to College Attendance Control System' page. It features a large image of a modern college building with multiple wings and a central entrance. Navigation links include 'Home', 'My Profile', 'My Marks', and 'My Attendance'. The page footer includes copyright information for 2012-2013 College Attendance Control System.

Mail & SMS received

This screenshot shows a mobile phone screen displaying a text message. The message is from 'BPS-080001' and reads: 'BPS-080001: Hello Shashank Shukla, Your attendance is low (60%) in Game Application 12.25 PM'. Below the message is a text input field with placeholder text 'Add text...' and a 'Send' button.

Figure 7: Some captures of the RFID Based Attendance Management System

V. CONCLUSION & SCOPE FOR FUTURE WORK

In conclusion, the objective to build an RFID based attendance system was successfully achieved. In terms of performance and efficiency, this project has provided a convenient method of attendance marking compared to the traditional method of attendance system. By using databases, the data is more organized. This system is also a user friendly system as data manipulation and retrieval can be done via the interface, making it a universal attendance system. Thus, it can be implemented in either an academic institution or in organizations.

However, some further improvements can be made on this RFID in order to increase its reliability and effectiveness. An indicator or an LCD screen can be incorporated into the system to indicate when any unregistered card is scanned. An IP camera can be integrated into this system to monitor the actions like buddy-punching wherein a person cheats by scanning for another person. Finally, this attendance system can be improved by adding a feature where the attendance system indicates when a student is late for work or classes as the case maybe.

Acknowledgment

With deep sense of gratitude I would like to thank all the people who have lit my path with their kind guidance. I'm very grateful to these intellectuals who did their best to help during my paper. It is my proud privilege to express deep sense of gratitude to my B.E Project Guide Prof. Rajesh S. Bansode for his continuous guidance and support throughout my project and paper. It would never have been possible for me to complete this project successfully without his guidance and support. I remain indebted to my H.O.D, Dr. Kamal Shah, IT Department, Principal, Dr. B.K. Mishra for their comments and kind permission to complete this paper. I would also like to thank them for their timely suggestions and valuable guidance. And lastly I would like to thank our friends and the people who are directly or indirectly related to my paper.

REFERENCES

- [1] Ononiwu G, Chiagozie, Okorafor G. Nwaji. "Radio Frequency Identification (RFID) Based Attendance System With Automatic Door Unit". in Academic Research International, ISSN-L : 2223-9553. 2012; 2(2).
- [2] Zatin Singhal and Rajneesh Kumar Gujral. "Anytime Anywhere-Remote Monitoring of Attendance System based on RFID using GSM Network". in International Journal of Computer Applications (0975-8887). 2012; 39(3).
- [3] Herdawatie Abdul Kadir, Mohd Helmy Abd. Wahab, Zarina Tukiran, Ariffin Abdul Mutalib. "Tracking Student Movement using Active RFID". in 9th WSEAS International Conference, ISSN : 1790-5117.
- [4] Elisabeth Ilie-Zudor, Zsolt Kemeny, Peter Egri, Laszlo Monostori. "The RFID Technology and its Current Applications" in MITIP-2006.
- [5] Mohd. Firdaus Bin Mahyidin. "Student Attendance Using RFID System". in University Malaysia, Pahang, May-2008.

*International Conference on
Advances in Optimization of
Decision Support Systems*

Soft Computing Based Empirical Analysis for Detection of Occlusion in Face Images

Rohit Raja

*Department of Computer Science and Engineering
Sreyas Institute of Engineering and Technology,
Hyderabad, India
drrohitraja1982@gmail.com*

Rajkumar Patra

*Department of Computer Science and Engineering
CMR Technical Campus,
Hyderabad, India
patra.rajkumar@gmail.com*

G. R. Sinha

*Department of Electronics & Communication Engineering
Myanmar Institute of Information Technology (MIIT),
Mandalay
gr_sinha@miit.edu.mm*

Su Su Hlaing

*Department of Computer Science and Engineering
Myanmar Institute of Information Technology (MIIT),
Mandalay
su_su_hlaing@miit.edu.mm*

Abstract— In the field of face recognition for detecting partial occlusion like glasses, beard, hairs, scarf or other accessories, a lot of research is being carried out. However, the problems like illumination field, facial expression, pose estimation and partial occlusion are not addressed in robust ways. This work includes enough human face image data with partial occluded image has been collected for detection of various types of occlusion. To support this process side-view of human face image with occlusion have been considered and a sufficient amount of features have been extracted. For face recognition of different type of occlusion, progressive switching pattern is used. Automatic Face Recognition (AFR) is used in detection of a partial occlusion is very difficult unless it has been brought closer to one's perception. Less amount of work has been carried out using soft-computing techniques. The present paper deals with an empirical analysis of side face image which is parallel to the surface plain using soft-computing technique of Artificial Neural Network (ANN). A robust object model is proposed using ANN and segmentation of the side face image and also using connected component method.

Keywords- Artificial Neural Network (ANN), Discrete Cosine Transform (DCT), SCOD (Soft-Computing Based Occlusion Detection) model, Connected Component method.

I. INTRODUCTION

Occlusion in context of facial imagery refers to the obstruction posed by wearable glasses, scarfs, hairs, or other accessories. Efforts made on face recognition under suitable condition has already been achieved in past studies but thus far the problems like illumination field, facial expression, pose estimation & partial occlusion has been the current concern of research for effective face recognition under varying condition. Part Based Methods include Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA), Non-negative Matrix factorization (NMF), Local Non-Negative Matrix Factorization (LNMF), Independent Component Analysis (ICA) and the other variations of it has

shown better recognition rates. Also, other methods like feature based method or fractal method usually considers the features around the essential facial parts like that of eyes, ears, or mouth region is incorporated in the recognition phase of this algorithm; though such methods show diverse set of analysis results. Occlusion is divided into two categories i.e., natural occlusion and synthetic occlusion. Natural occlusion refers to the non-intentional blockade of view of the objects whereas the synthetic occlusion to the intently posing of artificial obstruction between the views. Partial occlusion has sometimes severely affect the image processing of biometrics as it tends to disrupt the identification of the particular image. Tan, et al. (2009) has developed this method there was an assumption taken in advance before experimenting that the given image is occluded and the reference image is available with the probe image being locally occluded under the condition of partial similarity.

Tan et al. (2006) had presented a concept of eliminating the non-matching details from the non-parametric partial similarity between two sets of facial images; which helps in extracting interpersonal details. For Selective Local Non-Negative Matrix Factorization (S-LNMF) based method relies over divided PCA and 1 NN classifier employed to detect distortions, where LNMF was used for comparison over the same AR face recognition datasets cropped to 64 x 88 pixel sets has developed by Oh H. (2008).

Since, the probability is measured to find the correct image; consequently it deals with imprecise location and partial occlusions (Martinez et.al, 2000). In this type of methods the image is divided into distinct local parts and is then analyzed in isolation with the Eigen-spaces. This allows the learning of alignment of subspaces in the controlled lighting environment or occlusion. The experiment is carried over the A facial datasets for 100 greyscale images corresponding to 50 men & 50 women of size 120 x 170 pixels; giving the recognition rate of a range around 85-95% & ~80% for scarf occluded and sunglasses occluded images respectively. Another similar attempt are illustrated in (Farb et al.

sigmoid), O_i is the output of the i^{th} neuron and S_i is the weighted sum of the inputs.

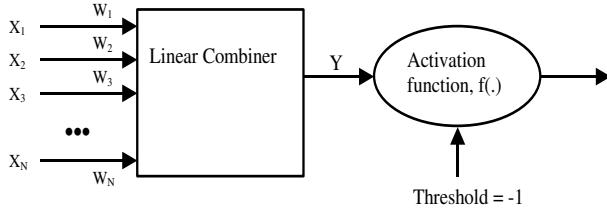


Fig. 2 An artificial neuron

A linear combiner is a function that takes all inputs and produces a single value. Let the input sequence be $\{X_1, X_2, \dots, X_N\}$ and the synaptic weight be $\{W_1, W_2, W_3, \dots, W_N\}$, so the output of the linear combiner, Y :

$$Y = \sum_{i=1}^N X_i W_i \quad (2)$$

An activation function will take any input from minus infinity to infinity and squeeze it into the range -1 to $+1$ or between 0 to 1 intervals. An activation function being treated as a sigmoid function that relates as:

$$f(Y) = \frac{1}{1 + e^{-Y}} \quad (3)$$

The threshold defines the internal activity of the neuron, which is fixed to -1 . In general, for the neuron to fire or activate, the sum should be greater than the threshold value.

AHFM signifies the population of genes or human-face parameters. Based on the assumption that the original image is additive with noise. To compute the approximate shape of the wavelet (that is, any real valued function of time, possessing a specific structure), in a noisy image and also to estimate its time of occurrence, two methods are generally used. The first one is simple-structural-analysis method and the second one is the template-matching method. Mathematically, for the detection of wavelets in noisy image, assume a class of wavelets, $S_i(t)$, $i = 0, \dots, N-1$, all possess certain common structural features. Based on this assumption that noise is additive, then the corrupted image has to be modeled by the equation,

$$X(m,n) = i(m,n) + G d(m,n) \quad (4)$$

where $i(m,n)$ is the clean image, $d(m,n)$ is the noise and G is the term for signal-to-noise ratio control. Next windowing the image and assuming $G = 1$, equation (4) becomes:

$$x_w(m,n) = i_w(m,n) + d_w(m,n) \quad (5)$$

Fourier transform of both sides of equation (5), yields:

$$X_w(e^{j\omega_1}, e^{j\omega_2}) = I_w(e^{j\omega_1}, e^{j\omega_2}) + D_w(e^{j\omega_1}, e^{j\omega_2}) \quad (6)$$

Where $X_w(e^{j\omega_1}, e^{j\omega_2})$, $I_w(e^{j\omega_1}, e^{j\omega_2})$ and $D_w(e^{j\omega_1}, e^{j\omega_2})$ are the Fourier transforms of windowed noisy, original-image and noisy-image respectively. To de-noise this image, wavelet transform has to be applied. Let the mother wavelet or basic wavelet be $\psi(t)$, which yields to,

$$\psi(t) = \exp(j2\pi ft - t^2/2) \quad (7)$$

Further as per the definition of Continuous Wavelet Transform CWT (a, τ) , the relation yields to,

$$\text{CWT } (a, \tau) = (1/\sqrt{a}) \int x(t) \psi\{(t-\tau)/a\} dt \quad (8)$$

The parameters obtained in equation (8) have to be discretized, using Discrete Parameter Wavelet Transform (DPWT). DPWT (m, n) is to be obtained by substituting $a = a_0^m$, $\tau = n \tau_0 a_0^m$. Thus equation (8) in discrete form results:

$$\text{DPWT } (m, n) = 2^{-m/2} \sum_k \sum_l x(k, l) \psi(2^{-m}k - n) \quad (9)$$

where ' m ' and ' n ' are the integers, a_0 and τ_0 are the sampling intervals for ' a ' and ' τ ', $x(k, l)$ is the enhanced image. The wavelet coefficient has to be computed from equation (9) by substituting $a_0 = 2$ and $\tau_0 = 1$.

Further the enhanced image has to be sampled at regular time interval ' T ' to produce a sample sequence $\{i(mT, nT)\}$, for $m = 0, 1, 2, \dots, M-1$ and $n = 0, 1, 2, \dots, N-1$ of size $M \times N$ image. After employing Discrete Fourier Transformation (DFT) method, it yields to the equation of the form,

$$I(u, v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} i(m, n) \exp(-j2\pi(um/M + vn/N)) \quad (10)$$

for $u = 0, 1, 2, \dots, M-1$ and $v = 0, 1, 2, \dots, N-1$

In order to compute the magnitude and power spectrum along with phase-angle, conversion from time-domain to frequency-domain has to be done. Mathematically, this can be formulated as, let $R(u, v)$ and $A(u, v)$ represent the real and imaginary components of $I(u, v)$ respectively. The Fourier or magnitude spectrum, yields to,

$$|I(u, v)| = [R^2(u, v) + A^2(u, v)]^{1/2} \quad (11)$$

The phase-angle of the transform is defined as,

$$\phi(u, v) = \tan^{-1} \left[\frac{A(u, v)}{R(u, v)} \right] \quad (12)$$

Segmentation of an image has to be performed using connected-component method. For mathematical formulation, let 'pix' at coordinates (x, y) has two horizontal and two vertical neighbours, whose coordinates are $(x+1, y)$, $(x-1, y)$, $(x, y+1)$ and $(x, y-1)$. This forms a set of 4-neighbours of 'pix', denoted as $N_4(\text{pix})$. The four diagonal neighbours of 'pix' have coordinates $(x+1, y+1)$, $(x+1, y-1)$, $(x-1, y+1)$ and $(x-1, y-1)$, denoted as $N_D(\text{pix})$. The union of $N_4(\text{pix})$ and $N_D(\text{pix})$, yields 8-neighbours of 'pix'. Thus,

$$N_8(\text{pix}) = N_4(\text{pix}) \cup N_D(\text{pix}) \quad (13)$$

A path between pixels ' pix_1 ' and ' pix_n ' is a sequence of pixels $\text{pix}_1, \text{pix}_2, \text{pix}_3, \dots, \text{pix}_{n-1}, \text{pix}_n$, such that pix_k is adjacent to pix_{k+1} , for $1 \leq k < n$. Thus connected-component is defined, which has to be obtained from the path defined from a set of pixels and which in return depends upon the adjacency position of the pixel in that path. Considering the first frame, with left-face(F_L) at the back and right-face(F_R) at the front, the coordinates with (x, y) for first frame, such that $F_L(x_1, y_1)$ and $F_R(x_2, y_2)$. Thus applying the Manhattan distance measures, the step-length has to be computed as,

$$|\text{Ear} - \text{Nose} - \text{length}| = |x_2 - x_1| + |y_2 - y_1| \quad (14)$$

Considering two measures, Accuracy and Precision has been derived to access the performance of the system, which may be

formulated as,

$$\text{Accuracy} = \frac{\text{Correctly Recognized feature}}{\text{Total number of features}} \quad (15)$$

$$\text{Precision} = \frac{\text{TPR}}{\text{TPR} + \text{FPR}} \quad (16)$$

where TPR = True positive recognition and FPR = False positive recognition.

Further the analysis has to be done for the recognition of behavioral traits with two target classes (normal and abnormal). It can be further illustrated that AHFM has various states, each of which corresponds to a segmental feature vector. In one state, the segmental feature vector is characterized by eleven parameters. Considering only three parameters: the ear-nose-length: distance, mean, and the standard deviation, the AHFM is composed of the following parameters

$$\text{AHFM}_1 = \{D_{s1}, \mu_{s1}, \sigma_{s1}\} \quad (17)$$

where AHFM₁ means an artificial human-face model of the first feature vector, D_{s1} means the distance, μ_{s1} means the mean and σ_{s1} means the standard deviation based on ear_nose_length. Let w_{norm} and w_{abnorm} be the two target classes representing normal behavior and abnormal behavior respectively. The clusters of features have been estimated by taking the probability distribution of these features.

B. Solution Methodology with proposed algorithm

The proposed algorithm called **SCOD (Soft-Computing Based Occlusion Detection) model** is:

- Step1. Read the unknown 90-degree oriented face image.
- Step2. Set the frame counter, fcount = 90
- Step3. Set the flag for best fit as fbest = 1
- Step4. Do while fbest <> 0
- Step5. Read the face_image[fcount]

Input: m × n RGB standard Image

Where, m & n are the row & column of the given image.

Output: rg normalized color scheme

Loop: **for** i to m

 Loop: **for** j:n //RGB normalization to rg scheme

$$r_{i,j} = \frac{R}{(R + G + B)}$$

$$g_{i,j} = \frac{G}{(R + G + B)}$$

end

end

Check & Segment:

if(Y_M ≤ Σ_M P(r_{i,j}, g_{i,j})) //To adjust contrast

P_{skin} = P(skin|rg, N) //for sampling skin texture

else

P_{background} = (background|rg, Y_M)

Enhance the image using DCT

Compute the connected components

Locate ROI and Crop the image

```

    Compute the relevant geometrical features
    Perform the best-fit matching using Genetic algorithm
    Compute the efficiency of matching of parameters If
    true then fbest = 0 and display 'ACCEPT' else display
    'REJECT'
    End do

```

III. RESULTS AND DISCUSSION

In order to form a FACE_MODEL for occlusion, a known face image as switching pattern is depicted in Fig. 3 has been analyzed for the extraction of relevant features. Fig. 3 shows the face image in different pose and angle with different occlusions. Fig. 4 shows Probability distribution of face features extracted from frontal face



Fig. 3 Data set with different pass and angle with different occlusions

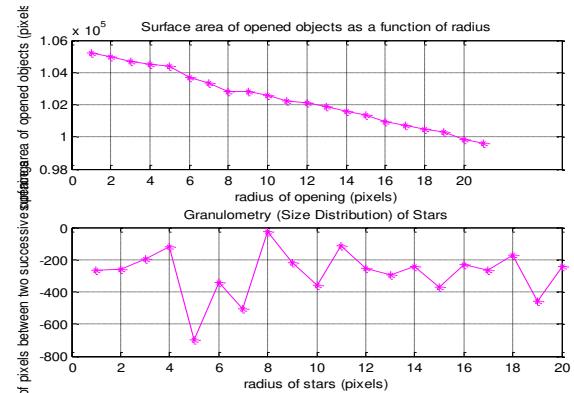


Fig. 4 Probability distribution of face features extracted from frontal face

The Normal Image, Image after preprocessing and Logical micro expression are shown in Fig. 5.

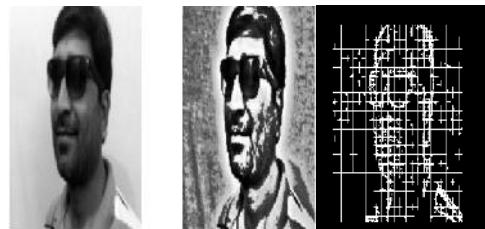


Fig. 5 Normal Image, Image after preprocessing and Logical micro expression

Fig. 6 shows classification of progressive switching pattern using support vector machine and the performance analysis is shown in Fig. 7.

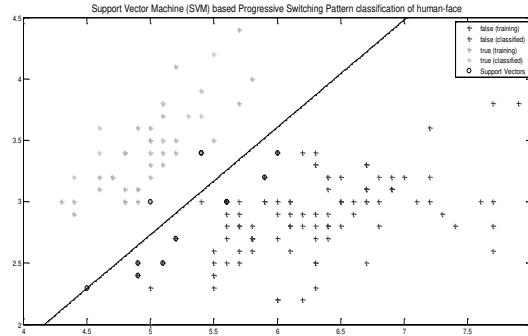


Fig. 6 Classification of progressive switching pattern using support vector machine of artificial neural network of the test image of the human-face captured from side-view with occlusion

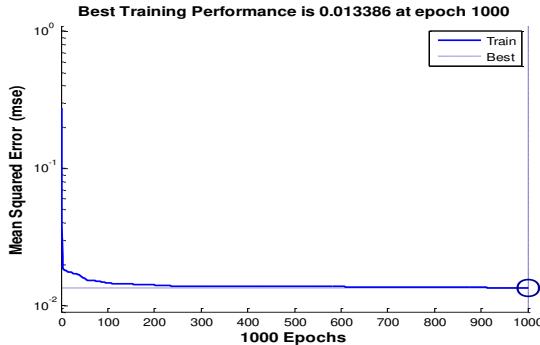


Fig. 7 Performance of training data sets

	SEN	SPEC	ACC	FPR	PPV	PPV	NPV	MCC	FDR
Occlusion 1	0.9231	1.9231	2.9231	3.9231	4.9231	0.9167	0.9167	0.9333	0.9231
Occlusion 2	1	2	3	4	5	0.8125	0.8125	1	0.8667
Occlusion 3	0.9643	1.9643	2.9643	3.9643	4.9643	0.8571	0.8571	0.9643	0.8929
Occlusion 4	1.9231	2.9231	3.9231	4.9231	5.9231	0.1875	0.1875	0	0.1333

Fig. 8 Different Features and Values of Different Occlusion

Fig. 9 shows different features and values of different occlusion whereas Fig. 10 shows its graphical representation.

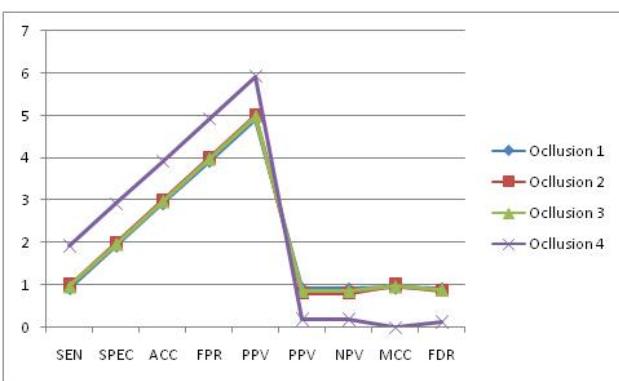


Fig. 9 Graphical representation of divergence values of metrics of the test occluded image of the human-face captured from side-view for different subjects

IV. CONCLUSIONS

In the present paper, the identification of occlusion from side-view with 90-degree orientation has been performed using a proposed algorithm called SCOD Model, which has been tested with the corpus developed and the result has been found very satisfactory. For further study, the proposed algorithm has to be tested with the image of the same subject by changing the getups and extracting the different feature from side-view for biometrical authentication.

REFERENCES

- [1] Tan X., Chen S., Zhou Z., and Liu J., "Face recognition under occlusions and variant expressions with partial similarity," IEEE Transactions on Information Forensics and Security, vol. 4, no. 2, pp. 217-230, 2009.
- [2] Tan X., Chen S., Zhou Z., and Zhang F., "Recognizing Partially Occluded, Expression Variant Faces from Single Training Image Per Person with SOM and Soft kNN Ensemble," in Proceedings of IEEE Transaction on Neural Networks, vol. 16, pp. 1-13, 2005.
- [3] Oh H., Lee K., and Lee S., "Occlusion Invariant Face Recognition using Selective Local Non- Negative Matrix Factorization Basis Images," Image and Vision Computing, vol. 26, no. 11, pp. 1515-1523, 2008.
- [4] Marsico M., Nappi M., and Riccio D., "FARO: Face Recognition Against Occlusions and Expression Variations," IEEE Transactions on System, Man, and Cybernetics-Part A: Systems and Humans, vol. 40, no. 1, pp. 121-132, 2010.
- [5] Abate A., Nappi M., Riccio D., and Tucci M., "Occluded Face Recognition by Means of the IFS," in Proceedings of the 2nd Intrnl. Conf. Image Analysis and Recognition, Berlin, pp. 1073-1080, 2005.
- [6] Martinez A., "Recognition of Partially Occluded and/or Imprecisely Localized Faces using a Probabilistic Approach," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, vol. 1, pp. 712-717, 2000.
- [7] Farb G., Delmas P., Morris J., Shorin A., "Robust Face Matching under Large Occlusions," Proceedings of 14th Intrnl. Conf. on Image Analysis & Processing, Modena, pp. 448-453, 2007.
- [8] Deng Y., Li D., Xie X., Lam K., and Dai Q., "Partially Occluded Face Completion and Recognition," in Proceedings of the 16th Image Processing IEEE International Conference on Image Processing, Cairo, pp. 4145-4148, 2008.
- [9] Yu X., Tian J., and Liu J., "Active Appearance Models Fitting with Occlusion," in Proceedings of the 6th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, Berlin, pp. 137-144, 2007.
- [10] Lee P., Wang Y., Yang M., Hsu J., and Hung Y., "Distinctive Personal Traits for Face Recognition under Occlusion," IEEE Conference on Systems, Man and Cybernetics, Taipei, vol. 5, pp. 4202-4207, 2006.
- [11] Lin J., Ming J., and Crookes D., "Robust Face Recognition with Partially Occluded Images Based on A Single or a Small Number of Training Samples," in Proceedings of IEEE International Conference on Acoustics, Speech & Signal Processing, Taipei, pp. 881-884, 2009.
- [12] Tilendra Shishir Sinha, Devanshu Chakravarty, Rajkumar Patra and Rohit Raja, "Modelling and Simulation for the Recognition of Physiological and Behavioural Traits Through Human Gait and Face Images", Computer and Information Science » Numerical Analysis and Scientific Computing » Discrete Wavelet Transforms-A Compendium of New Approaches and Recent Applications", 2013.
- [13] Rohit Raja, Raj Kumar Patra and Tilendra Shishir Sinha, "Modelling and Simulation of face from side-view for Biometrical Authentication using Hybrid approach", In the proceedings of 4th Intrnl. Conf. on Computer Applications in Electrical Engineering Recent Advances (CERA-09), IIT Roorkee, 2010.
- [14] Rohit Raja, Raj Kumar Patra, Tilendra Shishir Sinha, "Soft-Computing Based Modeling and Simulation For Face Identification From Side-View" International Conference on Systemics, Cybernetics and Informatics, 2010.

A Comprehensive Review and Efficient Implementation of Privacy-Preserving Associative Classification

Darshana H. Patel

V.V.P. Engineering College
Rajkot, Gujarat, India
darshana.h.patel@gmail.com

Avani R. Vasant

Babaria Institute of Technology
Vadodara, Gujarat, India
avani.vasant@gmail.com

Abstract: The enormous development in Information and Communications technology had increased the requirement for digital data to be stock up and communal securely. This immense quantity of data, if publicly available, can be employed for growth and development. However, data in its raw form comprises of sensitive information and advances in data mining techniques have increased the breach of privacy or confidential data. As a consequence, a field of privacy-preserving data mining emerged which deals with efficient conduction and application of data mining functionalities without scarifying the privacy of the data. Several studies in data mining have revealed that Associative Classification is superior to other traditional classification algorithms owing to its numerous favorable characteristics such as readability, usability, training efficient and excellent accuracy. This paper focuses on study of Associative Classification techniques in addition to privacy preserving techniques along with its pros and cons. In addition, related study of privacy preserving associative classification has been presented with an aim of prolific delve in this area. Furthermore, privacy preserving association classification has been implemented utilizing various datasets considering the accuracy parameter and it has been concluded that as privacy increases, accuracy gets degraded due to data transformation.

Keywords: Data Mining, Classification, Associative Classification, Privacy-Preservation

I. INTRODUCTION

With technological revolution, a huge amount of data is being collected and as a consequence data mining technologies are used for the extraction of useful information from huge compilation of digital data. However, this immense quantity of data, if publicly available, can be employed for growth and development as well as in several applications. In the field of data mining, enormous amount of data is processed to acquire minute quantity of fruitful data. Major techniques of data mining are classification, clustering and association [1].

Classification is one of the key techniques in data mining that allocates items in a collection to target class. Various classification methods such as Naive Bayesian classifier, Decision Tree, Neural Network, Associative Classification etc. exists [1]. Amongst

Classification, associative classification technique is simple for end users to understand and easy for prediction due to its various favorable characteristics. But it is apparent that the gathering and analysis of sensitive, personal as well as collective data causes a serious menace to privacy, confidentiality and freedom. Hence, to solve this problem, there subsists a field of privacy-preservation of data.

Privacy-Preserving technique deals with efficient conduction and application of mining without scarifying the privacy the data [2]. Hence, focus is made on the construction of class association rules generated by associative classification and on applying privacy-preserving techniques these rules to prevent its disclosure to uncertified population or nation.

A) Motivation

In the year of 2015, digital India is a campaign launched by the Government of India to ensure that Government services are made available to citizens electronically by improving online infrastructure and by increasing Internet connectivity or by making the country digitally empowered in the field of technology [3]. Due to this digital India campaign, all activities are going to be online and saved somewhere in the database. This data also contains individual's private information. This bulky assortment of data is also transformed into data warehouse, which can then be employed for data analysis purpose to retrieve information patterns. Every coin has two sides, if this information is leaked, it may be misused by the enemies for terrorist activities. So, privacy of this information should be maintained, with the intention that information cannot be misused by unauthorized party.

B) Applications

Privacy-preserving associative classification(PPAC) covers a wide range of applications such as credit card fraud detection, video surveillance, recommended system: product recommendation in online shopping, Stock trading data: finding signals to sell and buy, e-mail authorship attribution phishing detection, etc.[1][2][4].

C) *Generic View of PPAC*

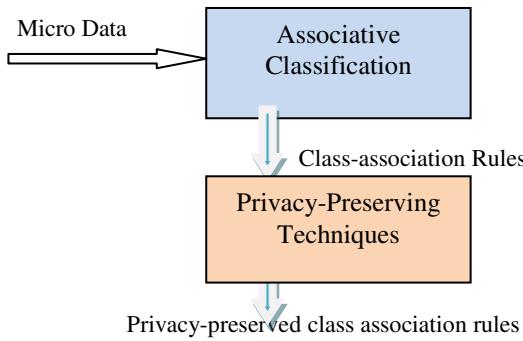


Fig. 1: Generic View of PPAC

Functionalities of data mining are utilized for proper analysis and fetching knowledge from the data for the sake of making certain decisions. Data mining task namely classification is utilized for numerous applications. Furthermore amongst classification techniques, associative classification is preferred due to its favourable characteristics such as readability, usability, training efficient and excellent accuracy.

Micro data i.e. publically available data consisting of sensitive or private information is provided for classification as shown in Fig. 1 to carry out associative classification and class association rules are generated as an output. Then, privacy preserved techniques are applied on such rules to hide or generalize the sensitive or private information and as an outcome privacy preserved class association rules are produced.

II. DATA MINING TASKS

Privacy preserving has been extensively studied by the Data Mining community in recent years. Currently, the PPDM algorithms are mainly used on the functionality of data mining such as Classification, Association and Clustering. [2]

(a) Classification: Classification [1] is the process of finding a set of model that differentiates data classes such that the model can be used to predict the class of objects whose class label is unknown. The common algorithms used for Classification [5] are Decision Tree, Bayesian classification, associative classification etc.

(b) Association Rule: It is a technique in data mining that recognizes the regularities found in bulky amount of data. Such a technique may identify and reveal hidden information that is private for an individual or organization. The common approaches used for Association are a priori, FP-Growth etc. [1]

(c) Clustering: Privacy consideration in data sharing utilizing the clustering functionality is a complex problem[4]. An important task is to protect the primary data values without interpretation of the similarity between objects under analysis in clustering functionality. The common algorithms used for Clustering are partitioning methods, hierarchical methods, grid methods etc.[1]

A) Classification techniques

Major techniques of classification, taken into consideration, are decision tree, neural network, naïve bayes and associative classification.

- (a) Decision trees are found to be flowchart like tree structures, where every internal node denotes a test on an attribute, every branch denotes a result of the test, and every leaf node holds a particular class label [5].
- (b) Neural networks has layers of interconnected nodes, and this node produces a non-linear job of its input and input to a node may arrive from different other nodes or straight from the input data. Also, several nodes are recognized with the output result of the network [6].
- (c) The Bayesian classifier is skilled of calculating the most feasible output based on the input. It takes into consideration, presence (or absence) of a meticulous attribute of a class, which is not related to the presence (or absence) of any other feature when the class variable is known [1, 6].
- (d) Associative classification mining is capable of utilizing the association rule discovery methods to build classification systems, which is also called associative classifiers [7].

Six real datasets for comparison of different classification techniques are taken into consideration. These data sets are available in UCI data repository [22] and its comparison considering accuracy parameter is as follows:

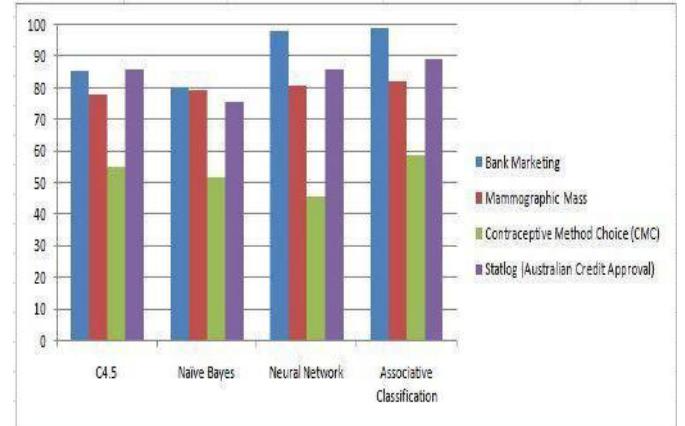


Fig. 2: Accuracy of various classification techniques

Above figure shows the accuracy (in %) of correctly classified instances of different datasets considering accuracy parameter on the various dataset is carried out using WEKA. It is observed that accuracy of associative classification and neural network is higher than any other classification method. However, Neural Network also has low processing speed as shown in Table: 1 and output of Neural Network is not interpretable as compared to the associative classifier which can be easily understood.

TABLE 1: TRAINING TIME OF DIFFERENT CLASSIFICATION TECHNIQUES

Sr. No	Dataset	Decision Tree	Naïve Bayes	Neural Network	Associative Classification
1.	Bank Marketing	0.03	0.05	22.23	0.03
2.	Mammographic Mass	0.05	0.04	77.02	0.02
3.	Contraceptive Method choice (CMC)	0.08	0.08	92.02	0.07
4.	Statlog	1.32	0.89	102.21	0.81

Table 1 demonstrates the resulted training time (in seconds) of decision tree, naïve bayes, neural network and associative classification techniques on the above listed datasets. The training time of neural network is high as it requires more time to learn i.e. more processing time compared to that of naïve bayes and associative classification. As a consequence, associative classification technique has been preferred or taken into consideration.

B) Associative Classification

Associative classification (AC) is a supervised classification method integrating association rule mining and classification [1]. The integration is done in order to get a special subset of association rules whose right-hand is restricted to classification class attribute [7]. These subsets of rules are referred as Class Association Rules (CARs).

Some of the basic techniques for associative classification are namely Classification based on Association (CBA), Classification based on Multiple Association Rules (CMAR), Classification based on Predictive Association Rules (CPAR), Classification TABLE 2: COMPARATIVE ANALYSIS OF ASSOCIATIVE CLASSIFICATION TECHNIQUES

based on Association Rules Generated in a Bidirectional Approach (CARGBA). Each techniques has its own advantages and disadvantages and are thus preferred over each other in view of applications. An comparative analysis of AC techniques are offered for the selection of techniques according to user's requirement.

III. DIFFERENT PRIVACY-PRESERVING TECHNIQUES

The Internet today is a widespread information infrastructure presently allowing many people all over the globe to communicate and better understand each other. However, at the same time, while sharing information private data may be revealed. Thus, the main goal of privacy preserving techniques is to preserve privacy while

maintaining the accuracy. They can broadly be classified into three categories mainly: Anonymization, Randomization and Cryptography [12][13][14].

Ref.	AC Technique	Advantages	Disadvantages
[8]	CBA	-Simple algorithm -More accurate than C4.5	-Redundant rules generation -Over fitting issue
[9]	CMAR	-Compact and efficient storage facility since CR tree is used; -Biasing issue is resolved as it considers multiple rules and weighted chi-square analysis	-Slower compared to CBA -Extra computational overhead due to weighted analysis
[10]	CPAR	-Qualitative rules are generated; -Usage of dynamic programming so works better on large datasets; -Over fitting issue can be resolved	-Complex algorithm -Greedy strategy usage leads to additional computational overhead
[11]	CARGBA	-Anomalies are also considered and exceptional rules are generated to cover it; -Works on heterogeneous data Consistent rules are generated	-Accuracy is less as both general and specific rules are generated; -Consumes much amount of time

(a) *Anonymization*: Anonymization aims to make individual record impossible to differentiate among a group of records by using following techniques [4][12]:

- (1) Generalization and Suppression: Replaces definite values from the original data.
- (2) Anatomization and Permutation: De-associates the correlation between attributes
- (3) Perturbation: Distorts the data by adding noise.
- (4) Bucketization: Separates the SAs (Sensitive attributes) from the QIs (Quasi-Identifiers) by arbitrarily permuting the SA values in each bucket.
- (5) Microaggregation: Unified approach consisting of partition and aggregation.

(b) *Randomization*:

Applied generally to provide estimates for data distributions rather than single point estimates. It distorts values of each attribute in a sample independently and doesn't use information about other samples

(c) *Cryptography*:

Cryptographic techniques are ideally meant for multiparty scenarios where most frequently used protocol is SMC (Secure Multiparty Computation) which mainly constitutes secure sum, secure union, secure intersection, etc. operations.

TABLE: 3 COMPARATIVE ANALYSIS OF PRIVACY PRESERVING TECHNIQUES

Privacy Models	Advantages	Disadvantages
Generalization & Suppression	-Simple Technique -Protects Identity Disclosure -More flexible	- Suffers from homogeneity attack & background knowledge attack - Significant loss of granularity - Not applicable for continuous data. - Suppression complicates analysis
Anatomization & Permutation	-More Accurate than Generalization -Certain aggregate computations can exactly be performed without violating the privacy of the data	- Linking Attack - Cannot be applied to High dimensional data without complete loss of utility - Lacks a formal framework for providing how much privacy is guaranteed
Perturbation	-Attributes are preserved independently -Direct protection for privacy of data is possible due to statistical nature of data mining	-Does not reconstruct the original data values but only data distributions. -Loss of information -Need to develop distribution based algorithm every time
Bucketization	-Used for high dimensional data -Better data utility than Generalization	-Does not prevent membership disclosure -Requires a clear separation between Quasi-identifiers & sensitive attributes
Micro-aggregation	-Unified approach unlike Suppression & Generalization -Reduces the impact on outliers -Reduces data distortion	-Finding an optimal partition in multidimensional micro-aggregation is NP-Hard problem
Randomization	-Simple and easily implemented at data collection phase -Efficient as compared to Cryptography -Doesn't require knowledge of distributions of	-High Information Loss -Cannot be used for multiple attribute databases -Treats all the record equally and reduces the utility of the data

	other records of data - Doesn't require trusted server	
Cryptography	- Offers a well-defined model for privacy - Better privacy as compared to Randomization - Vast toolset of cryptographic algorithms for implementing PPDM	- Complexity increases when more parties are involved - Does not address the question of whether the disclosure of the final data mining result may break the privacy of distinct records - Long Process

IV. RELATED STUDY

With the development of data analysis and processing techniques, the privacy disclosure problem about individual or company is inevitably exposed. This happens while releasing or sharing data to mine useful information, which in turn resulted into the field of privacy preserving data mining. Privacy-preserving data mining (PPDM) is one of the newest trends in privacy and security research.

Privacy issues have posed new challenges for novel uses of data mining technology. These technical challenges cannot simply be addressed by restricting data collection or even by restricting the secondary use of information technology. An approximate solution could be sufficient, where a balance between the need for privacy and knowledge discovery can be maintained. The momentary look of some related work done in privacy preserving associative classification has been offered in below table:

TABLE: 4 LITERATURE REVIEW OF PRIVACY PRESERVING ASSOCIATIVE CLASSIFICATION

Ref	Approach	Advantages	Disadvantages	Open Issues
[14]	Privacy preserving associative classification on vertically partitioned databases	-To transform efficient centralized mining models to work on horizontal and vertical partitioned databases Accuracy high	- Computational overhead increases due to distributed structure	Horizontal partitioning can also be taken into consideration
[15]	Privacy preservation for associative classification	-Proposed frequency based metric outperform as compared to available metric for data quality issue -Larger size of	-Fixed k-value has been considered for the experiments	Heuristics can be applied to efficiently search for near

		quasiidentifier makes generalize datasets harder to link		optimal problem			
[16]	Privacy preservation for associative classification: an approximation algorithm	-The proposed algorithm can transform the given data sets with $O(k \log k)$ -approximation factor with regard to the data utility	-Results vary depending on the types of dataset utilized for performance i.e. it is not necessary that for all types of data, efficient results are achieved.	Stochastic algorithms can be considered.	Classification on Horizontally Partitioned Data	SVM for horizontally partitioned data has comparable accuracy to that of an ordinary SVM	rs on datasets with more variables
[17]	Privacy Preserving of Associative Classification and Heuristic Approach	-Efficient and effective results are achieved using the proposed data quality metric	-The proposed data quality metric is tailored to a specific type of classification.	Other approaches can also be attempted for finding optimal solutions	[21] An Efficient Algorithm for Classification Rule Hiding	-Preserves the privacy of sensitive classification rules as well as maintains quality of the transformed data set also wherein rule based classification algorithms are used for generating rules.	-Overfitting of the data is an issue and it may not work well with small set of training data. Various available other parameters can be taken into consideration for implementation of PPDM
[18]	Incremental privacy preservation for associative classification	-It not only works for static data but also but also for incremental-data scenario wherein data is appended all time. - Computational complexity has been reduced	-Sensitive classification rules are hidden to preserve privacy which degrades the accuracy of data.	Comparable privacy models of k-anonymity like l-diversity can be utilized for preserving privacy			
[19]	Rule Based Privacy Preservation Method for Medical Data Sets	Comparatively better among other existing techniques like ID3 and C4.5	-Data requires to be discretized	Except medical datasets can be apply to other datasets too.			
[20]	Privacy-Preserving	-A novel privacy-preserving	Computationally expensive	applying new classifier			

V. IMPLEMENTATION

Various real datasets available in UCI data repository [22] are taken into consideration for implementation of associative classification technique and its details are described in Table 2 below:

TABLE 5: DATASET COMPOSITION

Name of Datasets	# Attributes	# Instances	# Class
Mammography	6	961	2
Contraceptive Method Choice (CMC)	9	1473	3
Adult	11	30719	2
Bank Marketing	17	45211	2

The work has been implemented using Windows operating system utilizing MATLAB tool. Preprocessing techniques namely discretization and sampling were carried out on these dataset. The initial step for associative classification process i.e. generation of association rule is conceded out using apriori algorithm. Apriori is a kind of association rule generation algorithm for mining frequent item set [1]. Then, class association rules are produced using algorithm namely CBA as a second step of associative classification process by considering the two measures namely support as 0.2 and confidence as 0.6 respectively. The generated set of CARs is evaluated by measuring the accuracy to build the appropriate classifier. Then, the induced classifier will be tested on unseen instances to assess the performance of a classifier. The following table represents the accuracy achieved for various datasets.

TABLE 6: ACCURACY OF VARIOUS DATASETS

Datasets	Accuracy (%)
Mammography	83.745
CMC	70.842
Census Income	72.081
Bank Marketing	75.173

Till now through the above process the class association rules are obtained. For privacy-preservation of this CAR, datafly technique of anonymization has been applied. Now, after applying k-anonymity principal on rules, following results have been obtained using MATLAB tool.

Then, at various privacy levels, accuracy has been inspected considering different database. However, it can be concluded that accuracy decreases with increase in privacy level. i.e. accuracy achieved at $k=2$ for a particular dataset is more than accuracy achieved at $k=4$ and $k=6$ as depicted in below graph.

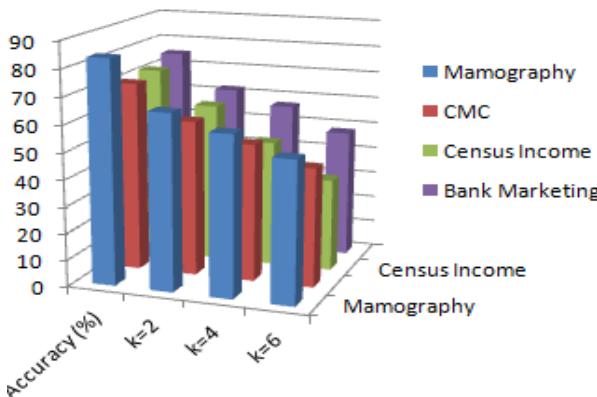


Figure 3: Accuracy and Privacy

Because as data is more transformed, privacy is secured better but at the same time due to transformation of data, accuracy is degraded.

VI CONCLUSION AND FUTURE WORK

In today's world, people are very much concerned about their sensitive information which they don't want to share. Privacy-preserving data mining (PPDM) is one of the newest trends in privacy and security research. This paper covers the fundamentals and basic methods of privacy-preserving and associative classification with its pros and cons. However, it can be concluded that anonymization and CBA technique can be preferred for the implementation of privacy preserving associative classification (PPAC) amongst several existing methods of privacy preserving and associative classification techniques respectively owing to its various favourable characteristics. Further, detailed tabular analysis of related study in privacy-preserving data mining has been carried out for exploring research in this area. In addition, it has been observed by implementing PPAC that as privacy increases the accuracy gets degraded due to data transformation.

As a part of future work, various evaluation parameters with different AC methods can be considered for the accomplishment of PPAC. However, different optimization approaches available in

data mining can be well thought-out for efficient result. Optimization methods essentially includes genetic algorithms, evolutionary search, simulated annealing, branch-and-bound, logical analysis of data, and mathematical programming.

REFERENCES

- [1] J. Han, M. Kamber, J. Pei (© Morgan Kaufmann 2012) 'Data Mining: Concepts and Techniques (3rd edition)' eBook ISBN: 9780123814807
- [2] A. Charu, Y. Philip (2008) 'A General Survey of Privacy-Preserving Data Mining Models and Algorithms', Springer, 11-52.
- [3] Wikipedia, "Digital-India", https://en.wikipedia.org/wiki/Digital_India.
- [4] D. Patel, R. Kotecha (2016) 'Privacy preserving data mining: A parametric analysis', *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, DOI 10.1007/978-981-10-3156-4_14.
- [5] M. Gupta., N. Aggarwal (2010): 'Classification techniques analysis', *National Conference on Computational Instrumentation* at Chandigarh, pp. 128-131.
- [6] S. Nikam ,(2015): 'A Comparative Study of Classification Techniques in Data Mining Algorithms.' in *Oriental Journal of Computer Science & Technology*, vol. 8, pp. 13-19, Techno Research, India.
- [7] B. Liu, W. Hsu, Y. Ma (1998):' Integrating Classification and Association Rule Mining'. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 80-86, AAAI, New York.
- [8] S. Gambhir, N. Gondaliya, (2012) 'A Survey of Associative Classification Algorithms', *International Journal of Engineering Research & Technology (IJERT)*, Vol. 1 Issue 9.
- [9] W. Li, J. Han, J. Pei, (2001) 'CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules' in *Proceedings of the IEEE International Conference on Data Mining*, IEEE, pp369-376.
- [10] X. Yin, J. Han, (2003), 'CPAR: Classification based on Predictive Association Rules', *Proceedings of the SIAM International Conference on Data Mining*, San Francisco; pp 369-376.
- [11] K. Gourab, M. Sirajum, F. Islam, M. Murase, K. Ishikawa, M. Doya, K. Miyamoto, H. Takeshi , (2007) 'A Novel Algorithm for Associative Classification', *International Conference on Neural Information Processing, (ICONIP), Kitakyushu-Japan*, Springer; vol.14, 13-16.
- [12] G. Nayak, S. Devi, (2011),'A survey on privacy preserving data mining: approaches and techniques', *International Journal of Engineering Science & Technology*.
- [13] V. A-Rivera, P. McDonagh, T. Cerqueus, L. Murphy, (2014) 'A systematic comparison and evaluation of k-anonymization algorithms for practitioners'. Transaction Data Privacy, 337-370.
- [14] B. Raghuvaran (2012)' Privacy preserving associative classification on vertically partitioned databases' in

- Advanced Communication Control and Computing Technologies (ICACCCT)*, IEEE; DOI: 10.1109/ICACCCT.2012.6320768.
- [15] N. Harnsamt, J. Natwichai, X. Sun, X. Li: (2014), ‘Privacy Preservation for Associative-Classification’. *Computational Intelligence* 30(4): 752-770 doi: 10.1111/coin.12028.
 - [16] J. Natwichai (2011) ‘Privacy preservation for associative classification: an approximation algorithm’ *International Journal of Business Intelligence and Data Mining; Volume-6 Issue 3*, Pages 283-301.
 - [17] S. Bowonsak, N. Juggapong, H. Nattapon (2010) ‘Privacy Preserving of Associative Classification and Heuristic Approach’, *11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, vol. 00, no., pp. 434-439, doi:10.1109/SNPD.2008.155
 - [18] S. Bowonsak, N. Juggapong (2009) ‘Incremental privacy preservation for associative classification’ *Conference: Proceeding of the ACM First International Workshop on Privacy and Anonymity for Very Large Databases, CIKM-PAVLAD, Hong Kong, China*.
 - [19] B. Murugeshwari and R. Sugumar (2016) ‘Rule Based Privacy Preservation Method for Medical Data Sets’ *Middle-East Journal of Scientific Research* 24 (8): 2640-2648.
 - [20] S. Tian, D. Hua, H. Guoping (2010) ‘Privacy-Preserving Classification on Horizontally Partitioned Data’, *International Conference on Computational Intelligence and security*. DOI: 10.1109/CIS.2010.56.
 - [21] S.Vijayarani and M.Divya (2011), ‘An Efficient Algorithm for Classification Rule Hiding.’ *International Journal of Computer Applications* 33(3):39-45.
 - [22] M. Lichman (2013), UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.

A Hybrid (Two-Tlbo) Meta-Heuristic Optimization Method for Data Clustering Application

Seyed Mohamad bagher davoodi
The Islamic Azad University, Science
and Research Branch,
Iran
mohamadbagherdavoodi@gmail.com

Mehdi Gheisari
School of Computer Science and
Educational Software, Guangzhou
University, Guangzhou,
Guangdong Province, China, 510006
mehdi.gheisari61@gmail.com

Mehrnoosh bazrafkan
Department of Computer, Islamic Azad
University, Science and research
Branch, Tehran, Iran
mehrnoosh.bazrafkan@gmail.com

Abstract— Contemporarily, clustering plays a very important role in many research areas and there are different clustering algorithms which in addition to advantages such as high speed and implementation facility are dropped into the local optima and don't always play optimal solutions. For this reason in this research, we attempted to extract a large percentage of the clustering results out of the local optima and make them approached the global optima by determining the optimal centers of the first clusters and using the proposed algorithm. In this research, the improved weed algorithm is proposed for solving the clustering problem. This algorithm is applicable to solve problems such as np-hard or np-complete. As well, the proposed algorithm is used in order to solve the clustering problems with different evaluation criteria which are tested with several data of the UCI dataset for solving the clustering problem. The experimental results of the clustering problem solving indicate that the proposed algorithm, outperform other algorithms based on different similarity functions and provides the clustering problem with suitable global search and as a result, converges better than other algorithms (invasive weeds, bees and teaching-learning and PSO) to the clustering problem solution.

Keywords — *Optimal clustering, Hybrid meta-Heuristic, Np-hard problem.*

I. INTRODUCTION

Evolutionary Improved Swarm-Based Hybrid K-Means Algorithm for Cluster Analysis: As one of the simple, efficient center based hard clustering techniques, K-means [1, 2] can be applied to different real world problems. There have been key interests among different scholars and researchers during recent decades in different applications such as pattern recognition, applications in biomedicine, image processing, fuzzy logic, cloud computing and so forth. Ensafi et al. [3] made use of fuzzy K-means with PSO in order to detect the network anomaly which resulted in better convergence. [4] Hybridized ACO, PSO and K-means and developed APSO-K-means clustering algorithm for speaker recognition. In another research, Kader [5] proposed a hybrid two-phase GA-IPSO with K-means data clustering algorithm. This algorithm performed faster in clustering of the data and could avoid the occurrence of premature convergence to local optima. Also, Ziangwie and Yuanjiang [6] in order to avoid the local optima in

clustering using K-means developed the improved PSO-based K-means algorithm. For getting optimal centers of clusters for analysis of cluster, Naik et al. [8] proposed a hybrid PSO—K-means clustering algorithm. Govindarajan et al. [9] developed a PSO-based clustering algorithm. Based on cloud computing, this algorithm can improve the quality of students' learning. Liao et al. [10] introduced a parallel K-means algorithm according to with map reduce by reducing the iterations in the algorithm. Bai et al. [11] presented the objects in the cluster centers of K-means algorithm, using local geometrical information. In order to retrieve the large-scale videos, Liao et al. [12] used a sample-based hierarchical adaptive K-means (SHAKM) clustering algorithm. In addition, by Jaganathan and Jaiganesh [13] introduced an improved K-means along with PSO algorithm, for web document clustering. Combining the K-means algorithm and mathematical morphology, an improved K-means method was developed by Yao et al. [14] in order to optimize fish images. Monedero et al. [15] proposed a variation of the celebrated K-means technique for learning in quasi-unsupervised fashion. This was proposed by controlling different cluster partitions size. They adjusted it using Levenberg–Marquardt algorithm. Moreover, Shahbaba and Beheshti [16] provided an introduction to a novel MACE clustering technique which was applicable in synthetic and real data. Naldi and Campello [17] proposed an evolutionary K-means clustering algorithm in order to investigate the distributed data and overcome the limitations of K-means. For case of data points at the border of clusters, Scitovski and Sabo [18] proposed analysis of K-Means algorithm. In addition, Pavithra and Aradhya [19] detected the multilingual text for video indexing by proposing the hybridization of three interesting techniques, wavelet transform, Gabor filter, and K-means clustering method.

In what follows, a literature review is provided for the clustering problem and the studies conducted in this regard are investigated. Then in section 3, the proposed algorithm is presented and a solution is proposed in order to avoid local optima trap of the weed algorithm. Then in section 4, the results of implementation methods are investigated. Finally, the conclusion remarks of this research are presented.

II. LITERATURE SURVEY

As clustering emerges in solution space and increase of the applications to the world as well as significant growth of the information and knowledge, it seems necessary to use this technology for different parts of the society. It can be said that clustering techniques are highly applicable to the pattern recognition

[20]. Clustering is defined as a process in which a set of data are divided into several subsets [21]. These subsets are called clusters. The internal members of clusters are mostly similar to each other and they are the least similar members to the members of other clusters. Unlike classification, clustering is an unsupervised separation process in which usually data are received by vectors in a multidimensional space and they are put into different and separate clusters. K-means clustering is one of these techniques whose algorithm, is applied to the clusters centroids by random selection of K-points from n-points. Based on the Euclidean distance criteria, each data is clustered which is the nearest to the centroid and mean of cluster is the representative. This algorithm iterated until the square scale of error (eq. 1) is minimized. The results is the entire distance of objects from their clusters:

$$E = \sum_{i=1}^K \sum_{p \in c_i} (p - m_i)^2 \quad (1)$$

Where m_i ith cluster centroid and k is the number of clusters

This algorithm has two K-mean advantages [21]

1. It is simply implemented
2. The time complexity is $O(n)$ where n is the data points' number appropriate for huge data. K-mean algorithm has three disadvantages [21]:
1. data-dependent performance.
2. High dependence on initial conditions based on greedy approach. K-mean achieves rapid convergence for optimal solutions.
3. user-based definition of number of classes

There are 2 major faults with K-means: 1) dependence on initial values of clusters and cluster centroid, and easily trapped in local optima and therefore doesn't produce optimal solution. Many researchers conducted studies in order to remove the K-means algorithm, faults by making correct choices of the initial cluster centroids. This is for avoiding production of local optima solutions or is an attempt to find solutions of optimal clustering [21]. In order solve these problems and enhance the efficiency of K-means algorithm, one should use optimization techniques. Many techniques have been proposed for optimization of different problems such as heuristic algorithms. In [22] it is defined that clustering is a NP-hard issue with P class which takes high complexity time to solve the problems. Therefore, for solving the clustering problem with exponential complexity time and Np-hard, similar patterns in data correctly have been recognized. The meta-heuristic algorithms have been proposed for solving these types of problems; since they are capable of solving the complex problems via random search. Some meta-heuristic approaches are: Genetic algorithm, Ant colony (ACO), Particle Swarm, Teaching-learning optimization, Optimization particle swarm, Flower Pollination Algorithm. MR Jensi and Wiselin Jiji [23] used hybrid k-mean algorithm and flower pollination to cluster data. Flower pollination algorithm is a new algorithm inspired by nature according to characteristics of flowering plants. Pollination and intersection operators are applied to enhance the variety of population in local optima and the local search was used to prevent from trapping capability using mutation operator based on improve preference. In this paper, ten datasets are selected and their effectiveness is investigated. The results of simulation indicates that flower pollination algorithm outperforms

the PSO, differential evolutionary, honey bee colony and cuckoo algorithms in achieving better results.

Genetic Algorithm (GA)

Genetic algorithms (GA) [24] are optimization and random search techniques in problem space which is formed in term of the natural selection theory and evolutionary processes and takes the following processes in order to find the best solutions of problem:

Gene's initialization: a set of chromosomes consists of K gene and are dispersed in problem space randomly. Then, based on fitness function, the cost function is calculated for all genes (evaluation phase) and then, based on Roulettewheel the best member of genes is selected and integrated with other genes (crossover phase). Therefore, each gene can only mutate toward another random solution with a contingent function. In fact in a complete chromosome a gene is dislocated with other gene and the mutation operation occurs (mutation phase). Finally, if optimal chromosome obtained, the algorithm ends; otherwise, the best gene selection phase, crossover and mutation iterate until the particular generation in the problem space will achieve the necessary convergence. In [25] Aibinu et al. implemented the GA in order to find the clusters' mean in K-means clustering problem based on Euclidean distance criterion. The authors stated that GA with lower mutation rate outperforms the GA with higher mutation rate. Also, in this research the multidimensional selection for crossover is used using which one can improve the algorithm's convergence rate and the algorithm converge to more optimal solution.

Particle Swarm Optimization (PSO)

The PSO algorithm like swarm intelligence techniques is an optimization technique for finding the global solutions in problem space which was proposed in 1995 by Kennedy and Eberhart as a deterministic search method. PSO algorithm is inspired by the birds motion [26] in which each bird is identified as a particle in the problem space and its knowledge is used to find food based on the best neighbor and the best particle is shared in the problem and next generation is generated with the best global position. Finally, the best global position in problem space is the best fitness in the entire problem calculated and the mentioned position is recognized as the best problem solution. In [27] Shafiq Alam could use the higher adaptability of Evolutionary Particle Swarm and Hierarchical Particle Swarm algorithms in order to find the best solutions in clustering problem. For this purpose, in order to cluster in these 2 algorithms different similarity measurement criterion was used.

In PSO [28] each subject is a particle and the populations are formed of these particles. Solving the problem area I PSO is a relevant way to the problem. In this population, particles working together attempts to find the best solution in search area. In addition, each particle moves based on its velocity. PSO clustering algorithm outperforms the K-means algorithm and FCM for low data sets; but this leads to high repetition and slow convergence for large scale data.

Teaching-learning based optimization algorithm

TLBO algorithm is np-hard problems solving optimization technique [29]. Like other optimization techniques, TLBO is a nature-based algorithm and works based on the effect of teacher's effect on

learning in class. This algorithm used a population of students' solutions for achieving optimal solutions.

The population is considered as a group of learners or students in a classroom. A teacher helps the students to enhance their knowledge level and students would achieve their favorite scores. In fact, a good teacher is who lead students to his/her knowledge level. Teacher is a high scientific level in society who shares his/her knowledge with students such that the best solution works in the same iteration of the TLBO algorithm, as a teacher. But, it is important to note that students obtain the knowledge based on the teaching quality by teacher and the students' situation in the classroom. As well, students are taught by their interactions which helps their knowledge enhancement. In TLBO algorithm, the classmates interactions are defined as quasi-codes as:

If $f(X_i) < f(X_j)$ $X_{new,i} = X_{old,i} + r_i(X_i - X_j)$

Else $X_{new,i} = X_{old,i} + r_i(X_j - X_i)$

End If

Where, X_i is the present population (student) and X_j is random student and $r_i(X_i - X_j)$ generates a random value and leads the previous population ($X_{old,i}$) to the more optimal solution.

Panigraph et al. [30] used the modified TLBO algorithm for K-means clustering based on Euclidean distance criterion. To this end, for training in learning phase each student is led to the best difference between 2 better students and to the student's difference with teacher. The researches indicate that modified TLBO algorithm outperforms the main algorithm and PSO and K-means in finding the optimal solution.

Ant colony optimization algorithm (ACO)

ACO is inspired by the studies and observations on the ants' colonies. This technique is inspired by the ants' behavior in finding the path between food and nest. ACO was first proposed by Dorigo et al., as a solution for optimization problems [31]. This research indicated that ants are social insects living in colonies and their behavior is mostly in line with entire colony survival. One of the most important and interesting ants' behaviors is their food finding. This ants' behavior is of a swarm intelligence which is recently considered by many researchers. Based on this algorithm, each ant leaves a footprint or pheromone and each ant is placed in a section of the problem area and detects the next path for its motion based on the probability function and where the pheromone is left in the path. Then, the best path is obtained for each ant's passing. In addition, in [32] the ACO was compared with 3 algorithms by Shelokar. This study indicated that the ACO escapes the local optima better than other algorithms converges better to the clustering problem solution, but the hybrid and genetic simulation algorithms are dropped in the local optima in global optima problems and couldn't converge to the main solution. In [33] Amiri and Niknam used the PSO and ACO in order to improve the clustering. This research could obtain better solutions for the problem by fitting the fuzzy technique for setting the input PSO algorithm parameters. Also, this method was compared to the GA, Taboo search, Honey bees mating, PSO and ACO and simulated annealing and it was indicated that proposed algorithm,

outperform other algorithms in finding the problem's optimal solution and converges to the solution.

Ant Bees colony algorithm (ABC)

ABC [34] is an optimization algorithm based on intelligent behavior of the Honey bees' population. In this algorithm, the bees are first initialized and then based on probability function, they move toward a different direction for honey search. Then, the best position for finding the honey is found and bees would randomly search for the honey in the garden area. Therefore, more bees would be assigned to the nest areas of the garden so that they can search in problem area using multifactorial technique. In next phase, the bees which found the best position would be selected to generate the new population and other bees are remained in problem area and would be randomly dispersed in the search space so that finding bets solution. These phases are iterated until the end condition is met. Zhang [35] implemented the ABC algorithm in order for K- means clustering and proposed K-NM-PSO and indicated that ABC algorithm has lower time complexity and outperforms other algorithms.

Invasive Weed optimization algorithm (IWO)

IWO algorithm is one of the relevant ones in solving optimization problems. This algorithm is inspired by the invasive weeds in nature in optimization technique. This was proposed by Lucas and Mehrabian [36] in 2006. IWO is fast and efficient in finding local regular points and acts based on primary features and natural features of the weeds such as seed generation, raising and breeding for more survival in a colony. The phases of this algorithm are as follow:

There is a primary population dispersed in the problem area and then, by reproduction, each member of population (weeds) can produce the minimum and maximum seeds between 2 predetermined values. Therefore, in spatial dispersion phase in farm or problem area, the generated seeds are randomly dispersed in the problem area. After the spatial dispersion, one has to select the best population and based on crossover, each seed can generate new seeds considering the reproduction and there is a score assigned to the seeds. Finally, seeds with lowest scores are removed. This is iterated until in next iterations the seeds will be converged to the optimal seeds. This algorithm was implemented in 2014 by Gou Pan et al. [37] for clustering based on Euclidean evaluation function. As well, this algorithm was compared to hybrid techniques such as fuzzy C-means, GA and possibilistic C-means and K-means. The results of implementation indicated that K-means invasive weeds technique converges to the solution more slowly and are more robust. In addition, Boobord [38] integrated the IWO and K-means technique and compared to the colonial competition, PSO and Taboo search, ACO and K-means. The results of implementation indicated that this algorithm, outperforms other algorithms in finding the best solution, but there are large standard deviations in some datasets relative to other methods which indicates that the IWO algorithm, is dropped in some cases in the local optima and cannot obtain the best global solution.

Clustering problem solving using meta-heuristic algorithms

In this section, the algorithms used for obtaining the initial clusters are explained and then, the main problem statement is provided. In order for clustering using meta-heuristic algorithm, it is began with a cluster containing one element and in each phase the distance of other elements with the clusters' centroids are calculated. This distance is a criterion for determining the next elements. The algorithm process is as follow:

Meta-heuristic algorithms such as IWO, TLBO, PSO and GA need 3 main phased for solving their problems:

1. Elements' or population's initialization
2. New population selection
3. Results evaluation
4. End of algorithm

The first phased of the algorithm is to determine the primary elements. In meta-heuristic techniques, the first element of population of dataset is considered as the beginning element. The initial element is determined to be the only element of cluster and it is removed from dataset so that is would not be considered in later

Distance function	Formula
Euclidean distance [39]	$D_{xy} = \sqrt{\sum_{i=1}^n x_i - y_j ^2}$
Minkowski distance [40]	$D_{xy} = \sqrt[p]{(x_i - y_j)^p}$
Jaccard distance [41]	$D_{xy} = \frac{\#[(x_{xj} \neq y_{yj}) \cap ((x_{xj} \neq 0) \cup x_{yj})]}{\#[(x_{xj} \neq 0) \cup x_{yj}]}$

calculations. The new population in next stage is calculated based on the distance between all elements of dataset and the center of gravity. As a result, the distance criterion is calculated based on their similarity criterion to the centroids of 3 clusters. Next, the algorithm iterates these processes; that is, addition process of 3 populations to the clusters and eliminating them from clusters have to have at least one element to cluster the data population.

III. THE PROPOSED ALGORITHM

In IWO algorithm, the seeds attempt to compete by producing new seed in order to achieve the best solution in population so that they evolve and they would be dispersed randomly in the problem area to achieve this evolution. This would lead to achieve the best population of seeds but in proposed algorithm as a hybrid form of reproduction and classmates' interaction (learning phase) in TLBO algorithm. Therefore, entire population attempt to generate new population by difference of the reproduction coefficient and learning coefficient. This causes that algorithm relocated toward a more random particle in order to generate its seeds and in this relocation; the entire population exchange its information with other population so that it is drawn to the best random member of the population. But, in this relocation it may be dropped in local optima of the problem which is improved with determining the inertia coefficient of the motion in students' interaction and the seeds would be evolved by better global searching. Therefore, the proposed algorithm (IWO-TLBO) would be as figure 1.

Phase1. The weed seeds are dispersed randomly in problem area.

Phase2. Each member of population produces the new population based minimum and maximum determined seeds.

Phase3. The population seeds are dispersed in problem area based on a random value.

The SD (σ) value for the normal distribution in each phase reduces from the primary defined value ($\sigma_{initial}$) to the final value (σ_{final}). The relationship between above parameters SD can be expressed in relation 1:

$$\sigma_{iter} = (\text{iter}_{max} - \text{iter}) / \text{iter}_{max}^n (\sigma_{initial} - \sigma_{final}) + \sigma_{final}$$

Where, iter_{max} is the highest number of iterations and σ_{iter} is the SD.

The new population would be produced based on previous population. In proposed algorithm, the value of present population is subtracted from each classmate's evolution value and there would be anew random value generated. In new function defined in this paper, the new population is obtained based on a random motion difference value in TLBO and IWO algorithm (relation 2):

$$If (X_i) < (X_j) X_{new,i} = X_i - r\sigma_{iter} - r(X_i - X_j) (2)$$

Else

$$X_{new,i} = X_i - r\sigma_{iter} + r(X_j - X_i) (2) \text{ End If}$$

Where, X_i is the seed value in present population and $X_{new,i}$ is the new produced population

Table1. Distance measurement criteria for clustering

Also in modified version, in order to increase the local optimization coefficients in weed mutation the value r_i is put in interval -10 to +10. This new position of the weeds seeds in problem area helps the local search so that there would be suitable balance in both local and global search; because by excessive local search it is possible that the algorithm doesn't converge to the solution of problem and by weak local search, it is possible that weed seeds are dropped in the local optima.

Phase4. After some iteration, the weak seeds are eliminated from the colony.

Phase5. The best population in entire population is selected as the best evolved seed. Also, the best population is calculated based on the distance function from distance measurement criterion in table 1:

1. Set the generation iteration, $t=0$;
2. Initialize the control parameters, max and min seed;
3. Create and initialize the population, $X(0)$ of n_s individuals;
4. While (stopping condition is not true) do
5. Evaluate the fitness, $f(x_i(t))$ clustering based on distance function
6. Production of the new population or random positions in solution space
7. Reproduction
8. Spatial dispersal and
- 8.1. If $(X_i) < (X_j)$ $X_{new,i} = X_{old,i} - r\sigma_{iter} -$

```

 $r_i(X_i - X_j)$ 
8.2. Else  $X_{new,i} = X_{old,i} - r\sigma_{iter} - r_i(X_j - X_i)$ 
8.3. End If
9. Competitive exclusion
10. if  $f(x'_i(t))$  is better than  $f(x_i(t))$  then
11. Best solution =  $x'_i(t)$ 
12. end
13. end

```

Fig1. The proposed algorithm (IWO-TLBP) quasi-codes

The algorithm evaluation similarity criterion. One of the similarity criteria studied in this research is the Euclidean similarity criterion as given in relation 3:

$$D_{xy} = \sum_{i=1}^m (\sum_{j=1}^k \|x_i - y_j\|)^2 \quad (3)$$

Where, y_j is the centroid of cluster j and all data are divided to k clusters. In example of figure 2, 3 points are considered as centroids and x_i is the clustering problem data.

Fig2. An example of clustering data with centrality data

In table 1, 2 criterions for algorithms' fitness functions are presented which are calculated from 3 inter- clusters distance measurement techniques. These similarity criteria perform as the clustering algorithm's evaluation function [41-46].

Mikowsky: distance between 2 criteria of the typical vector which is normalized and is for value $p=1$, there is a Manhattan distance criterion and for $p=2$, there is a Euclidean distance criterion.

Jaccard: similarity level of 2 data sets from the 2 sets' number of intersection divided by union of 2 sets .

DATA USED IN BENCHMARK TESTS

For accuracy and efficiency assessment of proposed algorithm, experiments are done on 5 standard dataset in order to determine the accuracy of the clustering algorithm. This dataset include Wine, Glass, WDBS, Sonar and Iris selected from UCI standard datasets [42] whose features are explained as follow:

- Iris: a dataset relevant to Iris flower recognition in which 3 different classes are contained each with 50 samples and each sample with 4 features.
- Wine: this is related to recognition of drinking with 178 samples classified in 3 different classes of 59, 59 and 48 samples and each sample contains 13 features.
- Glass identification database: this dataset includes information of different kinds of glasses with 214 samples and 6 classifications. Number of classes 1 to 6 contains 9, 13, 76, 17, 70 and 29 samples each with 9 features and continuous values.
- Diagnostic Breast Cancer (WDBC): this dataset include cancer tumor information for the breast cancer patients in Wisconsin University. There are 569 samples classified in 2 classes: benign (357 samples) and malignant (212 samples). Each sample includes 31 features such as patients' ID and other 30

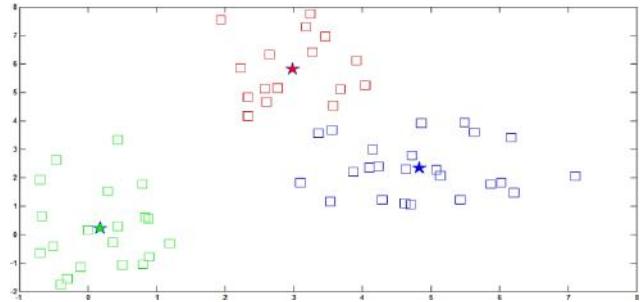
features are related to features of cancer tumor such as tissue, compactness, extent and so forth. These features are real numbers and classification is based on 30 features.

Table2. Clustering with Euclidean distance criterion

IV. RESULTS EVALUATION

We implemented IWO-TLBO algorithm in matlab 2016b in OS

windows 10 and hardware as Ram 6, Cori7. In this implementation with 3 clustering test functions and 4 test data, ABC, PSO, TLBOIWO and IWO- TLBO algorithms are attempting to find the best minimum solution from 3 test functions. Also, the implemented algorithms are run in 200 iterations and each algorithm is separately run 1000 iterations and the best and worst solutions and mean solution and solutions differences were obtained. Based on the clustering Euclidean test function (table2) it can be said that proposed algorithm in Iris data model obtained the optimal solution 60.8817 which is a more minimum value in this dataset in comparison with other meta- heuristic algorithms of the clustering with Euclidean criterion. Also, the minimum difference value of solutions in different iterations indicate that the proposed algorithm



has more robustness in finding best solutions of problem the attempts more than other algorithms to converge to better problem solution. Among other algorithms, it can be understood that the IWO and TLBO and ABC and PSO and GA had necessary convergence to the problem solution finding, respectively. In Wide test data, proposed algorithm could better converge to the solution than others but in finding the Worst solution, IWO and GA outperformed the proposed algorithm and following that, the most minimum value of solutions difference also as obtained by IWO and GA which is due to the fact that proposed algorithm is dropped to the local optima with excessive seeds random diversity in problem area and cannot always escape this trap. In this test data the IWO and LBO and ABC, PSO and GA obtained the best clustering problems' solutions, respectively. In test data Glass, the proposed algorithm minimized STD obtained from the problem solution in BEST, WORST and AVERGAE values and better converged to the solution. The proposed algorithm didn't drop in the local optima due to more balanced local and global search and outperformed other 5 algorithms in approaching the more optimal solution. Also, in this test function, IWO and GA and TLBO and ABC and PSO are the best algorithms implemented in this test data, respectively. In data set cancer, the proposed algorithm as well as the IWO obtained the optimal solution for problem which is due to the fact that both algorithms used the global search in order to find the solution, but in difference of solutions, the IWO obtained better value than proposed algorithm. This indicates that IWO mostly move in line with local solutions, but the proposed algorithm maximized the Worst value obtained from the algorithm by excessive search in problem area and moved away from the optimal

solution by excessive dispersion of weed seeds. But in other algorithms, it can be stated that TLBO, GA, ABC and PSO obtained the best solutions respectively, but among these algorithms GA could obtain the best value of solutions difference which is due to effective local search with more effective global search than ABC and PSO.

Comparing between Minkowski test function for 4 test data, it can be stated that the proposed algorithm obtained the best fitness

Euclidean distance					
Problen	Algorithm	Best	Average	Worst	STD
Iris	ABC	88.6075	89.3662	90.7948	1.292
	PSO	89.3208	90.4217	91.272	.9701
	TLBO	69.1029	69.0857	73.7255	2.7683
	IWO-TLBO	60.8817	60.9661	61.3437	.2432
	IWO	61.9725	62.1844	62.74	.9621
	GA	96.6582	97.9941	98.6263	.6067
Wine	ABC	16306.4918	16413.7587	16438.0475	70.265
	PSO	18083.7934	18087.9899	18313.5993	131.11
	TLBO	16296.7422	16297.0968	16301.9526	2.592
	IWO-TLBO	16292.9924	16293.2455	16298.0886	2.631
	IWO	16295.5971	16296.9639	16297.0795	.8921
	GA	16295.0285	16296.7014	16297.3129	1.8654
Glass identification	ABC	3916.793	3913.5852	3922.1631	4.331
	PSO	4114.0824	4272.4336	4294.0208	98.323
	TLBO	3895.4905	3897.7546	3906.6459	4.341
	IWO-TLBO	3890.9788	3892.71	3894.2981	1.6602
	IWO	3891.3132	3893.7158	3894.9571	1.821
	GA	3892.1104	3893.0596	3895.9977	2.022
WDBC	ABC	7.3489e+07	7.3777e+07	7.3796e+07	1.7202e+05
	PSO	7.3835e+07	7.3904e+07	7.4024e+07	9.5640e+04
	TLBO	7.3280e+07	7.3271e+07	7.3830e+07	3.2017e+05
function's	IWO-TLBO	7.2716e+07	7.3395e+07	7.3816e+07	5.5502e+05
	IWO	7.2716e+07	7.3759e+07	7.3178e+07	5.2263e+05
	GA	7.3830e+07	7.3830e+07	7.3838e+07	4.6188e+03

function's

value which is due to modified IWO global search. Also in test data Iris it can be stated that the IWO, TLBO, ABC, PSO and GA are the best algorithms for problem area search, respectively. But in this test data, GA with the worst solution couldn't obtain the most minimum solution difference after the IWO which is due to irrelevant performance of GA and its being dropped in the local

optima. But in test data Wine, GA outperformed IWO, PSO, ABC and TLBO in obtaining optimal solution which is due to the suitable performance of this algorithm[47-49]. Due to being dropped in local optima and lack of ability to escape it, the IWO couldn't obtain more optimal solution and in this dataset, GA and proposed algorithm maximized the STD value in different runs which is due to global search of both algorithms in searching for the most optimal centrality of the Minkowski test function. In test data Glass, proposed algorithm would converge to solutions better than other algorithms and has more robustness in avoidance of local optima; because this algorithm obtained minimum STD value of the solution. In addition, the GA and TLBO outperformed IWO, ABC and PSO algorithms in escaping the local optima and this is due to global search in mutation and learning phases. In dataset WDBC, the IWO, TLBO and ABC obtained better solutions than GA and PSO and this indicates that PSO is dropped in local optimal due to moving toward local optima more than other algorithms and due to lack of suitable performance in global search, it cannot obtain more optimal solution. Moreover, depending on excessive local search and lack of performance of this search in this dataset, GA didn't obtain more optimal solution than other algorithms.

Table3. Clustering with Minkowski distance criterion

Minkowski distance					
Proble m	Algorithm	Best	Average	Worst	STD
Iris	ABC	87.6804	87.9904	88.736	.765
	PSO	88.7521	89.1125	90.2155	2.611
	TLBO	67.497	69.6754	71.2539	2.654
	IWO-TLBO	61.4384	61.801	62.2153	1.395
	IWO	62.5929	63.2088	63.612	.2997
	GA	95.6797	96.860	97.2371	.884
Wine	ABC	16309.0352	16335.0125	16358.524	21.2552
	PSO	17215.5945	18112.5198	18326.7035	589.4790
	TLBO	16295.7889	16299.7718	16301.8273	2.9700
	IWO-TLBO	16293.2437	16294.8205	16295.7763	1.2489
	IWO	16296.6402	16296.7629	16296.7942	0.0794
	GA	16293.3354	16295.6594	16301.0413	3.548
Glass identification	ABC	3907.735	3910.0739	3912.5899	2.4280
	PSO	4228.2684	4233.8504	4279.4968	27.32
	TLBO	3894.2994	3896.0904	3913.9443	10.8619
	IWO-TLBO	3890.945	3891.1991	3892.2954	.748
	IWO	3908.7389	3910.5701	3911.7929	1.62
	GA	3892.3394	3892.6342	3894.6586	1.245
WDBC	ABC	7.2908e+07	7.3456e+07	7.3470e+07	3.2051e+05
	PSO	7.3604e+07	7.3872e+07	7.4403e+07	4.0583e+07
	TLBO	7.2794e+07	7.3830e+07	7.4110e+07	6.9621e+05
	IWO-TLBO	7.2716e+07	7.3258e+07	7.3830e+07	5.5969e+05

Jaccard distance					
Problem	Algorithm	Best	Average	Worst	STD
Iris	ABC	3.4833	3.5092	3.923	.223
	PSO	2.9667	3.012	3.23	.145
	TLBO	2.7833	2.7867	2.9202	.087
	IWO-TLBO	2.773	2.7799	2.8533	.053
	IWO	2.9287	2.955	3.0601	.0656
	GA	3.48	3.4621	3.694	.124
Wine	ABC	169	169.036	169.0714	.0379
	PSO	173.1286	173.5628	173.6429	.256
	TLBO	168.5	168.5801	168.8571	.185
	IWO-TLBO	168.5	168.545	168.5714	.034
	IWO	168.6294	168.6712	168.7143	.045
	GA	173.0714	173.490	174	.502
Glass identification	ABC	201.9535	201.9905	202.0467	.0489
	PSO	210.3636	210.6103	211.0109	.322
	TLBO	201.6689	201.8297	201.9015	.1191
	IWO-TLBO	201.32	201.4013	201.7868	.25
	IWO	201.5848	201.6772	201.8915	.122
	GA	204.5455	204.6389	207.1818	1.445
WDBC	ABC	345.2727	356.4887	378.43	16.434
	PSO	630.9091	630.9212	632.5291	.9339
	TLBO	339.2364	340.1216	350.3636	7.34
	IWO-TLBO	330	344.2535	350.1818	10.242
	IWO	332.9091	333.8058	333.8182	.603
	GA	500.9091	506.2479	558.9091	30.43
	IWO	7.2718e+07	7.3877e+07	7.4148e+07	7.5956e+05
	GA	7.3130e+07	7.3892e+07	7.3892e+07	4.3994e+05

In table 4, the proposed algorithm implementations for 4 datasets are presented which in each 4 datasets; the proposed algorithm outperformed other algorithms in searching. In dataset Iris and Wine, TLBO, IWO, IWO-TLBO, PSO, GA and ABC are the best algorithms for Jaccard clustering respectively. In addition, in dataset Iris, the TLBO, proposed algorithm and in dataset Wine, the proposed algorithm and IWO in different runs respectively outperformed other algorithms in finding similar difference values of solutions. In dataset Glass, the IWO, TLBO, ABC, GA and PSO respectively were the best algorithms successful in finding the global solutions and provided the best results in Jaccard clustering. On the one hand, ABC algorithm and PSO algorithm obtained the lowest and highest values of STD, respectively which indicate that ABC is dropped in local optimal and PSO couldn't obtain more optimal solution with maximum searching dispersion. Generally, it is concluded that the search dispersion in problem area generated worse solutions than local search, but local search may never obtain

better solution than this because it is dropped in the local optima. IWO, TLBO, ABC, GA and PSO respectively obtained the best results in cancer dataset clustering, proposed algorithm obtained worse solution difference value than other algorithms and this indicated that proposed algorithm attempted to find the best solution with search dispersion and in each iteration obtained different solutions. But IWO obtains the best STD value which is due to similar convergences obtained in problem area.

Table4. Clustering with Jaccard distance criterion

V. CONCLUSION

The main goal of this clustering problem is to find similar data and discover different data. In this research 6 meta-heuristic algorithms were implemented for this reason. In order to compare the results, 3 similarity or distance functions (Jaccard, Minkowski, Euclidean) were used. Also in this study, the modified IWO algorithm were implemented with different criteria and we indicated that these algorithms can find similar clusters and based on the global search of this algorithm, it can be stated that the proposed algorithm can avoid the local optima and using better global search, avoided being dropped in this trap. In addition, the IWO, TLBO, PSO and GA and ABC algorithms were compared with proposed algorithm each with best fitness in a different problem, but generally it can be stated that the PSO algorithm cannot obtain more optimal solution in majority of test functions and datasets. This is due to the fact that PSO algorithm is mostly dependent on local search. But, following the proposed algorithm, the performance of TLBO and IWO can be the most suitable clustering algorithms because both are better convergence in finding bets solutions.

REFERENCES

- [1] Hartigan, J. A. (1975). Clustering Algorithms Wiley New York Google Scholar.
- [2] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108.
- [3] Ensafi, R., Dehghanzadeh, S., & Akbarzadeh-T, M. R. (2008, March). Optimizing fuzzy k-means for network anomaly detection using pso. In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on (pp. 686-693). IEEE.
- [4] Sha, M., & Yang, H. (2009, November). Speaker recognition based on APSO-K-means clustering algorithm. In Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on (Vol. 2, pp. 440- 444). IEEE.
- [5] Abdel-Kader, R. F. (2010, February). Genetically improved PSO algorithm for efficient data clustering. In Machine Learning and Computing (ICMLC), 2010 Second International Conference on (pp. 71-75). IEEE.
- [6] Zheng, X., & Jia, Y. (2011, December). A study on educational data clustering approach based on improved particle swarm optimizer. In IT in Medicine and Education (ITME), 2011 International Symposium on (Vol. 2, pp. 442-445). IEEE.
- [7] Li, Y., & Wu, H. (2012). A clustering method based on K-means algorithm. Physics Procedia, 25, 1104-1109.
- [8] Naik, B., Swetanisha, S., Behera, D. K., Mahapatra, S., & Padhi, B. K. (2012, November). Cooperative swarm based clustering algorithm based on PSO and k-means to find optimal cluster centroids. In Computing and Communication Systems (NCCCS), 2012 National Conference on (pp. 1-5). IEEE.
- [9] Govindarajan, K., Somasundaram, T. S., & Kumar, V. S. (2013, July).

- Particle swarm optimization (PSO)-based clustering for improving the quality of learning using cloud computing. In Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on (pp. 495-497). IEEE.
- [10] Liao, Q., Yang, F., & Zhao, J. (2013, November). An improved parallel K-means clustering algorithm with MapReduce. In Communication Technology (ICCT), 2013 15th IEEE International Conference on (pp. 764-768). IEEE.
 - [11] Bai, L., Liang, J., Sui, C., & Dang, C. (2013). Fast global k-means clustering based on local geometrical information. *Information Sciences*, 245, 168-180.
 - [12] Liao, K., Liu, G., Xiao, L., & Liu, C. (2013). A sample-based hierarchical adaptive K-means clustering method for large-scale video retrieval. *Knowledge-Based Systems*, 49, 123-133.
 - [13] Jagannathan, P., & Jaiganesh, S. (2013, December). An improved K-means algorithm combined with Particle Swarm Optimization approach for efficient web document clustering. In Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on (pp. 772-776). IEEE.
 - [14] Yao, H., Duan, Q., Li, D., & Wang, J. (2013). An improved K-means clustering algorithm for fish image segmentation. *Mathematical and Computer Modelling*, 58(3), 790-798..
 - [15] Rebollo-Monedero, D., Solé, M., Nin, J., & Forné, J. (2013). A modification of the k-means method for quasi-unsupervised learning. *Knowledge-Based Systems*, 37, 176-185.
 - [16] Shahbaba, M., & Beheshti, S. (2014). MACE-means clustering. *Signal Processing*, 105, 216-225.
 - [17] Naldi, M. C., & Campello, R. J. G. B. (2014). Evolutionary k-means for distributed data sets. *Neurocomputing*, 127, 30-42
 - [18] Scitovski, R., & Sabo, K. (2014). Analysis of the k-means algorithm in the case of data points occurring on the border of two or more clusters. *Knowledge-based systems*, 57, 1-7
 - [19] Aradhya, V. M., & Pavithra, M. S. (2016). A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video. *Applied Computing and Informatics*, 12(2), 109-116
 - [20] Sheikholeslami, C., Chatterjee, S., Zhang, A. (1338) WaveCluster: A multi resolution clustering approach for very large spatial database. In: Proceedings of 24 th VLDB Conference, New York, USA.
 - [21] Ng, R., Han, J. (1334) Efficient and effective clustering methods for spatial data mining. In: Proceeding's of the 20 th VLDB Conference, Santiago, Chile.
 - [22] Rashidi, P., & Cook, D. J. (2003). Keeping the resident in the loop: Adapting the smart home to the user. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 33 (5), 343-353
 - [23] Jensi, R., & Jiji, G. W. (2015). Hybrid data clustering approach using K-Means and Flower Pollination Algorithm. arXiv preprint arXiv:1505.03236.
 - [24] Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press.
 - [25] Aibinu, A. M., Salau, H. B., Rahman, N. A., Nwohu, M. N., & Akachukwu, C. M. (2016). A novel clustering based genetic algorithm for route optimization. *Engineering Science and Technology, an International Journal*, 19(4), 2022-2034.
 - [26] Kennedy, J. (2011). Particle swarm optimization. In Encyclopedia of machine learning (pp. 760-766). Springer US.
 - [27] Alam, S., Dobbie, G., & Riddle, P. (2008, September). An evolutionary particle swarm optimization algorithm for data clustering. In Swarm Intelligence Symposium, 2008. SIS 2008. IEEE (pp. 1-6). IEEE.
 - [28] Izakian, H., & Abraham, A. (2011). Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. *Expert Systems with Applications*, 38(3), 1835-1838.
 - [29] Rao, R. V., Savsani, V. J., & Vakharia, D. P. (2011). Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems. *Computer-Aided Design*, 43(3), 303-315.
 - [30] Panigrahi, S. K., & Pattnaik, S. (2016). Empirical Study on Clustering Based on Modified Teaching Learning Based Optimization. *Procedia Computer Science*, 92, 442-449.
 - [31] Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE computational intelligence magazine*, 1(4), 28-39.
 - [32] Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. *Analytica Chimica Acta*, 509(2), 187-195.
 - [33] Niknam, T., & Amiri, B. (2010). An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Applied Soft Computing*, 10(1), 183-197.
 - [34] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, 39(3), 459-471.
 - [35] Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. *Expert Systems with Applications*, 37(7), 4761-4767.
 - [36] Mehrabian, A. R., & Lucas, C. (2006). A novel numerical optimization algorithm inspired from weed colonization. *Ecological informatics*, 1(4), 355-366.
 - [37] Pan, G., Li, K., Ouyang, A., Zhou, X., & Xu, Y. (2014). A hybrid clustering algorithm combining cloud model IWO and k-means. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(06), 1450015.
 - [38] Boobord, F., Othman, Z., & Bakar, A. A. (2015). A WK-means Approach for Clustering. *Int. Arab J. Inf. Technol.*, 12(5), 489-493.
 - [39] Elmore, K. L., & Richman, M. B. (2001). Euclidean distance as a similarity metric for principal component analysis. *Monthly Weather Review*, 129(3), 540-549.
 - [40] Ichino, M., & Yaguchi, H. (1994). Generalized Minkowski metrics for mixed feature-type data analysis. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4), 698-708.
 - [41] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In Proceedings of the International MultiConference of Engineers and
 - [42] <http://archive.ics.uci.edu/ml/>
 - [43] Mehdi Gheisari. Design, Implementation, and Evaluation of SemHD: A New Semantic Hierarchical Sensor Data Storage. Indian J. Innovations Dev., Vol. 1, No. 3 (Mar 2012) ISSN 2277 – 5390
 - [44] Mehdi Gheisari, Mehdi Esnaashari. A survey to face recognition algorithms: advantageous and disadvantageous, J. Modern Technology & Engineering Vol.2, No.1, 2017, pp.57-65
 - [45] Mehdi Gheisari, Guojun Wang, "A Survey on Deep Learning in Big Data", 15th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC 2017), July 2017, Guangzhou, China
 - [46] Mehdi Gheisari, Ali Akbar Movassagh, Yongrui Qin, Jianming Yong, Xiaohui Tao, Ji Zhang, Haifeng Shen, "NSSSD: A New Semantic Hierarchical Storage for Sensor Data", IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2016), May 4-6, 2016, Nanchang, China.
 - [47] Fazli, Mojtaba Sedigh, and Jean-Fabrice Lebrat. "A solution for forecasting pet chips prices for both short-term and long-term price forecasting, using genetic programming." The 2013 International Conference on Artificial Intelligence. Vol. 2. CSREA Press, 2013.APA.
 - [48] Fazli, Mojtaba Sedigh, Keykhosrow Keshavarzi, and Saeed Setayeshi. "Designing a Hybrid Neuro-Fuzzy System for Classifying the Complex Data, Application on Cornea Transplant." Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013.

Diagnosis of Heart Disease Using Neural Networks

Chaitanya Roygaga

*Computer Engineering Department,
Thadomal Shahani Engineering College,
Mumbai, India*

Swapnil Sampat

*Computer Engineering Department,
Thadomal Shahani Engineering College,
Mumbai, India*

Suraj Punjabi

*Computer Engineering Department,
Thadomal Shahani Engineering College,
Mumbai, India*

Tanuja Sarode

*Computer Engineering Department,
Thadomal Shahani Engineering College,
Mumbai, India*

Abstract- Heart diseases have been the cause of frequent deaths. It is difficult to diagnose heart problems at every medical centre, because of lack of technology and the cost to afford it. This problem has been increasing majorly in rural areas. That is why it is very important to develop an affordable and reliable technology. Artificial Neural Networks (ANNs) is intended towards developing such an intelligent system, which can diagnose whether a patient is suffering from a heart disease or not. The dataset is acquired from the UCI Machine Learning Repository. The training dataset was fed into the network. The learning network used in our analysis is Error Back Propagation algorithm. Artificial Neural network is used to classify and distinguish between absence and presence of disease. The performance measure taken into consideration is accuracy. The targets for the neural network have been classified as 0's (Disease is absent) and 1's (Disease is present) [4]. The results obtained from back propagation algorithm using varying number of neurons in hidden layer been compared in this research work. Our system has given the best accuracy (at 80.27%) of diagnosing heart disease when the neurons in the hidden layer are kept at four, with high sensitivity and specificity value. This system acts as a promising tool for heart disease diagnosis.

Keywords— ANNs; Heart Disease Diagnosis; Feed Forward; Error Back Propagation algorithm; Classification Accuracy; Database.

I. INTRODUCTION

Diagnosis of any disease is a significant challenge for all healthcare organizations. They aim to provide effective treatments to patients after diagnosing them correctly [3]. Decisions taken by medical practitioners must be accurate, else may lead to tragic consequences for the patient. Clinical tests must be provided at a lower and more affordable rate by the healthcare centres. In such cases, computer-based information and decision support systems can be useful. Diagnosis of Heart Disease is based on analysis of the various tests conducted on a patient. The tests may have results regarding the symptoms and the health examination of the patient.

After studying medicine, and through years of medical practice, doctors are able to predict heart disease. But, this prediction

cannot always be true. Human error is always a factor that needs to be considered in such a situation. Patient data is stored by medical centres, which can later be useful for analysing risk factors of various types of diseases. Artificial Neural Network is a flexible and powerful tool to help doctors for analysing and modelling of the patient data, so that it may be used for prediction purposes, and other applications in the medical field. Neural network can classify and organize the information given to it during training, and then use that data to solve complex problems.

II. EXISTING WORK

Cardiovascular disease dataset was analysed by Kumari Milan [5, 6] using data mining classification techniques. These techniques include Decision Tree, RIPPER classifier, Neural Networks and Support Vector Machine. RIPPER accuracy was 81.08%, Decision Tree accuracy was 79.05%, and ANN accuracy was 80.6%, and SVM accuracy was 84.12%.

Ibrahim Turkoglu and Resul Das [7] used SAS software to diagnose heart disease. Neural network ensemble was used to create new models by combining predicted values from multiple predecessor models. 89.01% precision was obtained during classification. Sensitivity value was 80.95%, and specificity value was 95.91%.

Niti Guru [8] used Artificial Neural Network to predict whether a person is suffering from heart disease. Experiments were performed on a dataset, with 13 inputs. This data was used as training data, and then to teach the network. He used backpropagation algorithm to teach the network. After the network was trained, if any doctor entered data unknown to the system, the system would analyse this data and would give list of diseases from which the patient could be suffering from. Human error could be avoided in such a system.

Hongmei [9, 10] utilized multilayer perceptron (MLP) neural network with three layers for developing a system to diagnose five types of heart diseases. 38 input attributes were used by him. Number of neurons in the hidden layer to be used for best accuracy was determined through the cascade learning process. Momentum rate and adaptive learning rate were used along with backpropagation algorithm by the author to teach the network. A

high level of accuracy (63.2 – 82.9%) was obtained on classification of heart disease.

John Gennari suggested the CLASSIT Conceptual clustering system. The Cleveland database was used by him for his research, and he could achieve a precision of 78.9% [11]. Detrano's [12] method used logistic regression, and he could achieve an accuracy of 77%. Rajkumar and Sophia suggested to diagnose heart disease using data mining algorithm. They achieved a precision of 53.33% [13, 14].

III. IMPLEMENTATION

A) Database

The dataset for the analysis was acquired from the UCI machine learning repository [15, 2, 1]. 13 vital attributes [16] were used for research out of all the available attributes. Those 13 attributes are as shown in Table 1 [16]:

Table I: Data Attributes

Age (Years)	1 (30-70), 2 (>70)
Sex	1 (Male), 0 (Female)
Chest Pain (CP) Type	1 (Typical angina), 2 (Atypical angina), 3 (Non-anginal pain), 4 (Asymptomatic)
Resting blood pressure	mm/Hg
Cholesterol level	vh (Very high), h (High)
Fasting blood sugar (>120 mg/dl)	1 (True), 0 (False)
Resting Echocardiographic result	0 (Normal), 1 (ST-wave abnormality), 2 (Probable or definite left ventricular Hypertrophy)
Maximum heart rate	ab (Abnormal), norm (Normal)
Exercise induced angina	1 (Yes), 0 (No)
Slope of ST segment	1 (Upsloping), 2 (Flat), 3 (Downsloping)
Number of vessels colored by fluoroscopy	0-3
Thal (Obtained defect)	3 (Normal), 6 (Fixed defect), 7 (Reversible defect)
Heredity (History of heart disease in family)	1 (Yes), 0 (No)
Class	0 (Healthy), 1 (Has heart disease)

B) Normalization of Attributes

In order to perform efficient classification, the values of the dataset need to be transformed into homogenous values that yield numerical stability [17, 18]. In this process we determine the highest value of each attribute. This highest value is then used to find the normalized value of that data. Normalization is done in order to get a value between 0 and 1, as shown in Equation (1). This will enhance the training of the network.

$$\text{Normalized value} = \frac{\text{Attribute value}}{\text{Corresponding Max value}} \quad \text{Eq. (1)}$$

C) Artificial Neural Network

It is a computer based model of the human nervous system. It tries to incorporate the working of a human brain and perform similar actions. The elementary computational model is known as a node. A node accepts inputs from other neurons, processes the information and gives the desired output. An ANN architecture consists of multiple nodes that are interconnected. These connection links are associated with weights that carry information [21]. All the nodes work simultaneously in order to solve complex problems. ANN needs to be trained initially and later tested. Therefore a dataset is used as: Training dataset and testing dataset. After training the network, it is tested to check the accuracy. Multiple algorithms are used in Neural Networks; Error back propagation algorithm is one of them.

D) Error Back Propagation

An example of supervised learning is the backpropagation neural network. It mainly consists of three types of layers: The input, middle and the output layers (Fig. 1) [19, 20].

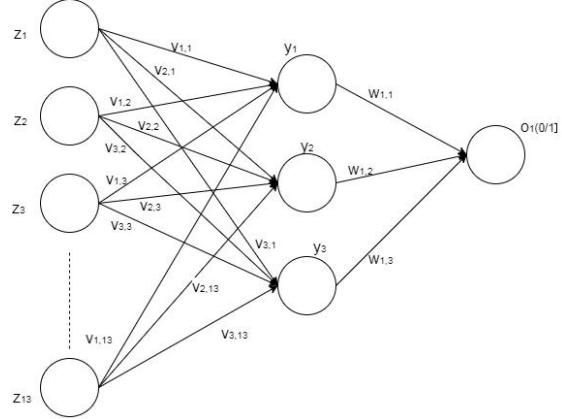


Fig. 1: Proposed Neural Network Topology

EBPT works in two phases: Feedforward phase and Feedback phase. The feedforward phase has following steps. Weights are randomly initialized. These weights are then multiplied with the input pattern and an activation function is applied to the product.

The output of the activation function serves as an input to the hidden layer. Sigmoid activation function is used, and is given by:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The same process is repeated again where the hidden layers acts as an input layer and the output layer as the hidden layer. This result is then compared to the desired result. This helps in determining the error.

$$E_{MSE} = \frac{1}{M} \sum (Desired\ output - Actual\ output)^2 \quad (3)$$

Where, E_{MSE} = Mean square error
 M = Number of training inputs

This ends the feedforward phase. In the feedback phase the error is sent backwards. We use this error to calculate the error signal vectors and the error vectors are used to update the weights [22]. Until error decreases to a minimum value, process is repeated.

E) Design of Network

297 data samples have been utilized to conduct this research. Training and testing samples are taken from the available dataset. The sample data is divided into half: 150 samples for training, and 147 samples for testing for the network input. The output is given as (1) for heart disease presence and (0) for heart disease absence.

The data is converted to a matrix which serves as an input to the network. A 150 x 13 matrix training set was obtained, with its corresponding target matrix of 150 x 1. Similarly, a 147 x 13 matrix testing set was obtained, with its corresponding target matrix of 147 x 1. 13 neurons are used to present the network with the data as input, as this is the number of attributes we have considered to give as input to the network. The learning rate was kept constant at 0.5.

The experimentation was conducted by varying the number of hidden layer neurons. We wanted to know the best number of neurons in the hidden layer to be implemented in the network, so that the network would give the best results in terms of accuracy, to correctly determine whether the patient whose test case is utilized has a heart disease or not. We started with three neurons, until we reached eight neurons. Four hidden neurons gave us the best result. At the output layer, one neuron was applied, which indicated the target of the network.

F) Performance Evaluation

Our research is based on changing the number of hidden layer neurons (3, 4, 5, 6, 7 and 8) in the backpropagation algorithm, to obtain a network which can best predict whether a test case has a heart disease or not.

The rate of learning used in this research work is 0.5, and the activation function is Sigmoid.

The diagnostic analysis has been implemented using the following terms: When a person is examined to check whether he has a heart disease; If heart disease is present and that person is diagnosed having a heart disease, this term is called as true positive (TP). Even though heart disease is absent in a person, and the test is claiming negative, then this is referred to as true negative (TN). But, in case a heart disease is present in a person, but the test claims the opposite, is referred to as false negative

(FN). False positive (FP) is a situation where the test shows the presence of disease even if the person is not suffering from it. To evaluate the efficiency of an intelligent system, a statistical measure must be considered. In this analysis sensitivity, specificity, and accuracy [2] can be considered as statistical measures.

Sensitivity is a measure of percentage of people living with heart disease and correctly diagnosed positive i.e. having a heart disease.

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100 \quad (4)$$

Specificity measures the proportion of people who have been rightly identified as not having heart disease, i.e. they are healthy and rightly identified as not having any heart disease.

$$\text{Specificity} = \frac{TN}{TN + FP} * 100 \quad (5)$$

Specificity, sensitivity and accuracy have been used to determine the performance of the network. Performance has also been measured using positive prediction value (PPV), which tells us how efficiently a positive test result determines presence of heart disease.

$$\text{PPV} = \frac{TP}{TP + FP} * 100 \quad (6)$$

Similarly, negative prediction value (NPV), which tells us how well a negative test result actually determines that the disease is absent.

$$\text{NPV} = \frac{TN}{TN + FN} * 100 \quad \text{Eq. (7)}$$

Table 3 shows the comparison between the different numbers of middle layer neurons used to carry out this research. Four middle layer neurons give the best accuracy.

Figures 2, 3 and 4 show the error graphs of 4, 6 and 8 middle layer neurons respectively. In Figure 2, we can see that when four hidden layer neurons were used in the network, it took 15900 iterations to reach an error of 0.01. Similarly, Figure 3 shows that when six hidden layer neurons were used, it took 13200 iterations to train the network. Also, (In Figure 4) when eight hidden layer neurons were used, it took 61500 iterations to train the network. This error graph comparison also shows us that we get the best accuracy in the least training iterations when we use four hidden layer neurons.

Table II: Comparison between the different numbers of hidden layer neurons

No. of hidden layer neurons	True Positive (TP)	False Negative (FN)	True Negative (TN)	False Positive (FP)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
3	44	26	68	9	76.19	62.86	88.31	83.01	72.34
4	54	16	64	13	80.27	77.14	83.11	80.6	80
5	50	20	67	10	79.59	71.43	87.01	83.33	77.01
6	52	18	63	14	78.23	74.29	81.82	78.79	77.78
7	52	18	61	16	76.87	74.29	79.22	76.47	77.22
8	57	13	59	18	78.91	81.43	76.62	76	81.94

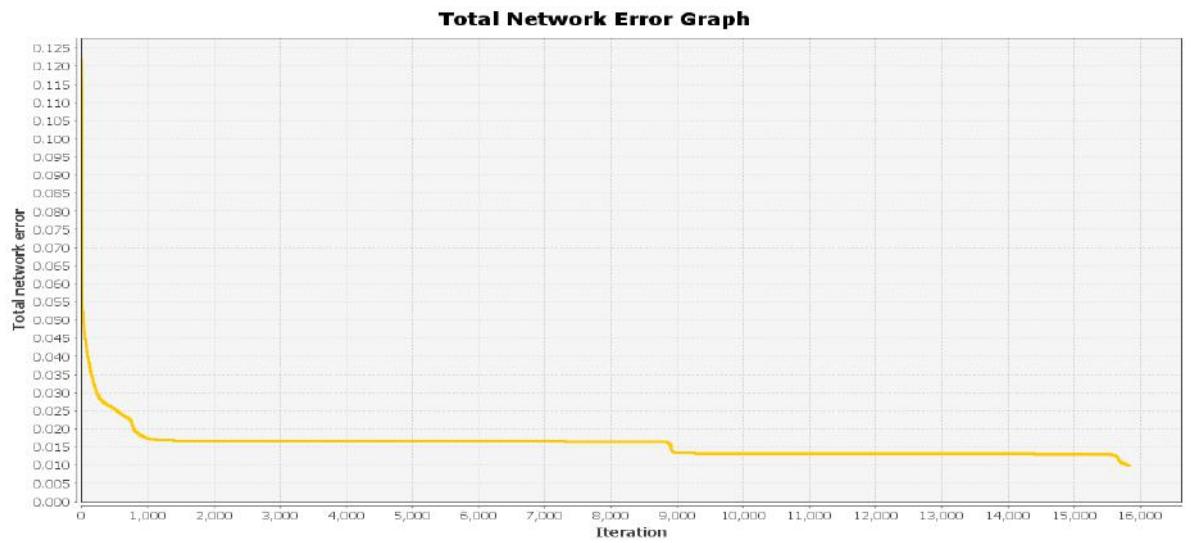


Figure 2 -Error Graph (4 hidden layer neurons)

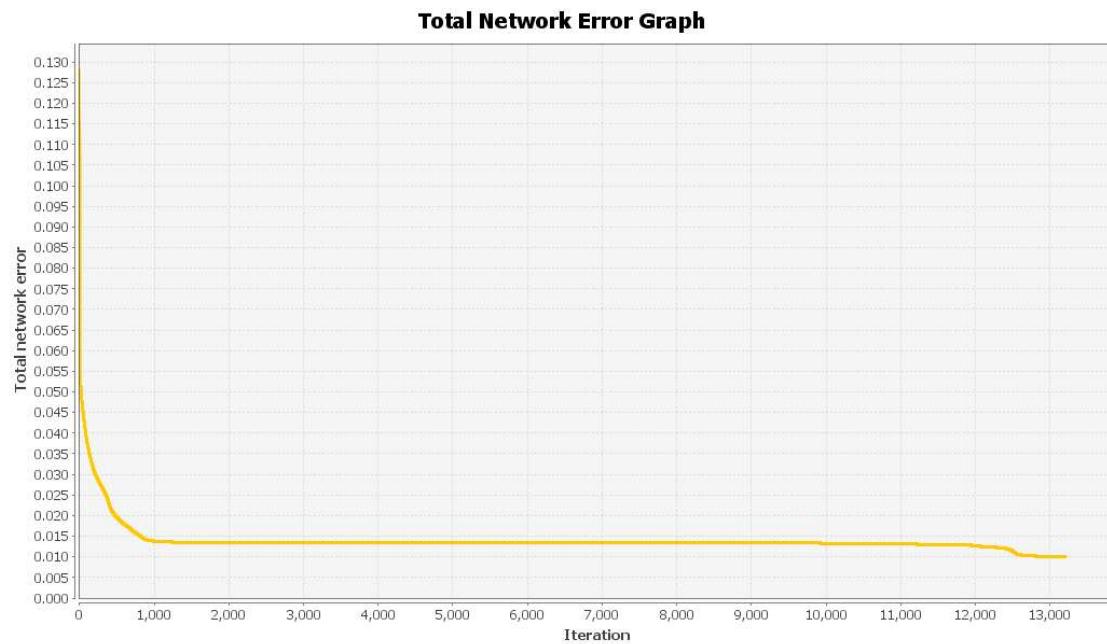


Figure 3 – Error Graph (6 hidden layer neurons)

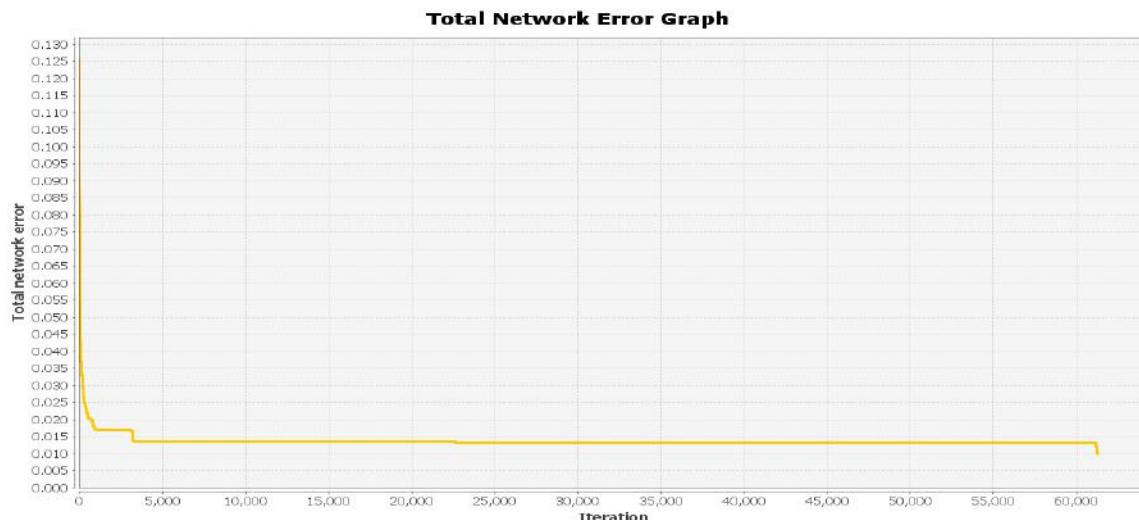


Figure 4 – Error Graph (8 hidden layer neurons)

IV. CONCLUSION

Multiple experiments were conducted using various samples to check how accurate and effective the proposed system is. In our result, we concluded that the system gives the best accuracy (at 80.27%) of diagnosing heart disease when the middle layer neurons are kept at four, with high sensitivity and specificity value.

Overall, the results should successfully and precisely determine the presence or absence of a heart disease. A decision support system, based on ANN, is implemented and tested to detect Heart Disease in a patient.

REFERENCES

- [1] Sameh Ghwanmeh, Adel Mohammad, Ali Al-Ibrahim, “Innovative Artificial Neural Networks-Based Decision Support System for Heart Diseases Diagnosis,” Journal of Intelligent Learning Systems and Applications, 2013, 5, 176-183 <http://dx.doi.org/10.4236/jilsa.2013.53019> Published Online August 2013 (<http://www.scirp.org/journal/jilsa>).
- [2] Ebenezer Obaloluwa Olaniyi and Oyebade Kayode Oyedotun, “Heart Diseases Diagnosis Using Neural Networks Arbitration,” International Jorunal Intelligent Systems and Applications, 2015, 12, 75-82 in MECS (<http://www.mecs-press.org/>) DOI:10.5815/ijisa.2015.12.08.
- [3] A. T. SAYAD, P. P. HALKARNIKAR, “DIAGNOSIS OF HEART DISEASE USING NEURAL NETWORK APPROACH,” International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009 Volume- 2, Issue-3, July-2014.
- [4] Noura Ajam, “Heart Diseases Diagnoses using Artificial Neural Network,” Network and Complex Systems www.iiste.org ISSN 2224-610X (Paper) ISSN 2225-0603 (Online) Vol.5, No.4, 2015.
- [5] M. Kumari, S. Godara, “Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction,” International Journal of Computer Science and Technology, vol. 2, Issue 2, June (2011).
- [6] N. Al-Milli, “Backpropagation neural network for prediction of heart disease,” Journal of theoretical and applied information Technology, vol. 56, pp.131-135, Oct 10, 2013.
- [7] R. Das, I. Turkoglu, A. Sengur, Effective Diagnosis of Heart Disease through Neural Network Ensemble, “Expert Systems with Applications” vol. 36 issue 4, pp. 7675-7680, May (2009). [Available]:10.1016/j.eswa.2008.09.013.
- [8] N. Guru, A. Dahiya and N. Rajpal, “Decision Support System for Heart Disease Using Neural Network,” Delhi Business Review, vol. 8, No 1, pp. 1 – 6, Jan – June (2007).
- [9] S. Prabhat Panday, N. Godara, “Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron,” International Journal of Computer Applications (0975 – 8887) Vol. 45– No.8, May (2012).
- [10] J. S. Sonawane, D. R. Patil and V. S.Thakare, “Survey on Decision Support System for Heart Disease,” International Journal of Advancements in Technology, vol 4, pp. 89-96, 2013.
- [11] Gennari, J. Models of incremental concept formation. Journal of Artificial Intelligence, vol. 1, pp. 11-61., 1989.
- [12] Detrano R., et al, International application of a new probability algorithm for the diagnosis of coronary artery disease, American Journal of Cardiology, 1989;64(5) 304-310.
- [13] A. Rajkumar and G. S. Reena, “Diagnosis of heart disease using data mining algorithm,” Global Journal of Computer Science and Technology, vol. 10, pp. 38-43, December 2010.
- [14] Vanisree K., Jyithi Singaraju, “Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks,” International Journal of Computer Applications (0975 – 8887) vol. 19-No.6, April 2011.

- [15] <http://archive.ics.uci.edu/ml/datasets/heart+Disease>
- [16] fftp.ics.uci.edu/pub/machine-learning-databases (last accessed: February 10, 2015).
- [17] E. O.Olaniyi, K. Adnan., “Onset Diabetes Diagnosis Using Artificial Neural Network”, International Journal of Scientific and Engineering Research, vol. 5, Issue 10, October (2014).
- [18] N. Kilic, B. Ekici, S. Hartomacioglu, “Determination of penetration depth at high velocity impact using finite element method and artificial neural network tools,” Defence Technology xx pp. 1-13, 2015. [online]: <http://dx.doi.org/10.1016/j.dt.2014.12.001>.
- [19] A. K. Nanila, A. P. Singh, “Fault Diagnosis of Mixed-Signal Analog Circuit Using Artificial Neural Network”, International Journal of Intelligent Systems and Applications, 07, pp. 11-17, 2015. DOI: 10.5815/ijisa.2015.07.02.
- [20] S. Bhuvaneswari, J. Sabarathinam, “Defect Analysis Using Artificial Neural Network” I.J. Intelligent Systems and Applications, 05, pp. 33-38, 2013.
- [21] M. Melanie, An Introduction to Genetic Algorithms. MIT press, pp. 20-64, 1999.
- [22] Magret H. Dunham, data mining Introductory and advanced topics, (2003).

Machine learning approach to predict farmer's loan/credit repayability using weather prediction and credit history

Waheeda Dhokley

*Department Of Computer Engineering,
M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Samad Ansari

*Department Of Computer Engineering,
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai, India*

Umair Shaikh

*Department Of Computer Engineering,
M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Nehal Ansari

*Department Of Computer Engineering,
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai, India*

Abstract- Our project is a comprehensive machine learning approach towards prediction of loan/credit repayability using weather prediction and credit history. It aims to develop a rating scale (0-10) based on the weather conditions and farmers credit history using machine learning. Through our portal, an individual could enter all relevant information into the designated fields and our algorithm developed via machine learning would give a rating on a scale (0-10), which would determine the likelihood of the individual being able to fulfill the loan/credit taken. Through our portal, we would try to provide knowledge to farmers, bankers and other microloan lenders to determine the best possible scenario in which maximum capital should be invested. Alternatively other friendly options/recommendations would also be provided in order to maximize the harvest/profits.

Keywords: Machine Learning, Weather Prediction, Credit history, banking sector.

I. INTRODUCTION

"Farming nowadays is an extremely risky occupation, with annual incomes often depend on the weather and seasons, and it's getting heavily devastated by the effects of climate change," said by a renowned Indian psychiatrist and mental health expert residing in Boston [1-5]. According to the latest NCRB reports, "bankruptcy and indebtedness" witnessed the sharpest rise in graph just 3 years ago, registering an almost threefold increase (3,097) as compared to just 2 years prior (1,163).

Approximately 60-70 percent of farmers belonging to the low income category committed suicide in 2015 because of various reasons such as bankruptcy and/or debts after taking loans from banks and registered micro-finance institutions as proven by official government reports and data. It's for the first time that farmers' suicides due to debt or bankruptcy based on the source of loans is been categorised by NCRB[7-9]. The interdependence of economic downfall sharp and rise of suicides

is astounding and, so is effect of climate change on suicide rates. Even Though, the Central government has announced a \$1.3 billion climate-based, crop insurance scheme partially hidden as suicide prevention policy, evidence to support such a scheme is still lacking [10-11].

Indian agriculture is heavily dependent on timely rains and its dependency is continuing to rise, land owned for farming being small and farmers struggling for investments, . NASA scientists predict temperature rise of at least 3 more degrees by 2035. This calls for an urgent and enhanced measures to improve rural farmer's financial situation, crop insurance cover and preventive strategies to avoid disastrous consequences in the long-term.

II. LITERATURE SURVEY

The primary reason in more than half of all suicides by farmers nationwide is that nearly 60-70 percent of India's 90 million agricultural households are in debt due to the business and household expenditure techniques adopted, by which they spend more than they earn on average each month, according to an analysis of various government data sponsored by India Spend .

The failing situation of such agricultural households across the nation, which are mostly indebted are also accelerated by the additional loans that families and households take to meet health and social issues, leaving them with exhausted ability to invest in farming and agriculture. Defaulted loans for health related issues increased threefold over a decade till 2014, and loans for farm business fell by about 50 percent over the same time period. These government sponsored and independent surveys help understand the nature of India's agricultural crisis in the light of the recent spike of farmer protests across various states to demand loan waivers and fair prices for their harvests and equipments. According to National Sample Survey Office, a situation assessment survey held in 2013, nearly 60 million households are spending more than they earn, had land holdings of one hectare or less. In contrast, 0.35 million (0.39 per cent) households owning more than 10 hectares of land had an

average monthly income of approximately Rs 42,000 and consumption expenditure of approximately Rs 15,000, thereby maintaining a monthly surplus.

According to an analysis of NSSO data conducted by the National Bank For Agriculture and Rural Development (NABARD) in 2015, apart from plummeting agricultural income, rising healthcare costs increase farmer's debt. Huge outstanding loans for health conditions have doubled in the last decade. Meanwhile, over the course of the decade loans for farm business fell by half, from 60 to 30 percent by 2012.

Bankruptcy is defined as a legal proceeding involving a person or business that is unable to repay outstanding debts. One of the major causes of suicides is bankruptcy. The process of bankruptcy begins with a petition filed by the debtor (farmer). The debtor's assets are then accessed, appreciated and evaluated, and the assets may be used to repay a portion of remaining debt. Likewise if farmer apply for the loan at the bank or microlender for need of the production ,but if there is not at much of the production as expected in the month or the value of product in the market is less than there is not much profit because of which the farmer is unable to pay the loan back to the bank. So when in order to pay the loan the farmer files a bankruptcy in which the bank abduct assets of the farmers which at the end leads them to suicide [12].

III. SCOPE

The part of project planning that involves determining and documenting a list of specific project goals, deliverables, tasks, costs and deadlines is called Project scope. The Project scope is related to the work necessary to deliver a product within a prespecified deadline. Project scope is defined by requirements and deliverables, and it is extremely important that the project stakeholders are completely in sync and updated regularly with the progress of the Project.

A) Current Scope

The current implementation of the portal will focus on developing an algorithm to take raw weather data and farmer credit history into account and put forward a rating on a scale. Will create 2 different profile login, for farmers and bankers, for them to access and store their information on our database, hence improving the accuracy of the algorithm. The portal will also feature government announcements related to agriculture and current market trends for harvest crops etc.

B) Future scope

Steps to be done for farmers to improve ratings, most preferred crops to be planted this season, maximum capital to be invested in supervised ways in order to maximize harvests, trending agricultural practices. Development of an android app as an expansion of this portal, with limited functionalist so as to keep check of important statistics and news.

IV. PROPOSED SYSTEM

Fig. 1 shows the proposed system for loan/credit repayability scale using machine learning

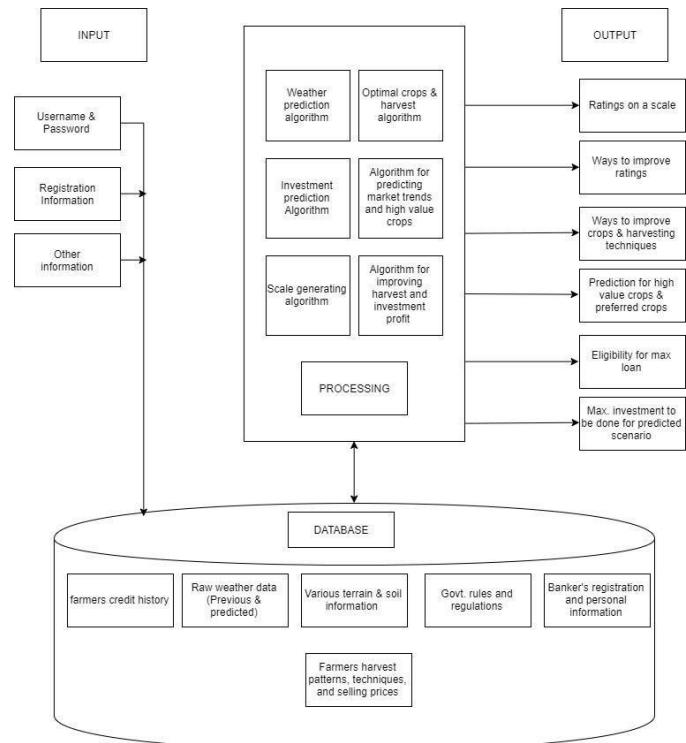


Fig. 1: Block diagram for loan/credit repayability scale using machine learning

V. EXISTING SYSTEM

Many papers have been published involving the use of machine learning in predicting weather forecast and pattern and predicting bankruptcy for farmers. However no research has been made by combining the two fields.

A) Machine Learning Applied to Weather Forecasting

Machine learning applied to weather forecasting by Mark Holmstrom, Dylan Liu, Christopher Vo from Stanford University. Weather forecasting is the task of predicting the state of the atmosphere at a future time and a specified location. Traditionally, this has been done through physical simulations in which the atmosphere is modeled as a fluid. The present state of the atmosphere is sampled, and the future state is computed by numerically solving the equations of fluid dynamics and thermodynamics.

However, the system of ordinary differential equations that govern this physical model is unstable under perturbations, and uncertainties in the initial measurements of the atmospheric conditions and an incomplete understanding of complex atmospheric processes restrict the extent of accurate weather forecasting to a 10 day period, beyond which weather forecasts are significantly unreliable. Machine learning, on the contrary, is relatively robust to perturbations and doesn't require a complete understanding of the physical processes that govern the atmosphere. Therefore, machine learning may represent a viable alternative to physical models in weather forecasting.

B) Automated weather event analysis with machine learning

Automated weather event analysis with machine learning by Nasimul Hasan .Weather forecasting has numerous impacts in our daily life from cultivation to event planning. Previous weather forecasting models used the complicated blend of mathematical instruments which was insufficient in order to get higher classification rate. In contrast, simple analytical models are well-suited for weather forecasting tasks. In this work, we focus on the weather forecasting by means of classifying different weather events such as normal, rain, and fog by applying comprehensible C4.5 learning algorithm on weather and climate features. The C4.5 classifier classifies weather events by building the decision tree using information entropy from the set of training samples.

We conducted experiments on LA weather history dataset; from evaluation results, it is revealed that C4.5 classifier classifies weather events with f-score of around 96.1%. This model also indicates that climate features such as rainfall, visibility, temperature, humidity, and wind speed are highly discriminative toward events classification.

C) Effectiveness of semi-supervised learning in bankruptcy prediction

Effectiveness of semi-supervised learning in bankruptcy prediction by Stamatis Karlos, Sotiris Kotsiantis, Nikos Fazakis, Kyriakos Sgarbas.Adoption of techniques from fields related with Data Science, such as Machine Learning, Data Mining and Predictive Analysis, in the task of bankruptcy prediction can produce useful knowledge for both the policy makers and the organizations that are already funding or are interested in acting towards this direction in the near future.

The nature of this task prevents analysts from collecting large amount of data for building accurate predictive models. Semi-supervised algorithms overcome this phenomenon and can perform robust behavior based on a few data. Experiments using data from Greek firms have been made in this work, comparing many semi-supervised schemes against well-known supervised algorithms and the results are promising.

D) Factors study of credit risks of farmer loans based on projection pursuit

Agricultural production output prediction using Supervised Machine Learning techniques by Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty. Because farmer loans have some features including a small amount loan, incomplete financial information, and large mobility, establishing a credit risk evaluation model that reflects characteristics of farmer loans has great significance for risk management of commercial banks.

This paper uses projection pursuit model to nonlinearly transform multiple evaluation indicators into a single indicator of credit risk, establishes measure model of credit risk for farmer loans. Moreover, the greater squared value of projection weights, the greater impact on credit risk for farmer loans. So this paper makes use of squared value of projection weights to obtain effect sizes of risk factors. The results show that the skills

of lenders and their families have the largest effect on credit risk for farmer loans.

VI. EXPECTED OUTCOME

The outcome of our project is to help the farmers overcome the difficulties they face before and after the cultivation and harvest season in the form of process of guiding the loan/capital requirement and investment and also the bankers/moneylenders to correctly lend the appropriate capital to the deserving agriculturalists.

- It will be made sure that the farmer is well guided and informed about the market and upcoming opportunities.
- A portal shall be developed to help farmers, bankers and moneylenders.
- Scale method will be used to make interaction simpler.
- Raw weather data and complex machine learning algorithms would be used.
- Awareness of impending unfavorable weather conditions on a certain region would be done.
- Support and suggestions on the current market trends and prices would be provided
- Farmers will be assisted in getting loans.
- Machine learning will be used to help the farmers and bankers alike.

VII. CONCLUSION

Through balanced understanding within the team members we have been successful in designing planning and documenting the application and website called “Machine learning approach to predict farmer’s loan/credit repayability using weather prediction and credit history”. We have acutely studied, analyzed and planned the features and function of the system. As we are focusing on helping the bank for repayability of loan and farmer for applying for the loan in the bank.

The application is designed using multimedia which will help the farmer to operate it properly. As we are focusing on making the website and the application. The main concept of while developing the application was the usefulness of the application towards farmer by implementing the multimedia.

The website is being developed so that it will help the Bank to grant the loan to the farmer on the basis of its previous record and the irrigation of the crop within the particular region based on the climate. So our main motive is to make ease in the process of applying for loan and repayability of the loan by the farmers by developing the application and the website.

REFERENCES

- [1] M. Holmstrom, D. Liu, C. Vo, “Machine learning applied to weather forecasting”, IEEE, 2016.
- [2] N. Hasan, T. Uddin, N. Karim Chowdhury, “Automated weather event analysis with machine learning”, IEEE.
- [3] T. Shakoor, K. Rahman, S. Nasrin Rayta, A. Chakrabarty, “Agricultural production output prediction using Supervised Machine Learning techniques”, IEEE, 2014.
- [4] S. Karlos, S. Kotsiantis, N. Fazakis, K. Sgarbas, “Effectiveness of semi-supervised learning in bankruptcy

- prediction" IEEE, 2012.
- [5] R. Caiming, Z. Hu, H. Li, M. Wicks, "Machine Learning Approaches", IEEE, 2010.
 - [6] S. Veenadhari, B. Misra, C. Singh, "Machine learning approach for forecasting crop yield based on climatic data", IJCSE, 2003.
 - [7] A. Aisjah, S. Arifin, "Maritime weather prediction using fuzzy logic in javasea", IJAR, 2012.
 - [8] X. Siu, "Farmers wilfully defaulting on loans in hope of waiver", Timesofindia.indiatimes.com
 - [9] Deepsiman Tiwary, "80% farmer-suicides due to debt, loans from banks, not moneylenders", indianexpress.com
 - [10] The Associated Press, "Farmer Suicides Rise in India as Climate" Warms Weather.com
 - [11] M. Somashekhar, "Climate change seen contributing to farmer suicides in India" thehindubusinessline.com
 - [12] D. Saha, "Farmer suicides hindustantimes.com", IJCSE, 2003.

Android Based Car Tracking System

Karan Pawar

*Thakur College of Engineering and Technology,
Mumbai, India*
karanpawar163@gmail.com

Aditya Gupta

*Thakur College of Engineering and Technology,
Mumbai, India*
aditya.gupta@gmail.com

Radhika Kotecha

*Thakur College of Engineering and Technology,
Mumbai, India*
kotecha.radhika7@gmail.com

Abstract- In today's era when the number of cars and traffic has increased substantially, it is extremely important to track the location of cars – to avoid thefts and for unwanted situations like accident. The paper presents a low-cost, simplified yet versatile and optimized vehicle diagnostic system that is compatible with all vehicles manufactured after 1996 (OBDII compliant). The graphical user-interface (display and command) is Andriod-based and utilizes the popular personal area network (PAN) communications standard Bluetooth to facilitate extraction and relaying of readings, diagnostic trouble codes (DTC), and commands. The paper targets smart phones as the computing device for the obvious growth and demand on such mobile devices, in addition to reducing the overall system cost by utilizing the built-in functionalities that is integrated in such compact devices. When installed properly, such devices can be a low cost alternative to integrated navigation systems. Through the approach proposed in this paper, new business opportunities that provides customer-oriented vehicle diagnosis and remote maintenance can be offered. Also, extraction of fault detection and sensor output signals from the engine control unit (ECU) according to the OBD-II standard, vital vehicle information can be displayed to users, and/or communicated to remote maintenance departments via cellular networks. Hence, if a correct diagnoses (or prediction) of a defect is available, instructions and commands can be sent to the driver on how to proceed in that situation. The results shown in the paper prove the same.

Keywords- Optimization, Defect Diagnosis, Car Tracking, Decision Making, Android

I. INTRODUCTION

While your vehicle is being driven on to the road, anything unfortunate can happen to the vehicle; it may meet some unwanted accident or may be theft. So, it is very necessary to track your vehicle. Keep i.e. track on the status and exact location of vehicle carrying valuable goods at anytime and anywhere is essential for both personal as well as business purpose. You can track your vehicle by calling to the driver over mobile phone but this not convenient as your driver might give you false information. So, at any time you want to be 100% sure about the location and status of the vehicle.

Here, various vehicle-tracking technologies based on the Global Positioning System (GPS) or radio frequency medium can track your vehicle location in real-time and provide better control over your assets[2]. Built-in automobile navigation systems are being equipped with various multimedia capabilities to satisfy the customer expectations of modern functions and features and offer diverse services in the current greatly competitive market state. This reflects on the overall system cost and limits the advantages to high-end vehicle models.

New business opportunities that provide customer-oriented vehicle diagnosis and remote maintenance can be offered. Further, extraction of fault detection and sensor output signals from the engine control unit (ECU) according to the OBD-II standard, vital vehicle information can be displayed to users, or/and communicated to remote maintenance departments via cellular networks where relatively newer cellular access technologies offer future solutions endless choices for high-end applications [4]. Hence, if correct diagnoses (or prediction) of a defect is available, instructions and commands can be sent to the driver on how to proceed in that situation. Also, if the system has knowledge of the location, in addition to an error category, the system can direct the vehicle to the closest vehicle service location in the vicinity.

II. LITERATURE SURVEY

Using GPS and CDMA we can track real time location and radio technology using microwave. Components use to track the location are expensive and coverage area might stutter because of 2G network [1]. This system controlled by an RFID module to switch on and off, GSM module will be used to send the location and GPS will to owner mobile phone. We can extend the use of this system to the car dispatching management field [2]. GPRS/GSM modules are used to find the real time location of the vehicle which is monitored on smartphone using Android User Interface. Power supply to the system is not constant which could go wrong if the battery dies[3]. Functionality of OBD 2 port increase in this project as it shows vehicle status using from ECU showing DTC (Diagnostic trouble codes.). The function is to send the error codes to the nearest Service Station. User can only track the

status of the vehicle using Bluetooth but can't track it remotely [4].

In all the existing systems, we can analyze that vehicle tracking integrates the GPS tracking system9 with existing vehicle alarm or provide alarm features when someone is tampering with vehicle. Before the vehicle is driven away, security threat is detected and the vehicle is able to be tracked over the internet which is done by utilizing Global Positioning Satellites. Through which, data such as Global Position, including latitude and longitude are transmitted over the Cellular network10. This information transmitted from the tracking device is stored in your private confidential account or sent over the wireless network. The data is cross referred on a street level map for viewing.

The crucial disadvantage of the existing system is that the system does not provide street wise address and speed of the vehicle and engine is no way controlled by the existing systems, thus exposing the vulnerability of a system that provides only tracking.

A detailed survey was conducted in this concern. The following are the key finding:

1. This system controlled by an RFID module to switch on and off, GSM module will be used to send the location and GPS will to owner mobile phone.
2. GPRS/GSM modules are used to find the real time location of the vehicle which is monitored on smartphone using Android UI.
3. Functionality of OBD 2 port increase in this project as it shows vehicle status using from ECU showing DTC (Diagnostic trouble codes.). The function is to send the error codes to the nearest Service Station
4. The commissioned a OBD-II linked cellular information-based system driving assistance. The information is run through a web based basic variable fusion algorithm which correlates events related to safety whilst interpreting and presents the results.
5. The information gained from the status of the road and vehicle is sent to the central cellular server and together with the GPS data which allows visual data to be acquired.
6. As developed an application which integrates the OBD-II connection to an external network. The driver is informed of the relevant variants to the OBD-II via the ECU in real-time.

III. PROPOSED SYSTEM

The network stutter faced in [1] can be solved by replacing CDMA module with the GSM module to access the 4g network, RFID tag [2] will make it complex as the car key sensors has to be changed hence the module will be connected to battery directly or OBD 2 port to power the whole device which also provides the solution in [3]. To track the location in [4], a user friendly interface can be made on Android OS which could be accessed by any android running smartphones. The planning process regarding the project is one of the main obstacle as it deals with many different configurations grade

of the hardware components and compatible software used for the project. Deciding to make it bulkier or small would create the main impact on the project. The IC components changes from time to time and the compatible issue would grow bigger as sometimes the software is not updated with the new gen hardware or vice versa. Selecting and designing the right compatible hardware. The options between OBD 2 ELC327 or OBD 2 adapter to USB 2.0 which can change the whole prospect of the project as ELC327 comes ship with Bluetooth 2.0 and interface to check the car error codes from ECU but it might not be compatible with the programmable Arduino Atmega 238 UNO R3.

The problem revolves around the software or programming languages used in the project. Arduino shield will used to program with GSM module but android based programs are written in java which might create a huge compatible issue. And the main issue might be faced while setting up the GPRS/GSM module with the server using http TCP/IP protocol. 7 Implementing on the cars before 1996 production would not be possible as the OBD 2 port was not introduced. The battery consumption for the Bluetooth module is higher which can drain the battery when car is in idle or off state.

This application makes use of the Bluetooth & GPS features of mobile phone as a solution for Vehicle tracking system. It is the technology used to determine the location of a vehicle using different methods like GPS and other radio navigation systems operating through satellites and ground based stations. Even data can be stored and downloaded to a computer from the GPS unit 3 at a base station and that can later be used for analysis. This system is an important tool for tracking each vehicle at a given period of time and now it is becoming increasingly popular for people having expensive cars and hence as a theft prevention and retrieval device. No doubt, Vehicle tracking system whether it is GPS based or any other wireless medium has brought one of the most important technological advances in today's communication field. Now one doesn't have to leave a place to know where a particular vehicle is at a given period of time. The automatic vehicle locating system only with the help of a tiny electronic device and tracking software can detect the real-time location of a vehicle by using the conventional cell phone network and Internet.

The features of the project are as follows:

- a) Real time monitoring of vehicle sensors
- b) Rules concept for checking state of sensors
- c) Digital Dashboard
- d) Acceleration Test
- e) Read Diagnostic Trouble Codes (DTCs)
- f) Clear DTCs and turn off the MIL (engine check light) if on
- g) Retrieval of car details such as OBD standard and Vehicle ID (VIN)

The following tools will be used in the implementation:

1. GSM module.
2. GPS module.

3. ELC327 OBD 2.
4. Arduino UNO R3.
5. SIM 808 Arduino shield (elements).

The entire implementation of the system can be divided into 4 phases, each of which are explained below:

Phase 1 (Design): The main goal for the design is to make the project much compact including all the necessary hardware components. The UI for operating the app is user friendly and simple as it instructs.

Phase 2 (Hardware components testing): Compatible IC Components such as GPS/GSM module, OBD 2 bluetooth and OBD 2 adapter to USB type A. Operating the OBD using ISO-9141-2 protocol and TTL/CMOS based ICs.

Phase 3 (Software components development): The uno r3arduino which is programmable in python language which will interact with GPS/GSM module following sending the data to server which will display the output on the smartphone using Android UI. Android apps are specifically made for programming in java.

Phase 4(Implementing and testing): Using the OBD 2 port to power the system continuously which will trigger if the car is not started and changes its initial position by sending the notification to user's cell phone in the created app by using TCP/IP protocol and https port 80 send by the gsm module to database and server.

Fig. 1 shows the proposed framework for connection from OBD-II to Mobile, which happens to be the heart of the system.

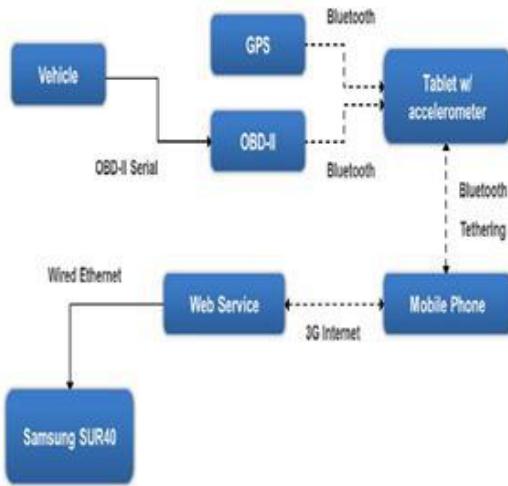


Fig.1 Proposed Framework for Connection from OBD-II to Mobile

VI. CONCLUSION

We propose a universal integrated system which is composed of a combination of a low-cost hardware unit and user-friendly Android-based mobile application software utilized to create an on-board vehicle diagnostic system. The mobile

application software will interact with the hardware interface unit wirelessly via Bluetooth to acquire desired vehicle parameters from the ECU of the vehicle. These readings will be displayed locally to the user then can be sent to a remote maintenance server as HTTP packets via a cellular internet connection. The packets received will be tabulated in the server, then made use of by the maintenance department which holds the server.

REFERENCES

- [1] D. Naikoji, Y. Patwa, M. Shilimkar, M.Nighot, "Car Tracking System Using Android Application", vol. 4,no. 4, UAR CET, IEEE, 2014.
- [2] L. Zhigang,Z.Angi,L. Shaojun, "Vehicle anti-theft Tracking system based on IoT", pp. 48-51, Qinhuangdao, China,IEEE, 2013.
- [3] L. SeokJu, T. Girma, K. Jaerock., "Design and Implementation of Vehicle Tracking System Using GPS/GSM/GPRS Technology and Smartphone Application", Electrical and Computer Engineering, Flint, MI, USA, IEEE, 2014.
- [4] T. Ashraf, S. Ahmad, J. Fouad, Q. Waleed, "Android-Based Universal Vehicle Diagnostic and Tracking System", Communication Engineering Department, Princess Sumaya University for Technology, Amman, Jordan, IEEE, 2012.
- [5] D. Enriquez, "CANOPNR: CANOBD programmable-expandable network enabled reader for real-time tracking of slippery road conditions using vehicular parameters", 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2012
- [6] P. Verma, J. Bhatia, "Design and Development of Gps-Gsm Based Tracking System". International Journal of Computer Science, Engineering and Applications (IJCSEA) vol.3, no.3, 2013.
- [7] Abid khan, Ravi Mishra, "GSM Based Tracking System", International Journal of Engineering Trends and Technology- vol. 3, no. 2, 2012.
- [8] M. A. Al-Khedher, "Hybrid GPS-GSM Localization of Automobile Tracking System", International Journal of Computer Science & Information Technology (IJCSIT), vol 3, no. 6, 2011.
- [9] A. Shrivastava, N. Sharma, R. Arora, S. Dikshit, "GSM Enhanced GPS Based Vehicle Tracking System", 2nd National Conference inIntelligent Computing & Communication organized by Dept. of IT, GCET Greater Noida, India, 2015.
- [10] B. Kodavati, V. Raju, S. Srinivasa Rao, A. V. Prabu, T. Appa Rao, Y. Narayana, "GSM And GPS Based Vehicle Location and Tracking System", International Journal of Engineering Research and Applications(IJERA), vol. 1, no. 3, pp. 616-625, 2014.
- [11] Y. Yalian, "Research and Development of Hybrid Electric Vehicles CAN-Bus Data Monitor and Diagnostic System through OBD -II and Android-Based Smartphones", Advances in Mechanical Engineering, 2013.

- [12] J. Zaldivar, "Providing accident detection in vehicular networks through OBD -II devices and Android-based smartphones", 36th Conference on Local Computer Networks (LCN), IEEE, 2011.
- [13] C. Emilian, "Embedded system for remote monitoring of OBD bus," 36th International Spring Seminar on Electronics Technology (ISSE), IEEE, 2013.
- [14] J. Jhou, "The Implementation ofOBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology", 2nd International Conference on Robot, Vision and Signal Processing (RVSP), 2013. IEEE, 2013.
- [15] B. Sung-Hyun, J. Jang, "Implementation of integrated OBD -II connector with external network", Information Systems, pp. 69-75, 2015.

Supply Chain Management And Data Mining Techniques: A Review

Pranjali Kasture

*Department of Information Technology
Thakur Collage of Engineering and Technology
Mumbai, India
kasturepranjali@gmail.com*

Abstract— Data mining in supply chain management (SCM) and logistics is going to make the world a better place, cheaper, quieter, and more profitable. It builds and directs the interactions between the different stakeholders along the value chain in the process of meeting customer needs and expectations. This paper has given a perspective on basic of supply chain management, issues in supply chain management along with how data mining used in SCM and applications of data mining in SCM.

Keywords- *Data mining, Supply Chain Management, logistics, Value chain, Outsourcing*

I. INTRODUCTION

The supply chain management is a great tool for an enterprise to lower costs, increase revenues, accelerate turnover and enhance core competency. It is difficult for an enterprise to find out the rules between suppliers and customers on the basis of its own business data, and then analyze and make decisions accordingly as before. Without a powerful tool for data analysis and processing, it is not possible for an enterprises in the supply chain to process the information in time, nor can they use the information to react quickly and correctly. Data mining with its powerful function of data analysis and processing will play an important role in supply chain management.

Data mining can have multiple uses for helping to address logistics problems: in the case of transportation, it can assist with selecting the transporters and logistics service providers, with evaluating their performance, and with scheduling vehicles during product distribution; and in warehousing, it can help to monitor products in different locations, and to identify their consumption patterns, among others.

The scope of data mining [1] reaches from anticipating the future purchases (forecasting), the optimal supplier base (procurement), and the status of production and the monitoring of the supply chain ecosystem to avoid disruptions (risk-management) [2]. In textile industry, study about effect of parameters like place, time and conditions of growing cotton, transportation duration and conditions, fabric manufacturing and storage conditions, etc. collected throughout the whole supply chain can specify on the success in applying specific colors on specific textures[3]. Considering the advantages and disadvantages of self-built and outsourcing, usually companies have to make a decision after trade-off. The factors which need to be considered include status of fund, sales volume, management level, and informationization capability.

II. BASICS OF SUPPLY CHAIN MANAGEMENT

A. Supply Chain concept

A supply chain is a network of facilities and distribution options that performs the functions of procurement of materials; transformation of these material into intermediate and finished products & distribution of these finished products to customers. Upstream supply chain are the transactions between an organization and its suppliers and intermediaries.

Downstream supply chain are the transactions between an organization and its customers and intermediaries. Successful SCM allows an enterprise to deliver the right product, at right place, at the right time. Supply chain has 3 main parts—

- **Supply side:** concentrates on how, where from, and when raw materials are procured and supplied to manufacturing.
- **Manufacturing side:** converts raw materials to finished products.
- **Distribution side:** ensures that finished products reach the final customers through a network of distributors, warehouses and retailers

Supply chain management (SCM) is handling and optimizing all the many complicated facets of a supply chain, involving goods and services. Even ensuring timely handoff from manufacturer to shipper to supplier to shipper to buyer is a massive task, but to do it cost effectively and build net value is truly a challenge.

B. Simple model of SC

An organization's supply chain can be viewed from a systems perspective as the acquisition of resources (inputs) and their transformation (process) into products and services (outputs) which are then delivered to customers. A change in supply chain thinking, and also in marketing communications thinking, is the move from push models of selling to pull models as shown in figure 1 or to combine push–pull approaches. The push model is illustrated by a manufacturer who perhaps develops an innovative product and then identifies a suitable target market. A distribution channel is then created to push the product to the market. Pull model is focused on the customer's needs and starts with analysis of their requirements through market research and close cooperation with customers and suppliers in new product development. Here the supply chain is constructed to deliver value to the customer by reducing costs and increasing service quality. The typical motivation for a pull approach is to optimize the production process for customer response, cost and efficiency.

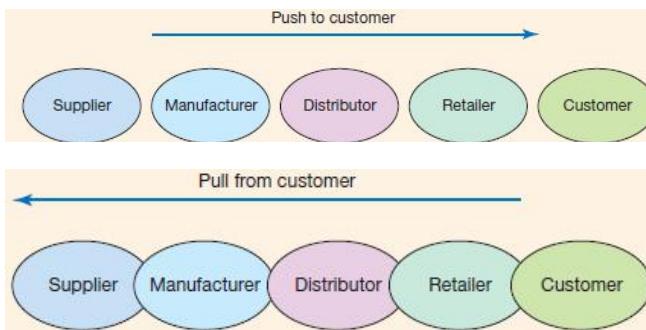


Fig.1 Push and Pull Approaches to supply chain management

C. Supply chain management examples

There are many examples of prosperous companies that have correctly developed the supply change management concept and that enforce efficient practices.

a) Amazon

Amazon offers about everything you can think of and their variety in offers and products along with their customer driven shopping and recommendations is a hit with customers. Their supply chain goes from the lowest levels of inventory, through the logistics of the order itself all the way up to an outstanding distribution chain of their products in an international scale. Amazon can currently ship close to 10 million different products. This diversity gives it an edge against competitors and makes it a perfect example of what efficient supply chain management can accomplish.

b) The Coca-Cola Company

Aside from having an extremely successful supply chain, Coca-Cola participates in sponsorships, partnerships, and alliances; thus creating a great management and marketing of their products.

c) Zara

Zara is one of the main clothing and accessory retailers internationally. They are mainly focused on new, trendy and cutting edge fashion for men and women. Zara is known for its energy saving efforts and techniques, as well as their minimization and efficient waste management.

D. Data Mining in Supply Chain

Data Mining is an ongoing processes. The supply chain model would continue to evolve, as new nuggets of knowledge are extracted from the resultant data. Before the full functional and economic benefits of a computer integrated manufacturing operation can be realized, a method of replacing the intelligence of the displaced equipment operators must be found. A mixture of traditional 'expert' computer control system technology and the emerging Data Mining analysis could be an invaluable tool to the endeavor to address this type of problem [4][5]. The study shows that, e-supply chain management had a positive effect in the electronic manufacturing services industry as these showed that the profits of the firm increased and internal communications were improved due to the implementation of e-supply chain management. Using e-supply chain, there is a rapid rise of Business to Business (B2B) transactions over the internet with the increasing use of e-procurement solutions by large organizations for purchasing [6] and collaboration [7]. RFID devices have been used to optimize the operation of warehouses, stores and transportation logistics [8].

III. ISSUES IN SUPPLY CHAIN MANAGEMENT

In the Supply Chain, risks can occur in supplying a product or service to a customer in terms of cost, timely delivery and impact on image. Some of the issues are mention below.

- All enterprises in the supply chain have to collect massive information on competitors and exchange and share the information continually, which will also increase the traffic in the supply chain.
- Dependence of enterprises on the supply chain has raised business risks. Network systems are built among supply chains with the help of advanced communication technology and the rapid and powerful tools for information analysis and processing.
- Purchasing raw materials and commodities is often underwritten by banks and financial institutions. Transparency of where product is sourced from, under what conditions it was produced, and where it is intended to go are important criteria for the financing decision.
- Internally, within a business, one of the main risks is in understanding the detail of the supply chain and being able to take decisions at the right time. Senior management may not have all the information available all the time to assess the risks in operating the processes.
- The principal problem for an enterprise to carry out supply chain management is to select cooperation partners.

IV. HOW DATA MINING USED IN SUPPLY CHAIN MANAGEMENT

The enterprise can evaluate all resources and historical data with the help of data mining tool. The enterprise can analyze its upstream suppliers and their suppliers, downstream retailers and logistic services with the tool to find out their competitive strengths and individual characteristics, so as to select the best partners with the foregoing analytical results of its own characteristics as references data mining can help the enterprise analyze its organizational structure, information flow, fund flow and logistics as well as those of its selected partners. Data mining technology can search models in the database, quickly and automatically find out the models and useful information hidden in the data, and help users understand reasons and make predictions. Data mining also analyzes customer needs and expectations. It help the enterprise to predict their needs and expectations by processing the collected historical information of customers with data mining.

A. A Key Business Problem

There are two issues that plague supply chain management

- Variation in the demand and supply
- Variation in the speed and extent of communication within the supply chain.

Variation in demand and supply is due to the inherent uncertainty also present in the processes. Accurately predicting the uncertainties in demand, supply, and processes and then formulating action plans around the prediction is the essence of supply chain management. Before we can address the problem of uncertainty in supply chain and explain the use of data-mining techniques, we need to understand the basic process of SCM and where uncertainty exists. In its most simplified form, a supply

chain can be depicted as the flow of information from a customer's customer to a supplier's supplier and then the flow of material in the reverse direction. The whole supply chain can be conceptually broken down as Supply – Process – Demand. Traditional forecasting planners of supply chains use demand and supply forecasting as a means of controlling uncertainty. However, there are three major drawbacks in those methods, like incorrect forecasting model, incorrect number of parameters and incorrect coefficients values of these parameters. Each of these three problems can be solved using data mining. The models are chosen from a finite set of predefined models in data mining. The model can be recreated as many times as needed in order to extract previously unknown patterns and relationships in data. When forecasting using data-mining techniques, the program can detect even minor effects of some parameters.

Data mining is the process of extracting ideas in data. It can also be defined as “a decision support process that tries to discover patterns and relationships that are beyond the realm of human experience and imagination in the large database and present them to a knowledgeable user for review and examination” or “as the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions”. Data mining not only uses a discovery-based approach in which pattern matching and other algorithms are employed to determine the key relationships in the data but also describes the steps that must be taken to ensure meaningful results.

B. Data Mining Methods

Data mining is used to build six types of models aimed at solving business problems:

Classification: A predictive model is generated based on the historical data. These models are used to assign instances to a group or class by calculating the value of a categorical variable. The value of this categorical variable is generally binary in nature. It can include multiple but discrete values.

Regression: Regression is used to predict values for categorical variables. The values are continuous, real numbers, and it has no fixed range in which the values of the variables are fitted.

Time-series forecasting: This method uses a series of existing values and their attributes to forecast future values, except that the values of the categorical variables are dependent on time. Using various data-mining tools, the distinctive features of time can be exploited.

Clustering: Clustering is used to segment a database into clusters, with the members of each cluster sharing a number of interesting properties. These clusters are not predefined and have two basic uses: 1) summarizing the contents of the target databases; and 2) as inputs to the other methods like supervised learning.

Association: Association is used to describe behavior that is captured in the database. This method relates the occurrences of various events by identifying patterns or groups of items.

Sequencing: Sequencing defines items that are likely to occur together on a sequence basis. This could help marketers in timing their promotions to correlate with the sequential buying order exhibited by their customers.

Figure 2 shows all phases of SCM. It includes Integration and information sharing, product development, procurement and manufacturing, logistics, customer service and performance measurement. When mining data in the supply chain management, the data mined refers to all the data collected at all

phases and the generation of the knowledge base that will help the enterprises to improve decision quality and the efficiency of the supply chain management.

V. DATA MINING APPLICATION IN SUPPLY CHAIN MANAGEMENT

a) Printing Industry

Printing industry which can be book publisher or newspaper or magazine is working on mass-customization in many printed items around us. The mass of promotional leaflets that arrive through the mail, most of which have been based on previous purchases. The regionalization of national newspapers and TV listing magazines, in the case of some subscription magazines they have begun to include various sections based on the customer segment you belong.



Fig. 2 Supply Chain management phases

b) RFID

RFID tags are a relatively recent innovation in e-SCM that are already widely used for logistics purposes. They can be attached to individual product items in a warehouse or in a retail location. With appropriate scanning technology they can then be used to assess stock levels. They can be read at a distance of 1 to 6 meters. RFID technology holds promise in transforming supply chain management by providing real time intelligence for tracking enterprise assets. RFID is to manage the entire life cycle of an asset by determining the time of initial asset acquisition, the asset's physical location, the asset's movement within a data center and the time of the asset's ultimate decommission [9].

c) Supply chain fraud management

Increasing use of internet transactions generate large data volumes and the inability to analyze them. It enables fraudulent activities to go unnoticed in supply chain management processes such as procurement, warehouse management and inventory management. Data mining technique used often in forensic accounting for fraud detection has been proven useful in supply chain fraud management.

VI CONCLUSION

Data mining has emerged as a vital tool in the case of supply chain management. It enables seamless integration of complex networks like inventory, production costs, and customer needs to

accelerate core competencies. In an industry, quick turn-around time is the key for gaining higher market share. It is essential for an enterprise to take the right decisions at the right time. Data mining is a tool, which facilitates easy integration of various verticals in the supply chain.

REFERENCES

- [1]. Shuanghong Liu , Jianguo Yu , "Supply Chain Management Research
- [2]. Based on Data Mining," The Conference on Web Based Business Management, 978-1-935068-18-1 © 2010 SciRes. 2010
- [3]. Matthew A. Waller1 and Stanley E. Fawcett, "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management," Journal of Business Logistics, 2013, 34(2): 77–84 © Council of Supply Chain Management Professionals
- [4]. Real Carboneau, Kevin Lafraimboise, Rustam Vahidov," Application of machine learning techniques for supply chain demand forecasting," European Journal of Operational Research 184 (2008) 1140–1154
- [5]. Grittner, D. H., & Valverde, R. (2012). An object-oriented supply chain simulation for products with high service level requirements in the embedded devices industry. International Journal of Business Performance and Supply Chain Modelling, 4(3-4), 246-270.
- [6]. Valverde, R. (Ed.). (2012). Information Systems Reengineering for Modern Business Systems: ERP, Supply Chain and E-Commerce Management Solutions: ERP, Supply Chain and E-Commerce Management Solutions. IGI Global.
- [7]. Stephens, J., & Valverde, R. (2013). Security of e-procurement transactions in supply chain reengineering. Computer and Information Science, 6(3), p1.
- [8]. Avédiassian, A., Valverde, R., & Barrad, S. An Extension Proposition for the Agent-Based Language Modeling Ontology for the Representation of Supply Chain Integrated Business Processes
- [9]. Adoga, I., & Valverde, R. (2014). AN RFID Based Supply Chain Inventory Management Solution for the Petroleum Development Industry: A Case Study for Shell Nigeria. Journal of Theoretical and Applied Information Technology, 62(1), 199-203
- [10]. Chen Lei, Tian Zhiyong, Research of RFID Data Mining Based on Supply Chain Management, 978-1-4799-4565-8/14/\$31.00©2014 IEEE

Survey on Improving Web Search By Analyzing Various Facets

Gunjan H. Deshmukh

*Department of Computer Science and Engineering
G.H.Raisoni Institute of Engineering and Management,
Jalgaon, India*

Govinda Borse

*Department of Computer Science and Engineering
G.H. Raisoni Institute of Engineering and Management,
Jalgaon, India*

Abstract— Query Faceted search is a way for searching users to find, analyze, and navigate through search data from online web pages. It is widely used in e-commerce and digital libraries. An effective approach for faceted search is the scope of this implementation. The majorities of the presented facets creation systems are based on a precise province or already defined facet categories. For example, Web search mining for an unsupervised contents by involuntary separate the facets which are relevant to the result that already search for personal web search as user search interest pattern from databases. Facet collection is generated for collection, rather for a single given query. Proposed facets searching system used for information search and media exploration in online search results. Proposed system extracts and aggregates the useful information from the specific knowledge database Wikipedia. In this paper, a proposed system explores to involuntarily find query related aspect of search for open-domain queries in the search engine. Facets of a query are directly mined from the top important searches related to the query without any additional domain knowledge required. As query facets are excellent summary of a query and are potentially helpful for users to appreciate the query and assist them discover information, they are probable data sources that enable a general open-domain faceted exploratory search.

Keywords- Facets generation, Query aspects, Query Reformulation, Seed Collection, Smart Crawler

I. INTRODUCTION

One important thing in query search is a collection of elements that describe and summarize an important aspect of a query. Here, a facet element is usually a collection of word or a phrase. A query can have multiple aspects that summarize the query information from various perspectives. The facets of the "look" query concern the knowledge of watches in five unique keywords, which include brands, gender categories, support characteristics, styles and colors [4]. Query aspects provide interesting and useful information about a query and, therefore, can be used to improve research experiences in many ways. First, we can show the facets of the query with the innovative results given by search engine appropriately. Therefore, users can understand some important facets of a query without go through number of pages [10]. In this paper, a proposed system scans to automatically identify the look related to searching for open domain queries in the Web search engine. The facets of a query are directly extracted from the consequences of the main query web search lacking the requirement for further domain knowledge. Because the aspects of the consultation are summaries in well manner and are truly helpful for users to know the query and aid them to get the

information, data sources are possible that allow a general multifaceted exploratory search of open domain.

II. MOTIVATION

To collect the aspects of the consultation, we suppose that the lists of the same website may contain duplicate information, while the different websites are independent and each one can contribute with a separate vote for the facets of the weighting. However, we have found that sometimes two lists can be duplicated, even they are coming from various websites. For example, mirror sites use different domain names, but they bring out duplicate content and contain the same lists. Some content originally created by a website might be re-published by other websites, hence the same lists contained in the content might appear multiple times in various websites. Another thing is, the various websites may bring out the content by making the use of identical software so there are possibilities that the software may generate similar lists of various websites. Here time to execute that all processes will be more. While searching on web user required more time and relevancy of result is not maintained.

III. OBJECTIVE

1. To generate automatic facet mining.
2. To cluster facet as per the various category.
3. To display ranked facets to user for making searching more efficient.

IV. REVIEW OF LITERATURE

- In this survey author designs solutions for collecting query aspects from search document for user expected search data. It assumes that query aspects are relevant search document parsed important list and using that list query facet can be mined . It is called automatic mining of query Facet by clustering from free text and HTML tags in search results. Author further apply fine grained similarity to avoid similar contents in list [10].
- In this thesis author consider a novel semantic presentation for query subtopic is implemented, which covers phrase embedding approach and query classification distributional representation, to solve those problems mentioned above. Additionally this approach combines multiple semantic presentations in the vector space model and calculates a similarity for clustering query reformulations. Furthermore, automatically discover a collection of subtopics from a query given by user and each of them are presented as a string that define and disambiguates the search intent of the original query. Query subtopic could be minded from various resources involving query suggestion, top-ranked search results and external resource [2].

- In this article, author represents concept of query aspect to know the user interest for search on various topics, where every facet presents a collection of words which explain a user intention for searching that a query. Investigated approach generates subtopics related to query factors and proposed faceted diversification approaches. The actual query aspects are investigated to give more specific search to user such as collecting facets and exploratory search. [1]
- In this thesis, author presents OLAP (On-line Analytical Processing) model for online analysis of user interest to extract query aspects with OLAP capabilities, existence of facet mining was related to the data over relational database, to the free text queries from metadata list style content. This is an extension shows efficiently facet extraction by a search engine to support correlated facets - a more difficult data model in which having the values related with a searches in numerous facets which are not autonomous [5]
- In this survey author proposes a random faceted search approach for searching query driven analysis on data with both textual content and structured attributes. From a facet query, user expected to dynamically choose a set of interesting aspect and present to a user. Similarly in OLAP exploration, author defines interestingness as how an collected value is, based on a given probability [6].
- Author of this term paper searched the new techniques based on a graphical representation to extract query facets from the collection of noisy data. The graphical representation tells how likely a candidate form is to be a aspect string and how two terms are clustered jointly in a query feature, and captures the dependency between the two factors. This work proposes two mechanism for aggregation of an inference on the graphical representation since exact deduction is obdurate [4].
- A hidden webpage extraction from an association makes easy to get to the maze by allowing end user to enter queries by a search engine. In other way, data collection as of such a cause is not by implemented in hyper link. As an alternative, data are collected by querying the boundary, and understanding the consequence page with randomly generated [3].
- This paper resolve problem of relevant search by using the stuffing of pages to focus the search on a subject; by prioritizing capable links within the topic and by also following links that may not lead to instant advantage. This paper recommend a new technique whereby searching involuntarily find out patterns of useful links and apply their spotlight as the creep progresses, thus mainly plummeting the amount of requisite physical setup and change [8].
- In this paper author design a two-stage crawler, namely Smart Crawler, for related harvesting deep web pages. In the primary stage, Smart Crawler performs web site (URL) based penetrating for hidden web pages with the be of assistance of search engines, avoiding visiting a huge quantity of pages. To accomplish extra efficient outcome for a focused crawl, Smart Crawler position webpage to prioritize extremely related data for a specified search query. In the next stage, Smart Crawler achieves rapid in site web crawling by extracting mainly related links with an adaptive link prioritizing [7].
- The paper designs the problem in the structure consisting of relevance model and type model. The relevance model shows whether the article is important or not to search query. The type model indicates whether or not a document belongs to the

collected or prescribed document type. This combines three methods for data collections as linear grouping of scores, threshold on the type score, and a hybrid of the previous two methods [9].

VI. EXISTING SYSTEM

- Query Reformulation and Recommendation:*

Query reformulation and query recommendation are two trendy customs to help users better illustrate their information necessitate. Query reformulation is the procedure of changing a query that can better match a user's information need and query recommendation techniques generate alternative queries semantically similar to the original query.

- Query-Based Summarization:*

Summarization algorithms are divided into various categories in terms of their synopsis building methods ,the amount of sources for the synopsis of information in the summary (indicative or informative), and the relationship between summary and query (generic or query-based). The distinction is that the majority presented summarization systems donate themselves to create summaries by means of sentences collected from documents

V. PROPOSED SYSTEM ARCHITECTURE

Fig.1. shows the input to the system, workings of each input fields describe are as follows:

1. Seed collection:

Here input to system is collect from online API. Which accepts the query and according to query it gives links according to query. After that reverse searching is performed to find seeds are relevant to query or not.

2. Unique website identification:

Here unique URL Only finds and that unique only passes to next step. We performing these step after getting seeds from seed collection by matching two pages content.

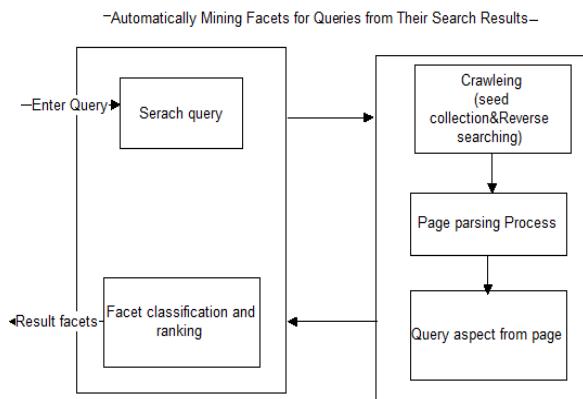


Fig.1 Proposed System architecture of facet mining

3. Page parsing process:

For a list extracted as of a HTML element like SELECT, UL, OL, or TABLE by pattern. That contain facet and links that will display to user.

4. Query aspects from page:

After performing page extraction we get facets and links. SELECT For the SELECT tag, we simply extract all text from their child tags (OPTION) to create a list. UL/OL

For these two tags, we also just take out text inside their child tags (LI). For a list extracted from a HTML element like SELECT, UL, OL, or TABLE by pattern HTMLETAG, its context is consist of the current factor and the previous and next element if any.

5. Facet classification and ranking:

Facets are clustered as per the various classes. It cluster data of similar facets and rank the facets high-quality facet should often come into view in the top results, a facet c is more important. Model (DOM) is applied over html document by parsing html tags. Design grained similarity to classify by comparing their similarity. record clustering alike lists are collected in concert to create a facet. For example various lists about watch gender types are collected because they split the same items men and women.

Which accepts the query and according to query it gives links according to query. After that reverse searching is performed to find seeds are relevant to query or not. For a list extracted from a HTML element like SELECT, UL, OL, or TABLE by pattern HTMLETAG are parsed and facets are finds. List and context extraction Lists and their context are extracted from each document in R. “men’s watches, women’s watches, luxury watches,” is an example list extracted. List weighting All extracted lists are weighted, and thus some unimportant or noisy lists, such as the price list “299.99, 349.99, 423.99, . . .” that occasionally occurs in a page, can be assigned by low weights. List clustering (Classification)Similar lists are collect to create a facet. For instance, poles apart lists about watch gender types are collected because they share the same items “men’s” and “women’s”. Facet and item ranking Facets and their items are evaluated and ranked . For instance, the facet on brands is ranked superior than the aspect on colors based on how repeated the facets occur and how pertinent the sustaining documents are. Within the query facet on gender categories, “men’s” and “women’s” are ranked higher than “unisex” and “kids” based on how frequent the items appear, and their order in the original lists.

VII. ADVANTAGES

1. Will applicable for facet extraction for data mining.
2. Facet mining for data extraction in big data and hadoop.
3. Recommendation system application can use it.
4. Users get relevant result
5. Online facet mining for user attention mining.

VIII. CONCLUSION

Mining information in the form of facets from record by parsing html labels from the report. Proposed mining accomplish fine grained features from search consequence for client search query relevant URLs are gathered by applying reverse search algorithm. This archive is grouped by facet mining. QD miner works for searching facets based over search result.

REFERENCES

- [1] Sha Hu, Zhi-Cheng Dou, Xiao-Jie Wang, “Search Result Diversification Based on Query Facets”, IJAR, 2012.
- [2] Lizhen Liu, Wenbin Xu, Wei Song, HanshiWang, “Query Subtopic Mining by Combining Multiple Semantics”, 2008.
- [3] Cheng Sheng1 Nan Zhang Yufei Tao1, Xin Jin, “Optimal Algorithms for Crawling a Hidden Database in the Web,” in Istanbul, Turkey. Proceedings of the VLDB Endowment, Vol. 5, No. 11, 2011.
- [4] Weize Kong and James Allan Center for Intelligent Information Retrieval, “Extracting Query Facets from Search Results,” Dublin, Ireland, 2013.
- [5] O. Ben-Yitzhak, N. Golbandi, N. Har’El, R. Lempel, A. Neumann, S. Ofek-Koifman, D. Sheinwald, E. Shekita, B. Sznajder, and S. Yogev, “Beyond basic faceted search,” in Proc. Int. Conf. Web Search Data Mining, pp. 33–44, 2008.
- [6] D. Dash, J. Rao, N. Megiddo, A. Alamaki, and G. Lohman, “Dynamic faceted search for discovery-driven analysis,” in ACM Int. Conf. Inf. Knowl. Manage., pp. 3–12, 2008.
- [7] Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang, Hai Jin, “SmartCrawler: A Two-stage Crawler for Efficiently Harvesting Deep-Web Interfaces,” in IEEE Transactions on Services Computing, 2015.
- [8] Luciano Barbosa, and Juliana Freire, “An Adaptive Crawler for Locating HiddenWeb Entry Points,” Banff, Alberta, Canada, 2007.
- [9] Jun Xu1, Yunbo Cao1, Hang Li1, Nick Craswell2, and Yalou Huang3, “Searching Documents Based on Relevance and Type,” in ECIR 2007, LNCS 4425, pp. 629 – 636, 2007.
- [10] Zhicheng Dou., Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song, “Automatically Mining Facets for Queries from Their Search Results”, IEEE, 2012.

A Literature Review on Drowsiness Detection System

Raees Ahmad Noor Mohammad

Department of Electrical Engineering
Theem college of Engineering, Boisar
shoaib128@gmail.com

Rahatullah Khan

Department of Electrical Engineering
Theem college of Engineering, Boisar
rahat600@gmail.com

Abstract— The number of deaths and injuries are increasing in traffic accidents every year. There are various reasons have been identified by the respective authorities and researchers. Drowsiness is found major reason in maximum accidents. Driving is a complex activity, needs active and fatigue free driver. Drowsiness is very complex in nature to identify as it may vary person to person. Though alcohol detectors are available to detect drunk-driver, which is the second largest cause of road accidents. Drowsiness minimizes the reaction time of driver as well as steering behavior. Due to drowsiness various physical, social and economic losses may occur. To avoid such situation, one way is to alert driver if he/she not concentrating on driving by giving any signal. In this paper, we discuss and review different drowsiness detection system for detecting driver's Drowsiness.

Keywords- *Intrusive, Non-Intrusive, Electroencephalogram, eElectrooculogram, etc.*

I. INTRODUCTION

India with more than 1.21 billion population stand second behind China. India's population is 1/6th of world population. The development of a country depends on its youth population, in this segment India is on top but still not completely developed. We are facing lot of problems which restricting development of the country. One of the problems is Road accidents. According to report presented by Ministry of Road Transport and Highways Government of India in 2011 country saw 4.97 lakh road accidents which is 1 accident per minute. Resulting 1,42,485 deaths in the year 2011. Some interesting figures have been collected from 2002-2012 in table1.1 [1]. Though in 2012 road accidents were reduced by 12%, but fatal accidents and Severity percentage was increased. The official numbers of year 2013, 2014, 2015 and 2016 collected and found that in 2014 India ranked 6th place worldwide in terms of percentage risk due to road accidents. From table 1.1 it can be observed that road accidents are not consistently increasing every year but severity level has been increased drastically. There are various reasons which are responsible for accidents to occur. The details of causes of Road accidents as given in [1] by Government of India is summarizes as follows.

For major of accidents occurred "Driver" is found responsible with different reasons. To address the issue of Road safety different pronged strategy may be adopted as:

- i) Engineering
- ii) Enforcement
- iii) Education
- iv) Emergency Care

From above discussion the engineering aspect which can be the solution to minimize the Road accidents as well as development in technology motivated to work on this project. As various functionalities have been added in different vehicles to avoid or sometimes to minimize the effect of accidents.

Driver's drowsiness or fatigue has been found as one of the main causes of accidents. Many of combination-unit truck crashes occurred due to drowsiness. Intensive research has been undertaken to try to assess the part of the total accident that could be a consequence of drowsiness. Two major areas are investigated.

Table 1.1: Accidents Report from 2002-2016

Year	No. of Accidents and No. of Persons involved: 2002-2016				
	Total	No. of Accidents		No. of Persons	
		Fatal	Killed	Injured	Severity*
2002	407497	73650(18.1)	84674	408711	20.8
2003	406726	73589(18.1)	85998	435122	21.1
2004	429910	79357(18.5)	92618	464521	21.5
2005	439255	83491(19)	94968	465282	21.6
2006	460920	93917(20.4)	105749	496481	22.9
2007	479216	101161(21.1)	114444	513340	23.9
2008	484704	106591(22)	119860	523193	24.7
2009	486384	110993(22.8)	125660	515458	25.8
2010	499628	119558(23.9)	134513	527512	26.9

2011	497686	121618(24.4)	142485	511394	28.6
2012	490383	123093 (25.1)	138258	509667	28.2
2013	486476	122589(25.2)	137572	494893	28.3
2014	489400	125828(25.7)	139671	439474	28.5
2015	501423	131726(26.1)	146133	500279	29.1
2016	480652	136071(28.3)	150785	494624	31.4

***Accident severity: No. of Persons killed per 100 accidents**

The first area intends to have better understanding of drowsiness mechanisms in relation with various factors in order to propose recommendations to reduce the risk. The second area aims at proposing in-vehicle alerting devices to counteract driver drowsiness. Devices can be grouped into two categories. In 1st category, the devices are designed to sustain alertness by some form of activation of the driver throughout the trip. The 2nd category includes devices, which detect drowsiness followed by a warning or an intervention. These devices are based on the continuous measurement of indicators used for assessment of drowsiness or of impaired driving performance. Several researchers considered the video-based system. Drowsiness can be measured from slow eye closure more scientifically i.e observing eye blinking of a driver we can predict the status of driver's fatigue.

Moreover, the same eye blinking action can be used to generate some types of signal and can be used to perform particular voltages on the scalp [4]. In this paper from EEG signal Electrooculogram (EOG) which is generated due to eye blinking/ eye movement is determined. To filter out EOG from EEG signals H^∞ adaptive filter is used. Two set of EEG signal were recorded by using electrode for 134 sec duration which is sampled at 256 Hz. EEG and EOG signals were filtered using two different 100-tap linear phase FIR band pass filter. From result it is concluded that presented algorithm based on H^∞ adaptive filter performs better than previous algorithm which was based on LMS algorithm.

In the year 2010, Junfeng Gao, Yong etc proposed a new technique for removal of Artifacts from EEG signal [5]. They have decomposed EEG signal into Independent component by Independent component analysis (ICA). Authors have developed Peak Detection algorithm of independent components to identifying the eye blink. They have experimented 15 different healthy subjects each for two minutes. Electrodes were placed on different 16 channels and got the sensitivity of 98% and specificity of 97%.

EEG and EOG signals are complex in nature and need to process precisely therefore number of methods have been

investigated by researchers. In [6] EOG signals were recorded between Cornea and Retina. Discrete wavelet Transform is used to decompose EOG signal in 4 levels. By setting a threshold value for coefficients Eye Blinking is determined. One EOG sample was taken from standard database and experimented for different Wavelet Transform thresholding which results in maximum of 100% efficiency for eye blink detection and minimum of 95%. It is concluded that Symlet based approach is best. Continuous Wavelet Transform in accordance with SVM method is developed for Eye blink detection from EOG Signals is introduced [7].

The combination of Brain and visual activity is presented [8] for drowsiness detection. Brain activity is measured using single EEG channel. Through eye blinking visual activity is recorded. Eye blinking features then extracted from EOG channel. Authors have calculated drowsiness from visual and brain activity separately and merged both results using fuzzy logic. Three different drowsy levels were indicated as "awake, drowsy and very drowsy", by using cascading rules. Different 20 targets were tested by the system and efficiency is found to be 80%. Finally, it is concluded that for high efficiency high frame rate video must be processed. A wireless Brain-Computer interface system [9] has been presented by Chin, Che-Jui Chang etc. In this paper physiological signal acquisition module is used. Physiological signals are processed using embedded signal processing module. Authors have adopted same technique to identify drowsiness as in previous paper except wireless physiological signal acquisition module and embedded module. The system's performance has been checked by recording sample EEG signals under alert and drowsy condition, these samples used as data base and matched with recorded sample in experiment.

Though efficiency of 76% - 88.7% is recorded but experiment was performed on 4- lane highway on a straight line, therefore system's performance on ordinary road condition and crowded traffic condition were not discussed. Hence practical application of the system seems difficult. A generalized EEG based drowsiness prediction system [10] has been proposed by using a self-organizing neural fuzzy system. As name shows the basic of drowsiness detection is same as in previous work except fuzzy system. In this paper Virtual action as in case of human-machine interface.

Physically challenged person can operate computer by blinking the eyes. Moving a cursor, opening and closing a file can be done on eye blink. Hence eye blink detection is considered as useful parameter by which we can extend technology at next level.

Loss of driver alertness is almost always preceded by psycho physiological and/or performance changes. Unfortunately, drivers themselves are often unaware of their own deteriorating condition or even when they are aware are often motivated to keep driving. Perhaps one

reason for their perseverance is the knowledge that alertness level often fluctuates during prolonged task performance.

The term “drowsiness” is used here to refer to the state of reduced alertness, usually accompanied by performance and psycho physiological changes that may result in loss of alertness or being “asleep at the wheel”. The term “driver fatigue” is also widely used to describe this condition, especially on Police Accident Reports and in accident data files. [2]

Another important distinction is that between “alertness” and “attention” Driver alertness (“awakeness”) is presumed to be necessary but not sufficient for an appropriate focus on external events i.e. attention or vigilance. Thus, drivers may be alert (i.e. awake) but still inattentive. In the context of driving, “inattentive” means that a driver has failed to perceive a visible crash threat due to “mind wandering,” distraction (internal or external to the vehicle), or “improper lookout” i.e. “looked but didn’t see”. The driver information processing error of inattention is widely regarded to be the most frequent principal causal factor in traffic crashes, greatly surpassing loss of alertness. [2]

As one of the most salient facial features, eyes, which reflect the individual’s active states and focus attention, are one of the most important information sources for face analysis. Efficiently and accurately locating the eyes positions in a given face image is therefore essential to a wide range of face-related research efforts, including face alignment, face recognition, gaze estimation, pose estimation, expression analysis etc. and has gained increasing attention from both the academic and industrial communities in the last three decades. However, the task of accurate eye localization is challenging due to the high degree of eye’s appearance variability. This variability may be caused either by intrinsic dynamic features of the eyes or by ambient environment changes. [2][3]

II. EXISTING METHODS

In the past decade, driver safety problems got special attention over many countries. Researchers are working on driver drowsiness detection. Various techniques, such as physiological detection, driver behavior monitoring, vehicle running status analysis and vision-based detection have been worked out. Though lot of work has been done, still the field remains talking point of researchers. The different methods can be classified broadly as

- i) Intrusive method
- ii) Non-Intrusive method

Intrusive Method

In this method electrodes are used to identifying drowsiness. The idea behind this technique is Reality (VR) based dynamic driving simulator is used to collect experimental result. Support vector Machine is employed here for solving the problem of function approximation

and regression estimation. Overall efficiency of system is found more than 95%.

As above literature reflects all the methods are based on EEG and EOG signal. The EEG and EOG signal were collected by connecting electrodes on the body of target which makes it annoying for target. Moreover, processing of such signals is tedious task hence Kamil, Krzysztof etc. have presented drowsiness detection system using Microwave Doppler Sensor [11]. The paper is based on eyelid detection, estimation of eye blink duration and eye blink frequency. The concept of Doppler Radar system has been employed. Multiple receiver architecture is used in sensor. A SSB modulation of transmitted signal has been applied in order to allow for beam steering at the intermediate frequency. The recorded signals by the sensor are analyzed in frequency domain using Fast Fourier transform (FFT) and Wavelet Transform is used to analyze in time-frequency domain. System’s efficiency of 84% - 94% were observed depending upon the position of head. Dongrui Wu and others have implemented Brain-Computer interface in real application in [12]. Online weighted adaptation regularization for regression (OWARR) algorithm is proposed by authors. It was found that different disturbances occurred randomly for various 16 subjects. The total experimentation is based on approximated input record which has been recorded for 3600 seconds. Authors have proved that this method has less estimation error as compared to previous works. Cheon and Kang proposed sensor-based system [13]. The PhotoPlethysmoGraphy (PPG) sensor has been employed to collect different bio signals. These bio signals have been processed through data segmentation and averaging before applied to support vector machine (SVM). The complete model is shown in figure 2.1. The results obtained in this paper while considering error rate, precision, accuracy and recall. The efficiency of 96.3% obtained by the author and also presented Driver’s condition model which can be used onboard.

Non-Intrusive Methods

Unlike intrusive method, images are used to detect eye position; eye blinking and eye blink frequency to calculate drowsiness detection. The general block diagram for this method is shown below:

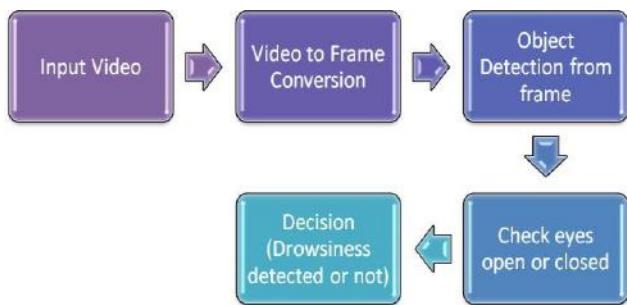


Fig. 1: Block diagram non-intrusive method

Many researchers worked to solve this problem and still it remains a topic to talk about. While calculating drowsiness various parameters should be considered by the designer. Face detection and tracking is presented using template matching [14]. In this paper after tracking the face, eyes were detected using a thumb rule that eyes will occupy the upper portion from centre of face. Moreover, eyes are always symmetrical in nature; hence by setting a threshold value left and right eyes were identified. The result shown by author are useful for application like human-machine interface, wouldn't be useful for drowsiness detection due to its rigid nature of computing.

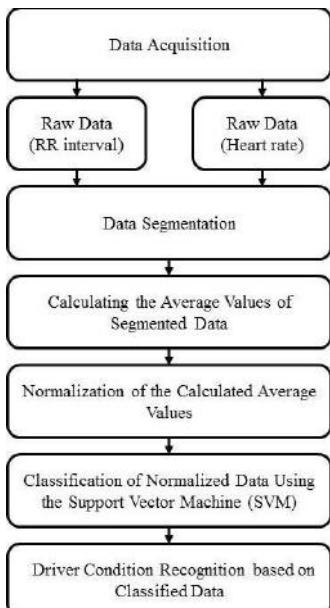


Fig. 2: Flow chart for Driver's Condition Classification

To overcome this limitation other method with Variable lighting condition and various face orientations were considered [15] by Zhiwei and Qiang Ji. Authors have divided system in two parts: eye detection and eye tracking. Bright/dark pupil effect under active IR illumination and the eye appearance pattern in ambient illumination using SVM accomplished the eye blink detection. Eyes were tracked using Kalman filter as well as mean shifting to improve the performance of the system. By processing on images with different head position with different illumination it is concluded that by combining two eyes tracker with SVM makes a robust system with high efficiency. In previous work SVM were used which makes the system more robust but on the price of increased complexity level. A passive method is presented without SVM for drowsiness detection [16]. In this paper simple web cam was installed on car dash board. Viola-Jones technique was used to recognize the face from image. Then eyes were searched on detected face portion. If eyes were found, stored as template; if not last saved template is used in template-matching procedure. From eyes, eye blinking was determined by converting eyes template in binary form. Then after a

specified time if eyes were closed or open continuously, it was concluded that the driver is in drowsy condition.

The combination of multiple eye detection and tracking is presented [17] by Francesco and Giancarlo. In this paper they have combined two different approaches to detect and track eyes. Simple competitive and boosted competitive approaches were combined considering that change in situation will lead to limit the working of one approach. Single camera remote eye tracker was used to perform experiments. It is found that using two different approaches we can get robust and efficient system to tackle the problem of eye detection and tracking which can be the principle operation in other application. Color based approach [18] is presented by Axel, Ayoub and Bernd. The authors have used Viola-Jones algorithm for face detection. By using cascade model eyes were detected and fixed ROI was set to avoid extra computation. A color has been used as potential feature. From mean color value around the eyes will gives the information about eyelid movement. Result which obtained by the authors shows the good efficiency and even system works target having very dark eye lashes. The problem of glass wearing target is also solved. The problems due to change in illumination and driver posture [19] were solved by Wei Zhang, Cheng and Lin. Authors have introduced robust algorithm to solve above problems. Percentage value of eye lid closure, maximum closure duration, blink frequency, average value of opening of the eyes and closing velocity of the eyes were calculated. Adaboost based face detector is used for face detection. Eyes were identified by active sharp model. The problem of illumination is resolved using self-quotient image instead of original image. Mean-shift algorithm is used to achieve robustness in the proposed system. The result shows the accuracy of more than 86% is achieved.

More efficient system is presented by Yuriy, Francesco and Mirabelli. Proposed system is based on single web camera placed on front of the face. For fast detection of eyes Viola-Jones algorithm with Haar-like features as input to classifier [20] is used. Eyes were tracked but not on all frames by assuming that person during work does not move frequently and rapidly with static background which saves the processing time. For eye blink detection ROI is converted into gray level and further binarized. With setting a threshold value eye blink is detected. Obtained results reflects the 94% overall efficiency. A comparative study on automatic eye blink detection is presented by Kyril, Stefanos and Maja PAntic. Viola –Jones algorithm is used by the author for face detection. Eyes are detected by employing eyeAPI [21] which uses isophote curvature. Appearance-based eye tracker is used to track the eyes and hence ROI. Eye blinking, Blink duration and Blink frequency are calculated. Authors have concluded that System with

Gabor filter generates good result. Large scale Naturalistic driving data processing [22] for Eye/ head tracking is proposed. Authors have developed algorithm which enhances the signal quality captured from video of driver data

and increases the data handling quality. Authors have processed the data which was acquired by SeMiFOT project. SeMiFOT project was involved with 44 unique drivers and 13 eye tracker equipped vehicle had been used to drive 10,000 trips. It has been concluded that while handling large database post-enhancement and quality handling are the critical parameters and need to be researched furthermore. In [23], VLSI architecture is utilized with image processing for better performance. The proposed flow chart by the author is shown in figure 3.

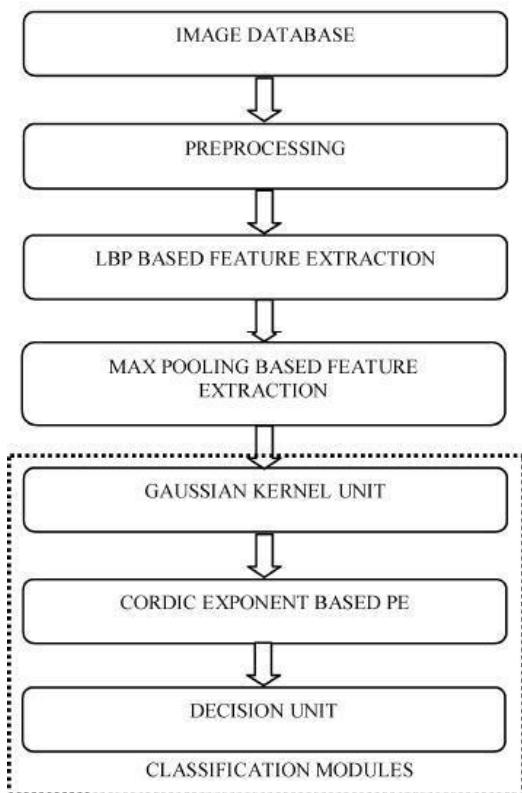


Fig. 3: Drowsiness Detection System Using VLSI

The decision-making unit is completely done by VLSI technology. Since threshold value is converted into binary values and hence digital comparator using VLSI makes fast decision resulting in reduced conversion time. This kind of technology incorporation makes system fast and closer to real-time which is the prime requirement of the problem considered. The decision-making unit is shown in figure 2.3. Raspberry Pi Camera with Raspberry Pi 3 module is incorporated with previous technology and accuracy of 99.59 were identified over 10 experiments [24]. Only yawning has been explored to identifying drowsiness [25]. This is the limitation of the proposed scheme because drowsiness can't be defined completely with this aspect. Deep Neural network is employed with embedded system for detection of drowsiness [26]. Due to

presence of Neural network, system has light weight with less structure for simple installation process. The results of 89.5% accuracy achieved at 14.9 frames per second.

Other Methods

Driver operation and vehicle behavior can be implemented by driver's behavior while driving. It is noticed that in abnormal condition the steering wheel movement, accelerator or brake patterns with vehicle speed as well as lateral acceleration and lateral displacement is observed as abrupt.

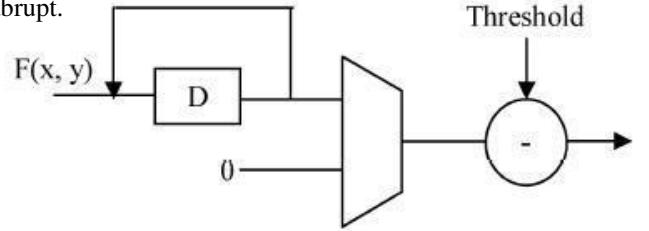


Fig. 4: Decision Unit based on VLSI technology

This kind of movement can be useful to determine driver's fatigue. These too are non-intrusive ways of detecting drowsiness, but are limited to vehicle type and driver conditions. The final technique for detecting drowsiness is by monitoring the response of the driver. In this method Driver is asked to give the signal regularly so that system will consider that driver is alert otherwise defined action will take place. The problem with this technique is that it will eventually become tiresome and annoying to the driver [30]. In [31], GSM is implemented onboard and controller has been incorporated with different sensors as shown in Fig. 5.

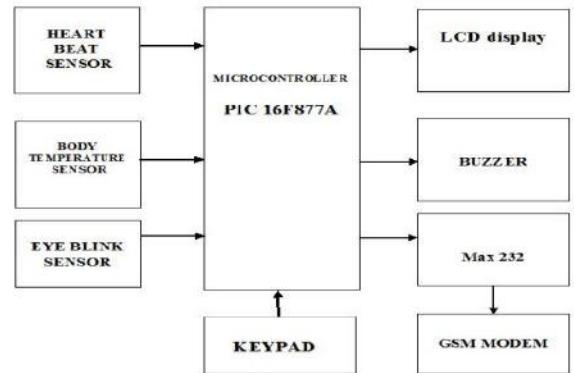


Fig. 5: GSM Based Drowsiness Detection System

This system will work only for ethical drivers. Kashyap Vasudevan and other have utilized vehicle telemetry data. These data include brake pattern, Speed and turn at regular interval. These parameters communicated through CAN bus to signal processing unit, which track the drowsiness order and give the signal to the driver. [32]

III. CONCLUSION

In this paper, we understood that simulation was performed for day time only. In the night time video/images-based system won't work properly with

some ambient light. Accidents may happen in early morning or in midnight, in such situation light intensity can't remain same; it changes drastically due to incoming traffic. The proposed systems can be extended further for tackling the drowsiness detection in night time, where the illumination varies drastically and frequently as per as non-intrusive method is concerned. The researchers are advised to take this issue in account to work further in this field. As well as simulation carried out considering person with spectacles but person with dark sunglasses is not considered. We suggest to researcher to account this situation also.

Moreover, it has been found that drowsiness affects rule-based driving rather than skilled-based driving. We suggest that in new drowsiness detection system skilled-based driving should also be considered. We can make method developed in this project generalize to detect other driving faults. Higher level cognitive functions involved in complex tasks need to be addressed in future studies. It can be observed that person's motivation can help significantly in complex tasks. Research Engineers and policy makers should share the desk. For example, driver's behavior under fatigue condition should be considered while training process. Information obtained from these kinds of tests will be useful for implementation of individualized drowsiness detection system

REFERENCES

- [1] "Road accidents in India Issues and dimensions", Ministry of Road
- [2] Transport and Highway Government of India, 2016.
- [3] Ronald R. Knipling and Walter W. Wierwille, "Vehicle-Based Drowsy Driver Detection: Current Status and Future Prospects", IVHS America Fourth Annual Meeting, Atlanta.
- [4] Fengyi Songa, Xiaoyang Tana, Songcan Chena, Zhi-Hua Zhoub, "A Literature Survey on Robust and Efficient Eye Localization in Real-life Scenarios".
- [5] S. Puthusserypady and T. Ratnarajah, " H^∞ Adaptive Filters for Eye Blink Artifact Minimization From Electroencephalogram", IEEE signal processing letters, vol.12, No.12, December 2005.
- [6] Junfeng Gao, Pan Lin, Yong Yang and Pei Wang, "Automatic Removal of Eye-Blink Artifacts Based on ICA and Peak Detection algorithm", 2nd International Asia conference in Control, Automation and Robotics 2010.
- [7] A.Sammaiah, B.Narsimha, E.Suresh and M. sanjeeva Reddy, "On the Performance of Wavelet Transform Improving Eye Blink Detection for BCI"IEEE proceedings of ICETECT 2011.
- [8] Andreas /bulling, Jamie A.Ward, Hans Gallersen and Gerhard Troster, "Eye Movement Analysis for Activity Recognition Using Electrooculography", IEEE Transactions on Pattern analysis and Machine Intelligence, vol.33, No.4, April 2011.
- [9] Antoine Picot, Sylvie Charbonnier and Alice Caplier, "On-Line Detection of Drowsiness Using Brain and Visual Information", IEEE Transactions on System, Man and Cybernetics-Part A: Systems and Humans, Vol.42, No.3, May 2012.
- [10] Chin-Teng Lin, Che-Jui Chang, Bor-Shyh Lin, Shao-Hung Hung, Chih-Feng Chao and I-Jan Wang, "A Real-Time Wireless Brain-Computer Interface System for Drowsiness Detection", IEEE Transactions on Biomedical Circuits and Systems, Vol.4, No.4, August 2010.
- [11] Fu-Chang Lin, Li-Wei Ko, Chung-Hsiang Chuang, Tung-Ping Su and Chin-Teng Lin, "Generalized EEG-Based Drowsiness Prediction System by Using a Self-Organizing Neural Fuzzy System", IEEE Transactions on Circuits and Systems-I: Regular Papers, Vol.59, No.9, September 2012.
- [12] Kamil Staszek, Krzysztof Wincza and Slawomir Gruszcynski, "Driver's Drwsiness Monitoring System Utilizing Microwave Doppler Sensor", 19th International conference on Microwaves, Radar and wireless Communications, May 21-23, MIKON 2012.
- [13] Dongrui Wu, Vernon J. Lawhern, Stephen Gordon and Brent J. Lance, "Driver Drowsiness Estimation from EEG Signals Using Online Weighted Adaptation Regularization for Regression (OwARR)", IEEE 2016.
- [14] Seong-Pil Cheon and Suk-Ju Kang, "Sensor-based Driver Condition Recognition using Support VectorMachine for the Detection of Driver Drowsiness", IEEE Intelligent Vehicles Symposium (IV), 2017.
- [15] John J.Magee, Margrit Betke, James Gips, Matthew R. Scott and Benjamin N. waber, "A Human-Computer Interface Using Symmetry Between Eyes to detect Gaze Direction", IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol.38, No.6, November 2008.
- [16] Zhiwei Zhu and Qiang Ji , "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations", Computer Vision and Image Understanding 98, ELSEVIER 2004.
- [17] Andrea Cristiani, Marco Porta, Davide Gandini, Gian MarioBertolotti and Nikola Serbedzija, "Driver drowsiness identification by means of passive techniques for eye detection and tracking", Fourth IEEE International Conference on Self-Adaptive and Self-Organizing SystemWorkshop, 2010.
- [18] Giancarlo Iannizzotto and Francesco La Rosa, "Competitive Combination of Multiple Eye Detection and Tracking Techniques", IEEE Transaction on Industrial Electronics, Vol.58, No.8, August 2011.
- [19] Axel Panning, Ayoub Al-Hamadi and Bernd Michaelis, "A Color Based Approach for Eye Blink Detection in Image Sequence", IEEE international Conference on Signal and Image Processing Applications, 2011.
- [20] Wei Zhang, Bo Cheng and Yingzi Lin, "Driver Drowsiness Recognition Based on Computer Vision Technology", TSINGHUA SCIENCE and TECHNOLOGY, Vol.17, No.3, June 2012.
- [21] Yuriy Kurylyak, Francesco Lamona and Giovanni Mirabelli,"Detection of the Eye Blinks for Human's Fatigue Monitoring", IEEE 2012. [21] Kyril Mincov, Stefanos Zafeiriou and Maja Pantic, "A Comparison of Different Features For Automatic Eye Blinking Detection With an Application To Analysis Of Deceptive Behavior", Proceedings of the 5th ISCCSP, Italy, May 2012.
- [22] Christer Ahlstrom, Trent Victor, Claudia Wege and Erik Steinmetz, "Processing of Eye/Head-Tracking Data in Large-Scale Naturalistic Driving Data Sets", IEEE Transactions on Intelligent Transpotation

- Systems, vol.13, No.2, June 2012.
- [23] D.Selvathi and N.Dhivya, "Realization of VLSI architecture to detect driver drowsiness for road accident avoidance system", IC-GET, IEEE 2016. [24] Oraan Khunpisuth, Taweechai Chotchinasri and Narit Hnoohom, "Driver Drowsiness Detection using Eye-Closeness Detection", 12th International Conference on Signal-Image Technology & Internet-Based Systems 2016. [25] Belhassen AKROUT and Walid Mahdi, "Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness", IEEE IPAS'16: international image processing applications and systems conference2016.
 - [24] Bhargava Reddy, Ye-Hoon Kim, Sojung Yun, Chanwon Seo and Junik Jang, "Real-time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks", IEEE Conference on Computer Vision and Pattern Recognition Workshops,2017.
 - [25] Ji Hyun Yang, Zhi-Hong Mao, Louis Tijerina, Tom Pilutti, Joseph F.Coughlin and Eric Feron, "Detection of Driver Fatigue caused by Sleep Deprivation", IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol.39, No.4, July 2009.
 - [26] Paul Viola and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features". IEEE CVPR, 2001.
 - [27] Lienhart R., Kuranov A., and V. Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection". In Proceedings of the 25th DAGM Symposium on Pattern Recognition. Magdeburg, Germany, 2003.
 - [28] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm". In Machine Learning: Proceedings of the Thirteenth International Conference, Morgan Kauman, San Francisco,1996.
 - [30] Gobhinath.S, Aparna V. and Azhagunacchiya R, "An Automatic Driver Drowsiness Alert System By Using GSM", 11th International Conference on Intelligent Systems and Control (ISCO), 2017.
 - [31] Kashyap Vasudevan, Anjana P Das, Sandhya B and Subith P, "Driver Drowsiness Monitoring by Learning Vehicle Telemetry Data", IEEE 2017.

Development of Pathology Informatics Tool Using Disease Symptoms Network: Identification of Ailment and Remedies

Lutful Islam

Computer Department

*M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Aiman Khan

Computer Department

*M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Himanshree Nagvekar
Computer Department

*M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Md. Ataul Islam

*University of Pretoria,
Pretoria, South Africa*

Veena Misar

Computer Department

*M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Mrinmayee Rane
Computer Department

*M. H. Saboo Siddik College of Engineering,
Byculla, Mumbai, India*

Abstract- Physicians and medical practitioners make Diagnostic decisions and treatment recommendations based on history, medical imaging, lab results and other text or multimedia records of patients. Our system allows to have faster recognition of diseases based on symptoms rather than having to go to professional experts and wasting a lot of time and money. Computer assisted Information retrieval may help support quality decision making and to avoid human error. Also, the human decision making is often optimal, it is poor when there are huge amounts of data to be classified. Also, efficiency and accuracy of diseases will decrease when humans are put into stress and immense work. The current project has been undertaken to develop computational algorithm-based support system to help the common people for early diagnosis and prescription. In this regard, a symptoms-based support system will be developed with the help of hybrid intelligent algorithm which proves the accuracy of disease prediction. Successful completion of the project will be helpful to the wider community of the general practitioners as well as common people.

Keywords: Clinical diagnosis, clinical prediction model, artificial neural networks (ANNs), Medical data mining, decision tree and Bayesian classification.

I. INTRODUCTION

Accurate diagnosis of disease is challenging and time and cost intensive method. The clinical diagnosis is done mostly by Doctors expertise and experience [1-3]. Clinical prediction is a rapidly growing field that is concerned with applying computer science and information technology to Medical and health data [4-6]. With the aging population on the rise in developed

countries and the cost of Healthcare, governments and large healthcare organizations are becoming very interested in the potential of clinical diagnosis to save time, money and human lives [6, 7]. Clinical prediction model can be applied to several challenging clinical scenarios: screening high risk individual for asymptomatic disease, predicting future events such as disease or death and assisting medical decision making and health education. Despite the impact of clinical prediction models on practice, prediction modelling is a complex process requiring careful statistical analysis and sound clinical judgement. Although there is no definite consensus on the best methodology for model development and validation a few recommendations and checklists have been proposed. After model development and vigorous validation in relevant settings possibly with evaluation of utility- usability and fine tuning, good models can be ready for the use in practice. Hippocrates emphasized prognosis as a principal component of medicine. Nevertheless, current medical investigation mostly focus is on etiological and therapeutic research rather than prognostic methods such as the development of clinical prediction models. Clinical prediction models can inform patients and their physicians or other Health Care providers of the patients' probability of having a developing a certain disease and help them with associated decision-making [7].

II. LITERATURE SURVEY

Authors of the above paper have stated that preliminary studies have shown that clinical prediction of diseases is more accurate than physicians in identifying acute myocardial infarction in patients presenting to the emergency department with anterior chest pain. Many biological and Partho psychological processes manifest chaotic behavior chaotic processes are not best

analyzed by classical linear methods the hope has been that the networks ability to identify multidimensional relationships in clinical data not a parent to other forms of analysis would allow the network to improve Diagnostic accuracy nonlinear statistical methods have been tried before but they were Complex and computation early intensive and never became widely accepted the advent of artificial neural network and the much greater speed of today's computers have changed this clinical diagnosis became one of the first areas to which the artificial neural network was applied acute Myocardial Infarction was one of the earliest applications but the range is wide from appendicitis to the examination of biopsy specimens the first application of artificial making. Example: facilitating patient doctor communication based on more objective information. Applying a model to a real-world problem can help with detection of screening in undiagnosed high-risk subjects which improves the ability to prevent developing diseases with early intervention [7]. Furthermore, in some instances, certain models can predict the possibility of having future disease or provide a prognosis for disease [7]. Clinical prediction models are used to investigate the relationship between future or unknown outcomes and baseline health states among people with specific conditions. Neural network for chest pain appeared in 1989 [2].

The purpose of this study was to develop a method of classifying cancers to specific diagnostic categories based on their gene expression signatures using artificial neural networks (ANNs). We trained the ANNs using the small, round-blue-cell tumors (SRBCTs) as a model. These cancers belong to four distinct diagnostic categories and often present diagnostic dilemmas in clinical practice. The ANNs correctly classified all samples and identified the genes most relevant to the classification. Expression of several of these genes has been reported in SRBCTs, but most have not been associated with these cancers. To test the ability of the trained ANN models to recognize SRBCTs, we analyzed additional blinded samples that were not previously used for the training procedure, and correctly classified them in all cases. This study demonstrates the potential applications of these methods for tumor diagnosis and the identification of candidate targets for therapy [3].

Artificial neural networks have featured in a wide range of medical journals, often with promising results. This paper reports on a systematic review that was conducted to assess the benefit of artificial neural networks (ANNs) as decision making tools in the field of cancer. The number of clinical trials (CTs) and randomized controlled trials (RCTs) involving the use of ANNs in diagnosis and prognosis increased from 1 to 38 in the last decade. However, out of 396 studies involving the use of ANNs in cancer, only 27 were either CTs or RCTs. Out of these trials, 21 showed an increase in benefit to healthcare provision and 6 did not. None of these studies however showed a decrease in benefit. This paper reviews the clinical fields where neural network methods figure most prominently, the main algorithms featured, methodologies for model selection and the need for rigorous evaluation of results [4].

Artificial neural networks are algorithms that can be used to perform nonlinear statistical modeling and provide a new alternative to logistic regression, the most commonly used method for developing predictive models for dichotomous

outcomes in medicine. Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms. Disadvantages include its "black box" nature, greater computational burden, proneness to overfitting, and the empirical nature of model development.

The research paper Predictive data mining for medical diagnosis an overview of heart disease prediction [5] intends to provide a survey of current techniques that are in use in today's Medical Research particularly in heart disease prediction. A number of experiments have been conducted to compare the performance of predictive data mining technique on the same data set and the outcome reveals that the decision tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but as a protective method like KNN, neural network classification based on clustering are not performing well. Conclusion is that the accuracy of the decision tree and Bayesian classification for the improved after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

Medical data mining has great potential for exploring the hidden patterns in the data sets of medical domain. These patterns can be utilized for clinical diagnosis however the available raw medical data are widely distributed heterogeneous in nature and voluminous these data need to be collected in unorganized from this collected data can then be integrated to form a hospital information system data mining Technology proverb provides a user oriented approach to novel hidden patterns in the data.

III. RESEARCH FINDINGS: DATA MINING IN HEART DISEASE PREDICTION

Three different supervised machine learning algorithms, that is, Naïve bayes, KNN, decision list algorithm have been used for analyzing the data set. Tangram tool is used to classify the data and the data is evaluated using 10 fold cross validation and the results are compared. Tangara is a data mining suite built around the user interface algorithms. The main purpose of Tangara tool is to give researchers and students and easy to use data mining software and allowing to analyse the real or synthetic data. It is a powerful system that contains clustering supervised learning, meta supervised learning, feature selection, data visualisation, supervised learning assessment statistics, feature selection and construction algorithms.

Decision tree is a popular classifier which is simple and easy to implement it requires no domain knowledge or parameter setting and can handle high dimensional data the results obtained from decision trees are easy to read and interpret. Naïve Bayes is a statistical classifier which assumes no dependency between attributes. KNN is a method for classifying objects based on closest training data in the feature space.

The experiment is performed using training data set consisting of 3000 instances with 14 different attributes. Naive Bayes appears to be the most effective as it has the highest percentage of correct prediction for patients with heart disease(86.53%) followed by neural network and decision trees decision trees however appears to be the most effective for predicting patients with no heart disease(89%) compared to the other two models.

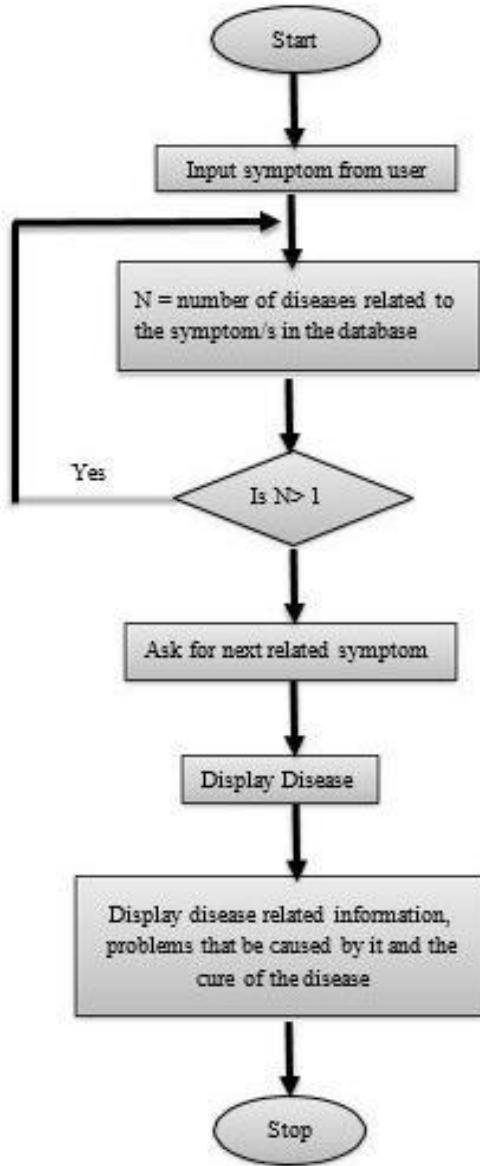


Fig. 1: Flowchart of proposed system

IV. PROPOSED SYSTEM

The proposed system for prediction of diseases (shown in Fig. 1) is basically a system where we will be able to predict the disease based on the symptoms the victim is facing. The system will also provide option to view information about doctors and hospitals near the user's location. Users will have to login into the system

to know the prediction of the disease. Login is not needed if the user wishes to only check for nearby hospitals and doctors. Changes to the database can only be done by admin.

Following are the features of the system:

- It will have the optimal time to predict the disease.
- It will generate related symptoms based on the symptom entered by the user.
- It will suggest the doctors based on the disease predicted.
- The system will be available all the time.
- Following are the advantages of the system over the previously discussed system:
- Will help patients have an idea of what disease they might be having.
- It is one of the first Automations in this field of health and medicine.
- It is more available then the previous system.

V. CONCLUSION

Medical diagnosis is regarded as an important yet complicated task that needs to be executed accurately and efficiently. The automation of the system would be extremely advantages. Regrettably all doctors do not possess expertise in every subspecialty and moreover there is a shortage of resource persons at certain places there for an automatic medical diagnosis system would probably be exceedingly beneficial. By bringing all of them together appropriate computer-based information and decision support systems can aid in achieving clinical test at a reduced cost efficient and accurate implementation of automated system needs a comparative study of various techniques available.

REFERENCES

- [1] W.G Baxt, MD, THE LANCET: Application of artificial neural networks to clinical medicine. Department of Emergency Medicine, University of Pennsylvania Medical Center, Philadelphia, PA 19104-4283, U.S.A., 1995.
- [2] Javed Khan, Jun S. Wei, Markus Ringnér, Lao H. Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R. Antonescu, Carsten Peterson& Paul S. Meltzer, NATURE MEDICINE: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine 7, 673–679, 2001.
- [3] John Moores Azzam F.G.Taktak, “NEURAL NETWORKS: The use of artificial neural networks in decision support in cancer: A systematic review” Royal Liverpool, 2001.
- [4] Jack V.Tu, “Epidemiolog advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes.”, Journal of Clinical, 1999.
- [5] Jyoti Soni,Uzma Ansari, “Predictive data mining for medical diagnosis an overview of heart disease Prediction”, International Journal Of Computer Applications, 2011.

- [6] International Journal of Computer Science Engineering and Information Technology Research(IJCSEITR) ISSN 2249-G831 Vol. 3 Issue 1, Mar 2013,239 – 248
- [7] How to Establish Clinical Prediction Models Yong-ho Lee, Heejung Bang, and Dae Jung Kim
- [8] Using methods from the data mining and machine learning literature for diseases classification and prediction a case examining classification of heart failure Original research article journal of clinical epidemiology Volume 66 Issue 4 April 2013 Pages 398 to 407 Authors: Peter C Austin, Jack V Tu, Jennifer E Ho, Daniel Levy, Douglas S Lee
- [9] Data mining approach for diagnosis of coronary artery disease. Original research article Computer methods and programs in Biomedicine Volume 111 issue 1 July 2013 pages 52 to 61 Authors: Roohalla Alizadehsani, Jafar Habibi, Muhammad Javad Hussaini, Hoda Mashayekhi, Reihane Boghrati, Asma Ghandeharioun, Behdad Bahadorian, Zahra Alizadeh Sani
- [10] Effective diagnosis of heart disease through neural network ensembles Research article expert systems with applications volume 36 issue 4 may 2009 pages 7675 to 7680 Authors: Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur.
- [11] Improving predictive models of glaucoma severity by incorporating quality indicators Original research article Artificial Intelligence in Medicine volume 60 issue 2 February 2014 pages 103 to 112 Authors: Lucia Sacchi, Allan Tucker, Steve Counsell, David Garway Heath, Stephen Swift.
- [12] Comparison of regression tree data mining methods for prediction of mortality in head injury original research article expert systems with applications volume 38 issue 12 November December 2011 pages 15534 to 15539. Authors: Necdet Sut, Osman Simsek.
- [13] Modern parameterization and explanation techniques and Diagnostic decision support system: a case study in Diagnostics of coronary artery disease original research article Artificial Intelligence in Medicine Volume 52 Issue 2 June 2011 pages 77 to 90 Authors: Matjaž Kukar, Igor Kononenko, Ciril Grošelj
- [14] An expert system for detection of breast cancer based on Association rules and neural network original research article expert systems with applications volume 36 issue 2 part 2 March 2009 pages 3465 to 3469
- [15] Classification of cardiotocogram data for anticipation of fetal risks using machine learning technique original research article applied soft computing volume 33 August 2015 pages 231 to 238 Authors: Hakan Sahin, Abdulhamit Subasi
- [16] Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules original research article journal of King Saud University computer and information Sciences volume 24 issue 1 January 2012 pages 24 to 40 Authors: P.K.Anooj
- [17] Discovering interobserver variability in the cytodiagnosis of breast cancer using Decision trees and Bayesian Networks. Original research article applied soft computing volume 9 issue 4 september 2009 pages 1339 to 1342. Authors: Nicandro Cruz-Ramírez, Héctor-Gabriel Acosta-Mesa, Humberto Carrillo-Calvet, Rocío-Erandi Barrientos Martínez.
- [18] Knowledge discovery in medicine: current issue and future trends review article expert systems with applications volume 41 issue 9 July 2014 pages 44 3424 4463. Authors: Nura Esfandiari, Mohammad Reza Babavalian, Amir-Masoud Eftekhari Moghadam, Vahid Kashani Tabar.
- [19] Medical image analysis with artificial neural networks original research article computerized medical imaging and graphics volume 34 issue 8 December 2010 pages 617 to 631 Authors: J. Jiang, P. Trundle, J. Ren
- [20] Automatic identification of cardiac health using modelling techniques: A comparative study original research article information Sciences volume 178 issue 23 1 December 2008 pages 4571 to 4582 Authors: U. Rajendra Acharya, Meena Sankaranarayanan, Jagadish Nayak, Chen Xiang, Toshiyo Tamura

Machine Learning Methods for Online News Popularity Prediction

Ravi Sitapara

I.T Department - TCET

Mumbai, Maharashtra

ravvisitapara007@gmail.com

Prashant Chaudhary

I.T Department - TCET

Mumbai, Maharashtra

pchaudhary31@gmail.com

Pradeep Kotian

I.T Department - TCET

Mumbai, Maharashtra

pradeepkotian1995@gmail.com

Shridhar Kamble

I.T Department - TCET

Mumbai, Maharashtra

shridhar.kamble@thakureducation.org

Abstract—Machine learning techniques are utilized to find the best model and set the features to predict the popularity of online news. Our data comes from Mashable, a well-known online news website. Using over 49000 sample articles from mashable.com different learning algorithms are implemented on the dataset, ranging from various regressions to SVM and Random Forest. Their performances are recorded and compared. Feature selection methods are used to improve performance and reduce features. Our work can help online news companies, editors, and bloggers to predict news popularity before publication. Accuracy & speed of different algorithms are being measured first. After running the algorithms on the dataset, it showed distinct results. The results were compared on the basis of their accuracy & speed to compute the articles.

Keywords— *Machine Learning, News Prediction, Random Forest, KNN, Likes, Comments, Shares, Views, Regression.*

I. INTRODUCTION

Online website has been the majority source for news to spread. Interesting news is going to be shared thousands of times through the Internet. With the expansion of the Internet, more and more people enjoy reading and sharing online news articles. Popularity prediction for news articles is a relatively novel problem and very few studies addressed this problem. However, a growing number of studies have been carried out on predicting the popularity of other types of online content. The number of shares, likes, comments under a news article indicates how popular the news is. In this project, we intend to find the best model and set of features to predict the popularity of online news, using machine learning classification techniques. Prediction will be based on the topic which will be the current trends in the online digital feed and depending upon this the news article will give a rank. As if the news article shares, likes and comments changes the ranking position will be changed. Depending upon the rich features the news article will predict the popularity and getting the top position in the news feed world.

II. OVERVIEW

The Online News Prediction of our project will be based machine learning algorithms & methods. Our model is based on random forest regression and learned on a set of features derived from the given time series data to capture the expected number of visits, rate of change. This project will be able to tackle the problem of ranking online news by using a two- phase procedure. First step consists understanding two underlying properties of news articles that are relevant to our ranking problem: articles' lifetime and distribution of popularity. Then, based on these observations, we recommend a more rigorous evaluation strategy adapted to the characteristics of news.

III. LITERATURE REVIEW

Using SVM, Naive Bayes, Trees, Bagging, Decision Tree & Regression it shows an accuracy of 84% in predicting the popularity range of a news article [1]. They have extracted 2.89 million tweets referring to 24 different movies released over a period of three months [3]. The Accuracy found for different algorithms were: i) Linear Regression -66% ii) Logistic Regression -66% iii) Random Forest (500 Trees)- 69%. Four classification methods were tested to predict three popularity classes (1 to 20 tweets, 20 to 100 tweets, more than 100) and results ranged from 77% to 84% accuracy, for Naive Bayes and Bagging, respectively and using over 39000 sample articles from Mashable.com [4]. K-Means clustering, Polynomial and Linear Regression was applied on 2510 movies released 1990 onwards to study and build a predictive model to get the expected revenue. They achieved accuracy of 36.9% [5]. They are able to achieve the least accuracy with neural network multilayer perceptron 79.07% and highest with simple logistic of 84.34% [6].

IV. PROBLEM DEFINITION

This project, intends to find the best model and set of features to predict the popularity of online news articles, using machine learning techniques. And which will be predicted based on number of shares, likes, comments under a news articles.

A. Phase 1(Information Gathering):

Extraction of more than 49,000 sample of article from mashable.com (which is UCI machine learning repository). It extracts 59 attributes (as numeric value) describing different aspects of each article, published in the last two years from Mashable website.

B. Phase 2(Analysis & Design):

Prediction models Random Forest, Linear Regression, Logistic Regression & Naïve Bayes Classification Methodology are done on articles for Preprocessing using Python. Prediction model's accuracy will be measured & compared among all the Classification Methodology. It would be correct if its value and the actual result have the same sign (both + or -).

C. Phase 3(Implementation & Testing):

Development of Graphical user interfaces (GUI) using Django Framework. Django will dynamically serve web pages without using database store information. Django framework will provide inputs like comment, share, likes, dislikes.

D. Phase 4(Deployment): -

Pre-Processed data will be integrated onto the Front-End using Django framework. A ranking table will be provided which will rank the articles according to the popularity. Graphical representation of factors which provides top rankings.

V. FEATURES

The News Prediction of our project will be based machine learning algorithms & methods. Our model is based on random forest regression and learned on a set of features derived from the given time series data to capture the expected number of visits, rate of change.

Prediction will be based on the topic which will be the current trends in the online digital feed and depending upon this the news article will give a rank. When the news article shares, likes and comments changes the ranking position will be changed.

Depending upon the rich features the news article will predict the popularity and getting the top position in the news feed world.

It tackles the problem of ranking online news by using a two-phase procedure. The first step consists in understanding two underlying properties of news articles that are relevant to our ranking problem: articles' lifetime and distribution of popularity.

Then, based on these observations, we recommend a more rigorous evaluation strategy adapted to the characteristics of news.

VI. METHODOLOGY

The Proposed methodology contains following steps: -

- Data Extraction
- Data Preprocessing
- Data Integration and Transformation
- Feature Selection
- Classification

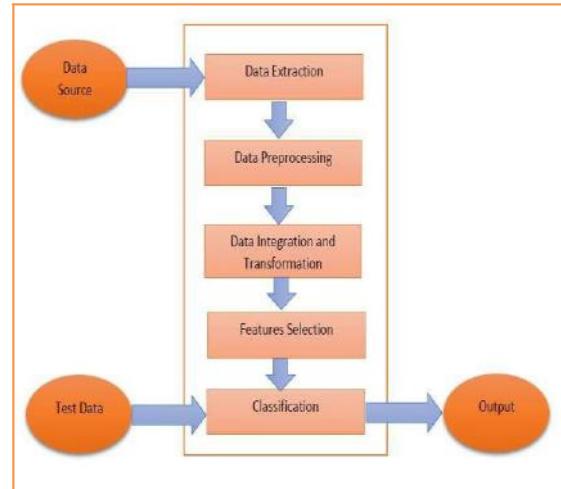


Fig. 1 Methodology

Random Forest:

In bagging (Bootstrap Aggregation), numerous replicates of the original dataset are created to reduce the variance in prediction. Random Forest uses multiple decision trees which are built on separate sets of examples drawn from the dataset. In each tree, we can use a subset of all the features we have. By using more decision trees and averaging the result, the variance of the model can be greatly lowered [9].

DECISION TREE:

Decision tree is commonly used in classification in machine learning, it is a tree like graph or model to make decisions and get consequent results. Decision tree implicitly perform variable screening or feature selection and the best feature of using it is easy to interpret and explain [9].

KNN:

K nearest neighbors is a non-parametric method used for classification and regression. The input normally consists of k closest training instances in the space. An object is classified to a class because it is most common among its k nearest neighbors. It is chosen because KNN don't need make any assumption on the underlying data distribution, it is also lazy algorithm and don't use training data to do any generalization [9].

NAÏVE BAYES:

Naive Bayes classifier is a probabilistic classifier based on Bayes' theorem. It depends on the assumption that some features are conditionally independent. It is super simple and can converge quicker than some other models, so it has a chance to achieve fast and easy process [9].

VII. FESEABILITY STUDY

- Technical Feasibility** – The application will operate on Web Browser. The system will be implemented using python, Django python framework, and web technology as required.
- Economic Feasibility**- In this project, we tried different methods to predict the popularity before its publication. By predicting the news it will helpful for news publication to put the relevant topic for good revenue. Therefore, the system is economically feasible.
- Legal Feasibility**-The project is developing under legal license of windows, python and other software's required. The data will not be used for some illegal purpose and will be used for making the news relevant to user as well as for publisher for making under certain constraint.
- Operational Feasibility** -An overview of the process portrayed that the proposed Machine Learning for Online News Popularity Prediction be beneficial for news publisher to monitor the growth and improvement in the future article content.
- Social Feasibility**-Easy to make allocation and accurately with this help online news publisher to predict the news popularity and also make user to read the current trending topic.

VIII. BLOCK DAIGRAM

This below block diagram shows the process of training the model using training data and implementing the model for getting the testing data.

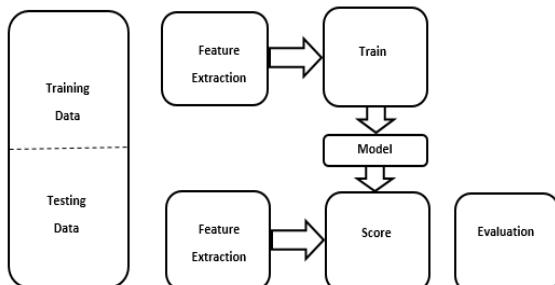


Fig. 2 Training and Testing data

By training the model it will be used for input testing the input data. This below diagram shows how the model in which the sample articles are being segregated into popular and unpopular articles using the model which is trained using the 49k sample.

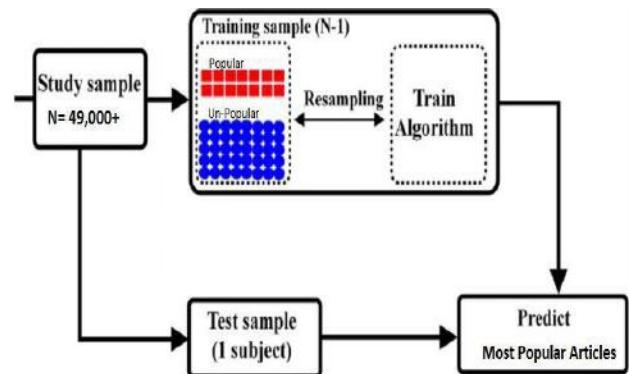


Fig. 3 Predict the popular and unpopular articles [14]

IX. METHODOLOGY

Below flow chart shows the collection of raw data from online repository and training the data using different machine learning algorithm like random forest, decision tree, KNN. Once the model is implemented then the testing data will be provided for predication the popularity and measuring the accuracy and time lapsed.

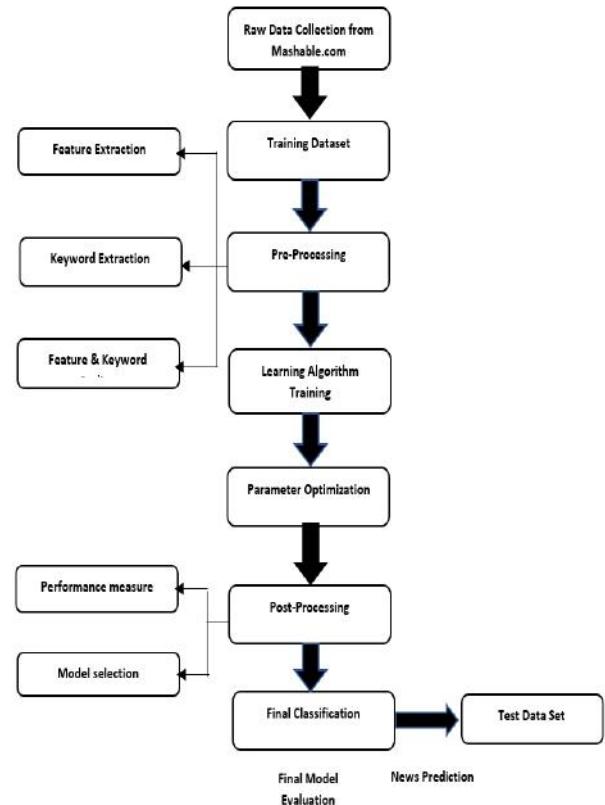


Fig. 4 Flow Chart

X. RESULTS

After implementing the following algorithms like decision tree, KNN, Random Forest, Naïve Bayes, it is being observed that random forest achieves better accuracy and Naive Bayes shows less time elapsed. Performance analysis carried out and its results are as follows:

TABLE 2: ACCURACY

Algorithm	Decision Tree	KNN	Random	Naïve Bayes
Accuracy (Original)	57.9%	57%	66.4%	61.2%
Accuracy (Cross Validation)	55.5%	54.9%	64.4%	60.8%

TABLE 2: SPEED

Algorithm	Decision Tree	KNN	Random	Naïve Bayes
Time elapsed (Original)	1.2s	3.23s	6.33s	0.07s
Time elapsed (Cross Validation)	8.44s	10.83s	37.71s	0.48s

From two tables above, it is observed Random Forest model achieves the highest accuracy (66.4%), but it consuming the most time (6.63s) among 4 models. Naive Bayes model achieves the least time elapsed (0.07s) and a properly medium accuracy. The other two models have bad accuracy. The K-fold cross validation has proved that the accuracy should be less than that gained from the original method. But considering its big time consuming, the difference can be ignored relatively under the purpose of comparing these 4 models.

XI. CONCLUSION

Current state-of-the-art was reviewed on News popularity prediction methods. We presented the different prediction methods, reported their performance, and suggested several applications that can benefit from these findings.

Various regression and classification models were trained for predicting the popularity of an online news article before its publication. We use shares to indicate popularity. The models were trained using Mashable data from UCI.

Overall, this project pursues a faster process; Naive Bayes is the best choice for this problem without sacrificing too much accuracy. Otherwise, if accuracy is considered first, obviously Random Forest is a better one.

REFERENCES

- [1] [1]. Alexandru Tatar, Panayotis Antoniadis, Marcelo Dias de Amorim, Serge Fdida. From "Popularity Prediction to Ranking Online News. Social Network Analysis and Mining", Springer, 2014, pp.4:174.
- [2] [2]. Babita M. Jangid, Chaitali K. Jadhav, Swati M. Dhokate,
- [3] Girish M. Jadhav, "Survey on Movies Popularity Prediction System Using Social Media Feature," IJIRCCE, VOL. 4,2016
- [4] Muhammad Hassan Latif, Hammad Afzal, "Prediction of Movies Popularity Using Machine Learning Techniques," IJCSNS, VOL.16 No.8.
- [5] He Ren, Quan Yang, "Predicting and Evaluating the Popularity of Online News," CS -229, 2016
- [6] Xuandong Lei, Xiaoti Hu Hongsheng Fang, "Is Your Story going to spread like a Virus? Machine Learning Methods for News Popularity Prediction," IEEE CS -299,2015
- [7] G.M.Bhandari Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WIAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
- [8] Fernandes, Kelwin, Pedro Vinagre, and Paulo Cortez. "A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News." Progress in Artificial Intelligence. Springer International Publishing, 2015. 535-546.
- [9] Statistical Learning and Data Sciences: Alexander Gammerman, Vladimir Vovk, Harris Papadopoulos (Springer).
- [10] Machine Learning in Python (Essential Techniques for Predictive Analysis): Michael Bowles. (Wiley)
- [11] The Definitive Guide to Web Development Done Right: drain Holovaty & Jacob Kaplan-Moss (Apress).
- [12] A Primer on Scientific Programming with Python: Hans Petter Langtangen.
- [13] UCI Machine Learning Repository - <http://archive.ics.uci.edu/ml/datasets/OnlineNewsPopularity> accessed on 17/7/2017
- [14] Github - <http://inseaddataanalytics.github.io/INSEADAnalytics/groupprocesses/January2017/MashableNews.html> accessed on 8/8/2017
- [15] https://www.researchgate.net/figure/288993135_Flow-diagram-showing-algorithm-training-and-testing-process-The-class-imbalance-problem accessed on 18/9/2017

Splunk for Big Data Analytics

Ratna Nayak

*Thakur College of Engineering and Technology,
Mumbai, India*

Shruti Mathur

*Thakur College of Engineering and Technology,
Mumbai, India*

Purvi Sankhe

*Thakur College of Engineering and Technology,
Mumbai, India*

Nishtha Mathur

*Thakur College of Engineering and Technology,
Mumbai, India*

Neha Patwari

*Thakur College of Engineering and Technology,
Mumbai, India*

Abstract- Most modern systems produce plentiful and diverse data. The lack of tools to efficiently store and cross-correlate diverse datasets makes it tedious to mine the data for analytic perceptions. In this paper, Splunk, a semi-structured database that can be used to index, search and analyze massive datasets. We share observations, lessons and case studies to demonstrate Splunk's power and elasticity for enabling understanding big data analytics.

Keywords- Big Data, IOT, Splunk, Cloud Computing,

I. INTRODUCTION

Every millisecond of every day applications, sensors, systems, web servers, networks, mainframe generate data. This machine data is the fastest growing, most complex areas of big data. It's one of the most valuable, containing definitive record of user transactions, customer behaviour, sensor activity, machine behaviour, security threats, fraudulent activity and more [1-3].

Making use of machine data is complex. It's generated by a gathering of disparate sources in unstructured, semi structured and structured formats. Because of unstructured data, makes it complex to fit into brittle pre-defined schemas. Organizations are challenged to ingest, process and analyze machine data using traditional data management methods or in a timely manner. Traditional business intelligence, data warehouse or analytics solutions are simply not engineered for this class of high-volume, dynamic and unstructured data.

Splunk software helps to unlock the hidden value of data. And with the ability to bring in insights from one to other tools, system get value from the full spectrum of data, not just a sub-set. Now can collect, index, search, analyze and

visualize all data in one place. Splunk provides a incorporated way to organize and extract real-time insights from massive amounts of machine data from virtually any source.



Fig. 1: Machine Data

II. RELATED WORK

Massive datasets are almost always timestamped, heterogeneous, and difficult to fit into traditional SQL database. Traditionally, datasets represent state information. They snapshot properties of real-world or virtual objects at some point in time. Some well-known examples of such stateful datasets include store inventories, airline reservations, and corporate personnel data. While these datasets often grow as large as billions of entries, they are bounded in size by real-world limits on the objects they represent. For example, a personnel database cannot grow larger than 7 billion entries as there are fewer than 7 billion people in the world. In addition, there is usually a set of properties (e.g. name, address, salary) for each entry in the database. While the values for each property vary from entry to entry and can evolve over time, the set of properties themselves is effectively fixed at the time the database is designed. Therefore, we refer to these as structured datasets. The general

techniques for analysing structured data are well established and the database community has provided a wealth of specialized tools for this purpose [8]. NOSQL-query based relational databases have served these structured datasets well.

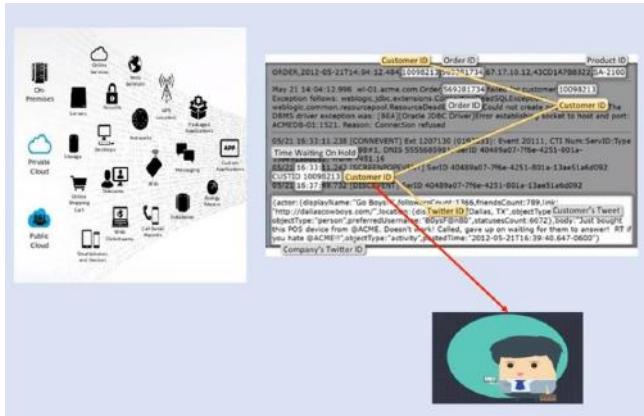


Fig 2: Challenges with Machine Data

Time is the best correlator for heterogeneous data sources. Large computing environments, ranging from corporations to governmental and academic institutions, generate tens to hundreds of different types of log data collected from thousands of components (data sources). Due to the unstructured nature of this data, there is usually no unique identifier (primary key) that can be used to join entries across different logs. For example, application server logs often contain Java stack traces that are directly or indirectly triggered by requests from a web-server. However, these logs contain no explicit identifiers to associate the stack trace with a specific web-server request. Each event is usually timestamped and thus the time at which each event occurs is often the only piece of information we can use to correlate various types of unstructured log data. Traditional relational database semantics are ineffective for temporal correlation because related events merely have timestamps that are “close” to one another and rarely have the exact same timestamp across various logs.

Relational database management solutions (RDBMS)

Converting from the semi-structured format of machine generated data to the dense and homogeneous structure demanded by traditional relational database tables is typically achieved by an Extract, Transform and Load (ETL) procedure [4, 5]. This procedure is often problematic for ad hoc analysis tasks in machine generated data sets. The designer of the ETL procedure must anticipate the gamut of questions to be asked of the data and correctly extract the rows and columns of the table. Such pre-planning is infeasible for two reasons. First, in many systems, there is no complete catalog of all

the messages that may be recorded. That is, never-seen-before messages may present themselves exactly when new problems occur. Second, the number of distinct messages in a large system can be prohibitively large and correctly extracting them all up front is often impractical [6-9].

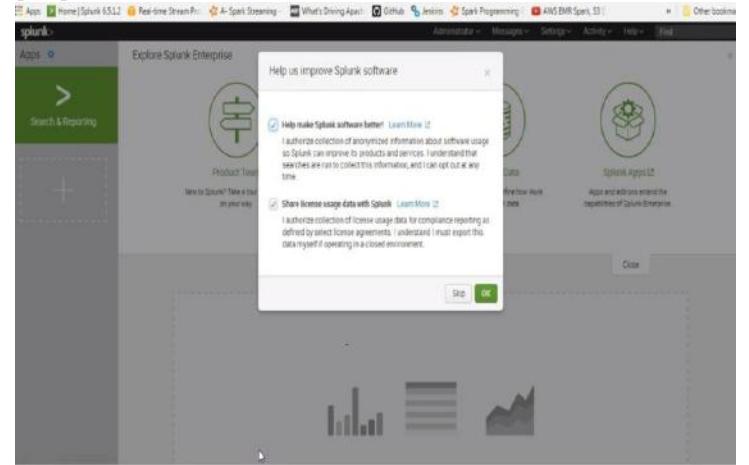


Fig. 3: Splunk Dashboard

Map Reduce implementations for big data analytics

The Apache Hive project [1] also seeks to put a general purpose, scalable interface on a semi-structured data warehouse. Hive provides ETL (Extract, Transform and Load), schematization and analysis of massive amounts of data stored in a Hadoop distributed file system (HDFS). Similar to Splunk, Hive uses the MapReduce paradigm, with Hadoop as the job management engine. Hive is designed primarily as a batch processing system and thus queries are not expected to be real-time for the operator. According to the Hive project, even the smallest jobs can take on the order of several minutes. Such lengthy job execution times are unattractive for IT and operations troubleshooting use cases where the time taken to solve a specific problem is very important [10-14]. Additionally, reducing the cycle time for iterating on query.

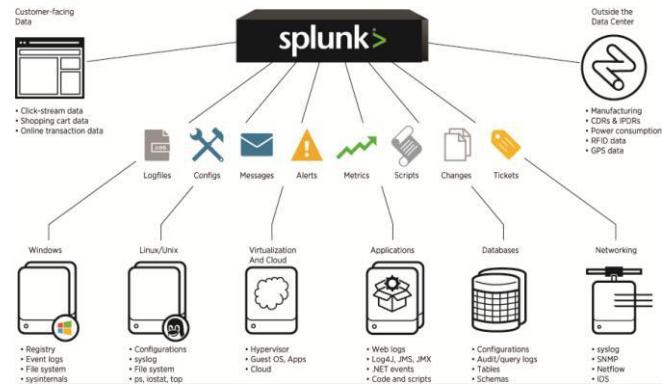


Fig 3: What Splunk Can Index

III. CONCLUSION

In this paper, presented Splunk, a big data analytics platform for massive datasets. Discussed why and how a semi-structured and unstructured database is the most appropriate mechanism for cross-correlating and mining heterogeneous datasets. Demonstrated the power and ease of using the Splunk search language for solving IT functioning problems.

Area knowledge is vital for successful data analysis. However, domain experts often lack familiarity with statistical/machine learning techniques to control them well. Splunk can abstract away the statistical expertise and thus enables domain expertise, rather than machine learning proficiency, to drive the analysis of any particular big dataset.

Many researchers have successfully leveraged machine learning techniques to gather system perceptions from data. Such techniques require a main fraction of time to be spent on collecting, parsing, and cleansing the data, and only a small fraction is attributed to applying the algorithm itself.

Splunk can be the foundation for newer infrastructures, but also has the latent to serve as the principal infrastructure to develop new big data mining techniques

REFERENCES

- [1] Apache Hive Project. <http://wiki.apache.org/hadoop/Hive>.
- [2] WELF. <http://www.m86security.com/kb/article.aspx?id=10899cNode=5R6Q0N>.
- [3] AMAZON.COM. Amazon Simple Storage Service (Amazon S3). <http://aws.amazon.com/s3/>.
- [4] ARCSIGHT. CEF. <http://www.arcgis.com/collateral/CEFstandards.pdf>.
- [5] DEAN, J., AND GHEMAWAT, S. MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM* 51, 1 (2008), 107–113.
- [6] FISHER, K., WALKER, D., AND ZHU, K. Q. LearnPADS: automatic tool generation from ad hoc data. In *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (New York, NY, USA, 2008), ACM, pp. 1299–1302.
- [8] FLINT, A., AND COOKSON, I. Data Warehousing on Oracle RAC Best Practices. White paper, Oracle Corporation, October 2008.
- [9] HELLERSTEIN, J. M., AND STONEBRAKER, M. Readings in Database Systems: Fourth Edition. The MIT Press, 2005.
- [10] IBM. CBE. <http://www-128.ibm.com/developerworks/webservices/library/ac-cbe1/>.
- [11] JACOB, B., LANYON-HOGG, R., NADGIR, D. K., AND YASSIN, A. F. A Practical Guide to the IBM Autonomic Computing Toolkit. IBM Redbooks, April 2004.
- [12] LONVICK, C. The BSD Syslog Protocol, 2001.
- [13] MITRE, SPLUNK AND LOGLOGIC. CEE. [http://cee.mitre.org/docs/Common Event Expression White Paper June 2008.pdf](http://cee.mitre.org/docs/Common%20Event%20Expression%20White%20Paper%20June%202008.pdf).
- [14] SORKIN, S. Large-Scale, Unstructured Data Retrieval and Analysis Using Splunk. Technical paper, Splunk Inc., 2009.

To Understand Drug Usage by Mining Social Media Data

Shwetha Yadav

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India.
shwethabaliyadav@gmail.com*

Abstract-This review paper in view of the research completed in the territory of information mining depends for overseeing mass measure of information with mining in web-based social networking on utilizing composite applications for performing more modern investigation utilizing cloud platform. Upgrade of online networking may address this need. The goal of this paper is to present such sort of hardware which utilized as a part of interpersonal organization to portrayed medication manhandle. This paper laid out an organized way to deal with break down web-based social networking so as to catch developing patterns in mediate mishandle by applying capable techniques like distributed computing and Map Reduce show. This paper portrays how to bring critical information for examination from interpersonal organization as Twitter, Facebook, and Instagram. At that point enormous information procedures to separate helpful substance for examination are talked about.

Keywords- *social media, data mining; Big data, K-NN algorithm.*

I. INTRODUCTION

Interpersonal organization (social media) is one to removing the data from the web. These days it is utilized for separating the information of patient's to know the comprehension of patient indications. Web-based social networking, order from singular informing to live for as, is giving inconceivable chances to patient to speak their encounters with medication and gadgets. Web-based social networking permits message commitment, gathering data and circulation in the medicinal services space. Medicinal services is one which contains the data of patients with their authorization. It gives a viable person to person communication condition. The best possible method for mining data and float from the learning is cloud. Utilizing system based examination technique it demonstrate the online networking, for example, Facebook, Twitter, WebMD. Nowadays the logical test regularly requires mass measure of calculation amid reproduction and information preparing. Execution of super PC is expanding rapidly. It permits tackling logical issue via programmed computational through gathering or cluster list which is developed by set of sensors. These days electronic instrument is developing in late situation..

Kamal Shah

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India.
kamal.shah@thakureducation.org*

II. DATA MINING TECHNIQUES

The information extracted permits forecasting the behaviour and future behaviour. This enables the entrepreneurs to take positive information drive choices. Information mining is implemented in various space like FMCG, economy, restorative, training framework and so on. Learning is gotten from the already truth by applying design acknowledgment, measurable, numerical strategies those outcomes in skill type of actualities, patterns, affiliation, examples, abnormalities and special cases. There are a few zones where information mining is connected.

Data Pre-processing: Data pre-processing makes the natural data ready for mining process.

Data Mining: Mining is the way of separating some specific patterns which is important from a vast amount of data.

Pattern Evaluation: This procedure assesses the pattern that is produced by the information mining. The examples are assessed by engagingness measure given by the client or framework.

Knowledge presentation: It utilizes representation procedures that picture the interesting patterns and help the client to see and translate the resultant patterns.

III. DATA MANAGEMENT

Administration of information from social chain which gives bound together way to deal with understanding logical assignments. Since friendly chain contains plentiful of information Author utilized Big Data ideal models to mine and examine it. Initially, information from friendly chain is mined utilizing the crawler or motor which spares it into Hadoop group/bunch. Also, enormous volume of mined information is sifted and amassed to get nearly minor datasets of data that is appropriate to the fathoming errand. At long last gathered information is utilized as a contribution for complex applications which perform last and refined information investigation. To oversee computational method for the unpredictable application Author utilized AaaS (Application as a Service) display, which is executed in this condition for conveying registering based cloud stage.

SOCIAL MEDIA DATA

The information from the various sites, the data originating from the clients, or the people posting something on the social media. All this data which is called as raw data is called as social media

data. Social media data can also be said as raw data because it consists of large amount of data. The raw data also consists of social media activities like shares, likes, comments, mentions, conversations, impressions and many more.

IV.WHY THERE IS A NEED

In the event that online networking information are your raw materials, at that point the web-based social networking investigation are your recipe. With these investigation, you can answer some critical inquiries concerning the achievement of your online networking exercises, for example,

1. Which type of contents attract the customers to click, share and convert?
2. Which organization contribute the most ?
3. Are the people taking interest ?

By answering these inquiries, you're basically obtaining bits of knowledge or 'web-based social networking insight' to illuminate your future choices and activities. At this stage, the raw materials have been cooked into a prepared to-eat dinner. Also, as with any feast that you cook; on the off chance that you need it to taste great, you have to utilize amazing fixings, not straightforward ones.

CLASSIFICATION TECHNIQUES:

There are many classification techniques which are used to classify the instances. It is used to predict that in which class they belong. This classification technique is the concept of machine learning and now a days it is widely used for analysis purpose. Many organization use this technique for analytic purpose like where they belong in the industry, what is their image in social media by mining the reviews. Some are k-nearest neighbor, K-NN classification, SVM and many more.

V. RELATED WORK

From Noemie Elhadad, et al[5], agreeable chains are a noteworthy hotspot for customer created criticism on about all items and administrations. Clients much of the time accept on social bind to reveal now and then genuine episodes as opposed to going to social correspondence channels. This vital, significant, customer made certainties, if extricated genuinely and powerfully from the social chain, can possibly have the positive effect on basic applications identified with social wellbeing and security, and past. Shockingly, the creation of data from social chain where the yield of the extraction procedure is utilized to take solid activities in the genuine world are not very much upheld by existing innovation. Customary data creation approaches don't function admirably finished the exceedingly casual and ungrammatical sentence structure in social chain. They don't deal with the generation and collection of uncommon substance. In our progressing aggregate undertaking between Columbia University and the New York City Department of Health and Mental Hygiene (DOHMH), this paper intend to address these distinction in research and innovation for one essential general wellbeing.

From Erwan Le Martelot et al [6], today wherever organize is accessible. The people group exposure got an expanding consideration as an approach to uncover the arrangement of

systems and associated inside than externally. However the vast majority of the powerful techniques accessible don't think about the conceivable levels of association, or scales, a system may incorporate and are in this manner restricted. In this paper Author said in regards to perfect with worldwide and neighborhood criteria that empowers quick multi-scale group finding. The strategy is to clarify with two calculations, one for each sort of measure, and executed with 6 known standard. Disclosure people group at different level is a computationally extravagant assignment. Consequently, this activity puts a solid consideration on the lessening of computational unpredictability. A few heuristics are initiated for accelerate reason. Trial display the competency and correct of our way regarding singular calculation and model by testing them against substantial outcomes in multi-scale arrange. This work likewise offers an appraisal amongst criteria and between the worldwide and nearby methodologies.

From Hari Kumar and Dr. P. Uma Maheshwari [7] Big information is the term that portrayed by its expanding volume, speed, assortment and veracity. Every one of these qualities make handling on this huge information an unpredictable undertaking. Along these lines, for preparing such information Author need to do it any other way like Map Reduce Framework. At the point when an association trades information for mining helpful data from this Big Data then protection of the information turns into an imperative issue in the earlier years, a few security safeguarding models have been given. Anonymizing the dataset should be possible on numerous operations like speculation, concealment and specialization. These calculations are for the most part reasonable for dataset that does not have the attributes of the Big Data. To propagate the protection of dataset a calculation was proposed recently. An creator speaks to how the development of enormous Data qualities, Map Reduce structure for security safeguarding in eventual fate of our exploration.

From E. Srimathi, K. A. Apoorva [8], as of late numerous web administrations expect customers to share their private electronic wellbeing records for inquire about examination or information mining, which prompts security issues. The size of information in cloud foundation ascends as far as nature of Big Data, in this manner making it a contention for conventional programming instruments to process such mass information inside a tolerable slipped by time. As a consequence, it is a contention for current anonymization strategies to save protection on classified extensible informational collections because of their deficiency of scalability. An Author speaks to an extensible two-stage way to deal with Anonymizing versatile datasets utilizing dynamic Map Reduce structure and LKC security display.

VI. PROBLEM IDENTIFICATION

Author have learned several things from this study (work). First, define the programming paradigm makes it simple to correlate and distribute computations for fault-tolerant.

Second, network channel is a limited resource. A number of surge in this model targeted at reducing the amount of data sent among number of client's node. Third, redundant execution may use to reduce the impact of slow machines, and to handle machine failures and data loss.]

VII. PROPOSED METHODOLOGY

Step 1: Input / Load data set

Step 2: Apply supervised learning for feature extraction
Received Extracted data as output

Step 3: Generate Training and Testing data set (By applying techniques)

Step 4: Apply Machine learning algorithm to training dataset (MLR)

Step 5: Build the Classifier / model using the “training” Dataset

Step 6: Apply Classifier on testing data se

Step 7: Perform / Obtain Prediction (classification) of the testing set.

Step 8: Utilize the “test” set predictions to calculate all the performance metrics (Measure Accuracy and other parameters

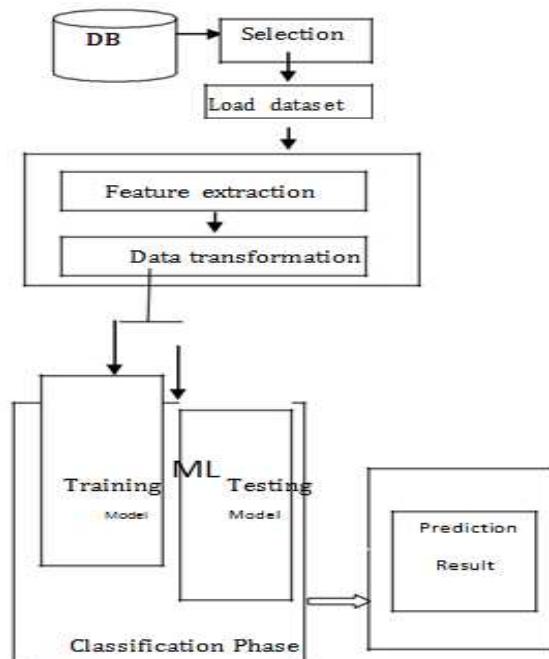


Fig. 1 Proposed architecture

The figure 1 above is the proposed srchitecture which is used to mine the social media data. The diagram consist of database where the social media data is stored and from that the particular data is selected and extraction of the data is done. Data extraction is nothing but extacting a particular data of feature is selected. After this step data transformation is done, during this phase transforming a particular data is done. And thiose data is the send to process by the machine learning model where those data is tested in which class they belong. There is a training set of instances present in the machine learning model which is used

to predict the new test instances where or which class they belong.

The algorithm is given in this survey paper, to quickly understand the flow of the process. As it is listed above in steps the first step is to collect the raw data from the social media as the input. The next step after the input is to apply supervised learning algorithm for feature extraction. And at this particular step the output is also generated as the extracted data. After this step the training set is generated and even the testing data is created by applying various algorithm. Next step is to apply machine learning algorithm to the training dataset to train the model for getting the correct predicted output. Aftyer this step are main aim is to build the classifier model to classify the test data. Its upto us that which classifier algorithm we can use. There are many classifier algorithm, they are listed above in this paper. After building the classifier model the next step is to apply that model for classifying the training data sets. The last step is we get the output as the predicted class of that particular training data set.

K-NN classification

This algorithm is simple to implement which stores all instances and the classify the new instances based on their similarities. It uses Euclidean distance formula to calculate the distance between the instances. It stores all the instances hence it does not required any learning.

1. Calculate “ $d(x, x_i)$ ” $i=1, 2, \dots, n$;
where d denotes the Euclidean distance
between the points.
2. Arrange the calculated n
Euclidean distances in non-decreasing order.
3. Let k be a +ve integer,
take the first k distances from this sorted list.
4. Find those k -points corresponding
to these k -distances.
5. Let k_i denotes the number of points belonging
to the i^{th} class among k points i.e. $k \geq 0$
6. If $k_i > k_j \forall i \neq j$ then put x in class i .

VIII.CONCLUSION

This paper exhibited our approach for mining and overseeing information from social chain which relies on mix of mass measure of information from interpersonal organizations which depends on blend of huge information and foundation ideal models. In this paper the data is mined for understanding the drug usage from social media. Since the data we are extracting is

mined from the social media and hence the data is in the form of raw data. We are proposing the approach to classify the extracted data by using the K-NN classification algorithm.

REFERENCES

- [1] The Fourth Paradigm, in: T. Hey, S. Tansley, K. Tolle (Eds.), "Data-Intensive Scientific Discovery," Microsoft, 2009.
- [2] S.C. Glotzer, et, al, WTEC Panel Reprt "On international assessment of research and development in simulation based engineering and science", World Technology Evaluation Centre, Inc,2009.
- [3] A.V. Boukhanovsky, S.V. Kovalchuk, S.V. Maryin, "Intelligent software platform for complex system computer simulation: conception, architecture, and implementation", Izvestiya, VuZov, Priborostroenie10 (2009) , 5-24, in Russian.
- [4] J.R. Rice, R.F. Boisvert, "From scientific software to problem-solving environments", IEEE, Computational Science and Engineering 3 (3) (1996) 44-53.
- [5] Noemie Elhadad, et al "Information extraction from social media for public health".
- [6] Erwan Le Martelot, "Fast multi scale detection of relevant communities".
- [7] Hari Kumar.R M.E (CSE), Dr. P. Uma Maheshwari , Ph.d, "Literature survey on big data in cloud," International Journal of Technical Research and Applications e-ISSN: 2320-8163.
- [8] E.Srimathi, K.A. Apoorva "Preserving identity privacy of healthcare records in big data publishing using dynamic MR", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, Issue 4, 2015.
- [9] [9] V. Borkar. M.J. Carey, C. Li., "Inside big data management: ogres, onions, or parfaits?", Prof. 15th Int. Conf. Extending Database Technol., 3-14, 2012.
- [10] X. Sun , B. Gao, Y. Zhang, W. An, H. Cao, C. Guo, et al, "Towards delivering analytical solution in cloud: business models and technical challenges", E-Bus. Eng. (ICEBE), 2011 IEEE 8th Int. Conf.
- [11] D.J. Abadi, "Data management in cloud: limitation and opportunities", IEEE Data Eng. Bull, 32: -12, 2009.
- [12] Matthew Kerland et al, "A review of data mining using big data in health informatics".
- [13] Swathi Yadav, Shwetha Yadav,"Text Mining of Voot Application Reviews on Google Play Store ",IRJET, vol. .5, no.1,pp.56-72,2018.

Project Title Allocation and Monitoring System

Prince Jain

*Thakur College of Engineering and Technology
Mumbai, India*

Jagdish Choudhary

*Thakur College of Engineering and Technology
Mumbai, India*

Pravinsingh Jodha

*Thakur College of Engineering and Technology
Mumbai, India*

Neha Kapadia

*Thakur College of Engineering and Technology
Mumbai, India*

Abstract- As part of the senior level of many undergraduate degree courses, students are required to undertake some form of project work. For that, project needs to be allocated to the students. The idea of our project is about making the allocation process web-based and assign the projects automatically. Based on the preference and criteria of the students, the projects will be automatically allocated. This will reduce the tedious time consuming paper work process and allow the faculty to complete the task within less time. This system will be basically developed for educational organizations where the projects need to be assigned to the final year students. Students, Faculty and project details will be stored in the database so that whenever required it can be instantly available. This process is mainly carried out for final year students. It also aims at increasing communication between faculty and students and making the process more efficient. Faculty can track progress of each group under him/her. Students can also enter the status and problems of their project if any ,this status and problem will be updated on the faculty's homepage.

Keywords: Automatic Allocation System, Monitoring System, Project Allocation

I. INTRODUCTION

Project Title Allocation and Monitoring system can be used by organizations to make the entire process online. Project have five roles in this system mainly, an administrator, a professor, a student, a lab assistant and a viewer. An administrator logs into system and can register a professor who belongs to that institution. Administrator is also approving for making changes to the database. Students register in this system and get user-id and group-id. Students have to provide details like their marks and Attendance and first preference first serve can be

added according to institute. And not only aggregate will be the factor but also attendance will be a key factor (Aggregation + Attendance = preference criteria).

preferences about project titles. According to this information project and mentor will be assigned to them. After allocation, students are responsible for fulfilling the milestones set by their mentors. Faculty members, after login, can view the groups under them and can track their progress by checking if deadlines are met proper attendance is met and schedule reports are submitted. Lab assistants responsible for attendance of students. He will also be register by administrator. A viewer need not log in. He simply can view the gallery aspect of project.

II. BACKGROUND

In universities project title allocation to students is done manually. The faculty allocates the project title and assign guide to that group. All work from allocation to monitoring is done manually. Each year institutes allocates no of project title to the students for their major project. Till date there are many universities and colleges which manually allocate and trace all allocates a student group id, project details and a mentor who monitors student's progress. This entire job manually becomes very difficult. It also includes setting of milestones, submission of reports and deadlines.

A mentor may have to monitor more than one group's progress. This entire job manually becomes tedious and time consuming. The Project Allocation System is a web application that allows the admin to automatically allocate the project. A message will be sent to the project department with group id and project id.(Reduced by 25%) to reduce the 100% use internet(ISSN2168-1323).

Till now there where many colleges and universities which manually allocate projects assigned to the students. This process is basically carried out for the final year students. It means that the college allocate the student group id, project details and a mentor who guides the project.

III. PROBLEM DEFINITION

Till date there are many universities and colleges which manually allocate and trace all the projects assigned to students. Previously according to ref3 the project title allocation to the members was done by board members, this system aims at providing a framework, for automatically allocation of project title to groups based on their preferences. The Project Allocation System is a web application according to ref1, that allows the input data to take place securely over the Internet .

The 24/7 internet connectivity is reduced by sending a message to the project department with group id and project id. According to ref4 the project title allocation were based on preferences and aggregate, this project aims of adding more choices by providing the project title according to their preferences, aggregate & attendance. Project monitoring manually becomes very difficult. It also includes setting of milestones, submission of reports and deadlines. A mentor may have to monitor more than one group's progress. This entire job manually becomes tedious and time consuming. This project aims at making this job web-based so that it saves time and efforts both on student's side as well as on faculty's side and mapping of attendance can be done.

Phase 1-

- Analysis- Detailed Study of project topic with the help of Literature survey from references and journals, finding the key gap from the previous work done on this topic. Defining the goals and scope of project, creating a workflow to complete the project.
- Design- Designing the methodology with the help of flowchart and algorithm. Developing the DFD, USE CASE diagram for the mentioned system.
- Implementation- Front end for Student and Teacher side will be scripted where student can add his/her project topics & and can see his/her updates regarding the project.Teacher side scripting will be done where he/she monitor the student project development and growth.Admin side scripting will be done where adding of project topics as per message received by the student.Server side scripting will be done where all the data will be stored.

Phase 2-

- Testing- Validation should be done at both allocation and monitoring part to prevent input of false or null data. Student and teacher login should be validated to prevent unauthorized access. There should be verification from the server side where the entered CGPA is accurate or not. User machine should have internet access and a browser to access the website.
- Deployment-After complete implementation and testing part the product will be deployed where in the project title allocation and monitoring part can be done.

IV. METHODOLOGY

The below figure shows the overall working of allocation and monitoring system ,from allocation to monitoring and how the system will be able to perform the given task and deliver the desired output as required by the user. The process starts from authentication till the allocation of the project title and monitoring the allocated project by the project guide & coordinator.

Firstly the student database will be created which will consist of their name and groupid and some useful data. Allocation of the title will be done according to the algorithm FCFS.Based on this allocation the monitoring system will be created which will handle the allocated project to the students and give weekly updates about their work.Attendance will be given based on their weekly updates.Monitoring system will consist of student and guide interaction with the help of chatbox. Fig. 1 shows the general block diagram.

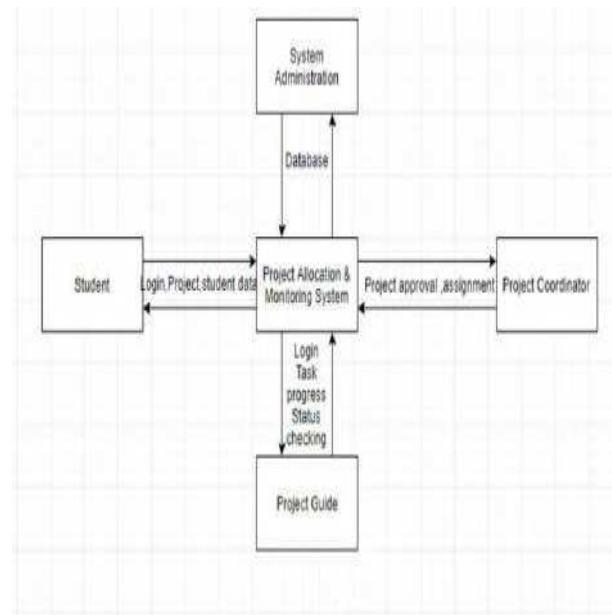


Fig. 1 Block diagram

V. RESULTS & CONCLUSION

Automatic Project Allocation System will be allocating the title of the project automatically according to the preferences given by the groups on the basis of parameters decided. Project Allocation system can be used by organizations to make the entire process online.. Efficient allocation of the project will be carried out using this system. All these conclusion were drawn according to the results derived from this process. The developed system solves the problem of unity, and lack of communication.

The system also breaches the gap between the students and the project guide by offering a platform whereby the students can monitor the progress of the system. It is recommended that this web based project allocation & management systems should be deployed wherever the need to manage projects efficiently arises. They are convenient to use, save time and resources, and reduce both stationary and labor costs. For further studies, this system can be modified by other researchers to be adapted in their various areas of study.

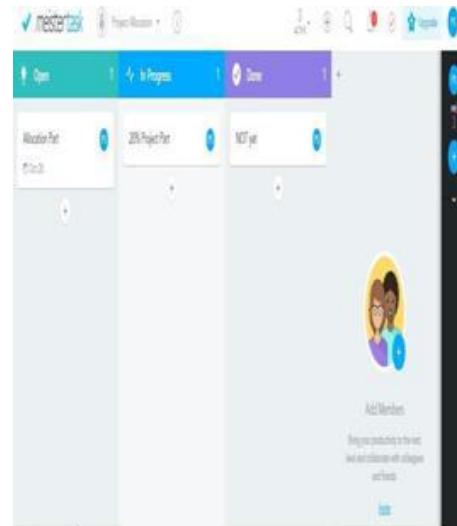


Fig. 2: Homepage

Allocation	
Group Id	<input type="text" value="1"/>
Project Id	<input type="text" value="4"/>
Aggregate	<input type="text" value="55"/>
<input type="button" value="Add"/>	

Fig. 3: Title Allocation

Fig. 2 shows the homepage appearance, Fig. 3 shows how titles are allocated. Aggregation of marks is shown in Fig. 4.



groupid	projid	aggregate
5	3	60
7	5	80
9	4	90
1	5	50
2	3	40
10	3	95
8	4	55

Fig. 4: Aggregation of marks

VI. CONCLUSION

This project is aimed at developing a web-based system which manages the activity of “Student project allocation and status tracking”. This system will manage the database and maintain a list of all student groups that have registered on this site. Broadly it can be effectively deployed in following real world instances: Helpful for universities and colleges to maintain and allocate students projects. Can also be used in educational organizations apart from universities.

REFERENCES

- [1] G Razis, “On-line Project Allocation System”, Proceedings of IEEE International Conference on management, 2012.
- [2] O. Nicholas, “eProject Tracker”, Washington, United States of America, 2011.
- [3] G. Pradeepa, S. Senevirantna, “Web based project collaboration, monitoring and management system”, 2014.
- [4] N. Kapadia, A. Kamat, P. Chaturvedi, J. Jain, “Project Allocation and Status Tracking System”, International Journal of Applied Information System(IJAIS)”, 2014.
- [5] N. Kapadia, R. Bubna, M. Jadhav, A. Joshi ,”Project Allocation and monitoring system”, International Conference on Cloud Computing and Virtualization,283-28, 2013.

Santander Customer Satisfaction

Rajat Nareshkumar Khemka
Department of Computer Science
California State University- Los Angeles
Los-Angeles, USA
rkhemka@calstatela.edu

Abstract—Benchmarking to be the leader in competitive market requires focus on total customer satisfaction. The goal of this competition was to predict if customers are satisfied or dissatisfied with their banking experience. Our solution is based on applying different classifier training model and then predicting the results for testing data. Random forest with cross-validation resulted in most accurate prediction for testing data

Keywords: accuracy, boosting, dataset, goal, test, train

I. INTRODUCTION

Customer satisfaction is the most important aspect for organizational success. From frontline support teams to C-suites, customer satisfaction is a key measure of success. Unhappy customers don't stick around. The organizations should have a constant check over the degree to which customers are satisfied. What's more, unhappy customers rarely voice their dissatisfaction before leaving. Santander is asking Kagglers to help them identify dissatisfied customers early in their relationship. Doing so would allow Santander to take proactive steps to improve a customer's happiness before it's too late. In this competition, you'll work with hundreds of anonymized features to predict if a customer is satisfied or dissatisfied with their banking experience [1].

II. PROJECT DESCRIPTION

A. Santander bank

Santander Bank is one of the country's top retail banks by deposits and a wholly owned subsidiary of one of the most respected banks in the world. Santander Bank was founded in 1902 as Sovereign Bank, a savings and loan in Wyomissing, Pennsylvania Banco Santander, their parent company. Santander Group serves more than 100 million customers in the United Kingdom, Latin America, and Europe. They are in the Northeast, they are a team of 9,800 individuals all committed to a single mission: to help their customers make progress towards their goals. They aim to make customer's banking hassle-free by providing simple ways for customer to spend, save and manage your money. Santander Bank has \$54.7 Billions of deposits, 650 Branches, 2000 and more ATMs and 9800 employees [2].

B. Data

The data for this project was provided with an anonymized dataset containing a large number of numeric variables. The actual dataset has 370 features and 76020 rows. The target column consisted information of satisfied and unsatisfied customers which is represented as 1 for unsatisfied customers and 0 for satisfied customers.

C. Goal

The goal of this competition was to predict if customers are satisfied or dissatisfied with their banking experience.

III. CLASSIFIERS

In machine learning and statistics, classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set. A training set is a set of observation that a machine learning algorithm uses to learn from data in case of supervised learning. Testing set is used to evaluate the accuracy of model. The unsupervised learning algorithm applies clustering of data based on some similarity present in data.

A. Classifiers Used for this project are:

1. Decision Tree is a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Decision Tree Classifier is capable of both binary (where the labels are [-1, 1]) classification and multiclass (where the labels are [0...K-1]) classification. Decision Tree Classifier is a class capable of performing multi-class classification on a dataset. Decision trees can also be applied to regression problems using the DecisionTreeRegressor class [4].
2. Gradient Boosting builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage $n_{classes_regression}$ trees are fit on the negative gradient of the binomial or multinomial deviance loss function. Binary classification is a special case where only a single regression tree is induced. Gradient Boosting Classifier supports both binary and multi-class classification. Gradient boosting can handle mixed type of training data set, can provide more predictive power for given training dataset and is also robust to outliers in output space which is achieved by robust loss function. Gradient Boosting Regression trees are less scalable due to sequential nature of boosting it can hardly be parallelized. Gradient Boosting attempts to solve this minimization problem numerically via steepest descent: The steepest descent direction is the negative gradient of the loss function evaluated at the current model F_{m-1} which can be calculated for any differentiable loss function [3]:

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_F L(y_i, F_{m-1}(x_i)) \quad (1)$$

3. A random forest is a meta-estimator that fits a number of classifiable decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and

- control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True. In random forests each tree in the ensemble is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. As a result of this randomness, the bias of the forest usually slightly increases (with respect to the bias of a single non-random tree) but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model [6].
4. Cross-Validation provides a good estimator to rate the generalization power of a model. This is a good solution to find which model is adapted to a certain data set. It can then be used to select the appropriate parameters to get the appropriate complexity. The first intuition to improve a model is to try to minimize the empirical error, i.e. the error made on the data used to train the algorithm. A current pitfall when we try to do this is to stick too closely to the data, we learn irrelevant details of the training set, which leads to a wrong generalization. This problem happens when we have too little data and a too precise model. Cross-Validation provides a good estimator to rate the generalization power of a model. This is a good solution to find which model is adapted to a certain data set. It can then be used to select the appropriate parameters to get the appropriate complexity. This way, we can avoid over fitting [7].
 5. Multi-layer Perception (MLP) is a supervised learning algorithm that learns a function $f(\cdot) : R^m \rightarrow R^o$ by training on a dataset, where m is the number of dimensions for input and o is the number of dimensions for output. Given a set of features $X = x_1, x_2, \dots, x_m$ and a target y , it can learn a non-linear function approximate for either classification or regression. It is different from logistic regression, in that between the input and the output layer; there can be one or more non-linear layers, called hidden layers. Multi-layer Perceptron has capability to learn from non-linear models and also to learn non-linear models in real-time using `partial_fit`. MLP Classifier implements a multi-layer perceptron (MLP) algorithm that trains using backpropagation. More precisely, it trains using some form of gradient descent and the gradients are calculated using backpropagation. MLP Classifier supports multi-class classification by applying softmax as the output function. Further, the model supports multi-label classification in which a sample can belong to more than one class. For each class, the raw output passes through the logistic function. Values larger or equal to 0.5 are rounded to 1, otherwise to 0. For a predicted output of a sample, the indices where the value is 1 represents the assigned classes of that sample [8].
 6. Logistic Regression despite its name, is a linear model for classification rather than regression. Logistic regression is also known in the literature as logit regression, maximum-entropy classification (MaxEnt) or the log-linear classifier. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function. The implementation of logistic regression in scikit-learn can be

- accessed from class Logistic Regression. It is used to predict a continuous value output for each, for logistic regression it is required that labels should be continuous (numeric) like stock price, housing price [5]
7. Variance Threshold is a simple baseline approach to feature selection. It removes all features whose variance doesn't meet some threshold. By default, it removes all zero-variance features, i.e. features that have the same value in all samples. As an example, suppose that we have a dataset with boolean features, and we want to remove all features that are either one or zero (on or off) in more than 80% of the samples. Boolean features are Bernoulli random variables, and the variance of such variables is given by $\text{Var}[X] = p * (1-p)$ so we can select using the threshold $.8 * (1 - .8)$ [9].
 8. The principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point, and predict the label from these. The number of samples can be a user-defined constant (k-nearest neighbor learning), or vary based on the local density of points (radius-based neighbor learning). The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice. Neighbors-based methods are known as non-generalizing machine learning methods, since they simply "remember" all of its training data (possibly transformed into a fast indexing structure such as a Ball Tree). Nearest neighbors has been successful in a large number of classification and regression problems, including handwritten digits or satellite image scenes. Being a non-parametric method, it is often successful in classification situations where the decision boundary is very irregular. The classes in `sklearn.neighbors` can handle either Numpy arrays or `scipy.sparse` matrices as input. For dense matrices, a large number of possible distance metrics are supported. For sparse matrices, arbitrary Minkowski metrics are supported for searches. Neighbors-based classification is a type of instance-based learning or non-generalizing learning: it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point `scikit-learn` implements two different nearest neighbor. `KNeighborsClassifier` implements learning based on the k nearest neighbors of each query point, where k is an integer value specified by the user. Radius Neighbors Classifier implements learning based on the number of neighbors within a fixed radius r of each training point, where r is a floating-point value specified by the user. [14]
 9. An AdaBoost classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases. The core principle of AdaBoost is to fit a sequence of weak learners (i.e., models that are only slightly better than random guessing, such as small decision trees) on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote (or sum) to produce the final prediction. The data modifications at each so-called boosting iteration consist of applying weights w_1, w_2, \dots, w_N to each of the training

samples. Initially, those weights are all set to $w_i = 1/N$, so that the first step simply trains a weak learner on the original data. For each successive iteration, the sample weights are individually modified and the learning algorithm is reapplied to the reweighted data. At a given step, those training examples that were incorrectly predicted by the boosted model induced at the previous step have their weights increased, whereas the weights are decreased for those that were predicted correctly. As iterations proceed, examples that are difficult to predict receive ever-increasing influence. Each subsequent weak learner is thereby forced to concentrate on the examples that are missed by the previous ones in the sequence [13].

B. Classifiers that didn't work for this project

1. Linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space. PCA is used to decompose a multivariate dataset in a set of successive orthogonal components that explain a maximum amount of the variance. In scikit-learn, PCA is implemented as a transformer object that learns n components in its fit method, and can be used on new data to project it on these components. The PCA object also provides a probabilistic interpretation of the PCA that can give a likelihood of data based on the amount of variance it explains. As such it implements a score method that can be used in cross-validation. The PCA object is very useful, but has certain limitations for large datasets. The biggest limitation is that PCA only supports batch processing, which means all of the data to be processed must fit in main memory. The Incremental PCA object uses a different form of processing and allows for partial computations which almost exactly match the results of PCA while processing the data in a minibatch fashion [10].
2. The implementation of C-Support Vector Classification is based on libsvm. Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. Support vectors are effective in cases where number of dimensions is greater than the number of samples. Support vectors are versatile, different kernel functions can be specified for decision functions. Common kernels are provided but it is possible to specify custom kernels. The fit time complexity is more than quadratic with the number of samples which makes it hard to scale to dataset with more than a couple of 10000 samples. The method of Support Vector Classification can be extended to solve regression problems. This method is called Support Vector Regression. There are three different implementations of Support Vector Regression: SVR, NuSVR and LinearSVR. LinearSVR provides a faster implementation than SVR but only considers linear kernels, while NuSVR implements a slightly different formulation than SVR and LinearSVR [11].
3. XGBoost is eXtreme Gradient Boosting classifier. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code

runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples. It is open sourced tool which performs computation in C++ and the interface is provided by R/python/Julia. It is a tree based model which is variant of gradient boosting machine. It is hosted on github server and code is written in R studio [12].

IV. APPROACH TOWARDS THE GOAL

A. Pre-processing data

1. Analyzing the dataset
2. Eliminating the redundant data both manually and by using feature selection.
3. Initial data set consisted of redundant columns and constant values in many columns these columns were eliminated
4. After pre-processing I was able to reduce the number of features from 370 to 212 in training dataset while number of rows was left unchanged
5. The prediction for each classifiers on testing data was stored inside submission folder on my github account

B. Models

1. 10 fold cross validation with Random Forest resulted in most accurate results among all other classifiers that we used for our project
2. Gradient Boosting algorithm is described in [3]. When compared to accuracy with decision tree on training, gradient boosting gave better result
3. Submitted results predicted by training data set on testing dataset to kaggle for this competition and reported the accuracy on testing model
4. Applied logistic regression on pre-processed data and plotted ROC curve
5. Applied K-nearest neighbors with cluster size varying from 10 to 1000 applying each cluster size and noticed that best accuracy was provided when cluster size is 1000
6. Applied random forest with depth and estimators of 10. Depth of 10 was selected for a tree because when increased depth there was a decrease in accuracy
7. Applied cross validation with 10 folds on training data and reported result.
8. Applied adaboost classifier with decision tree as base classifier [13].

V. RESULTS AND DISCUSSION

On applying feature selection using variance threshold. The plotted features importance graph is represented as shown in Fig. 1. The receiver operating characteristic curve for the project is represented through Fig. 1.

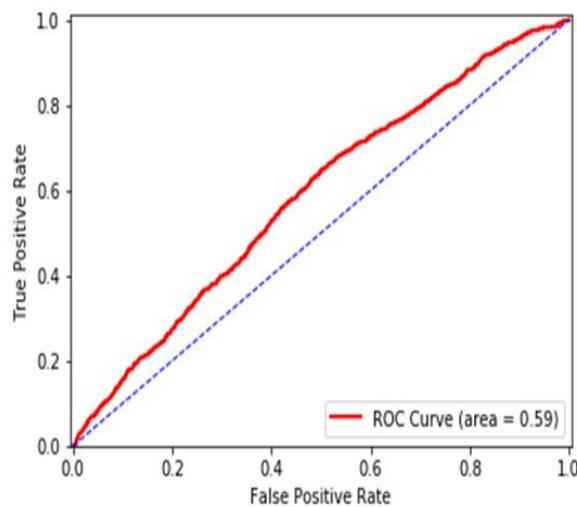


Fig. 1: Receiver Operating Characteristics

The accuracy for each algorithm that we recorded is reported in table below. Sorted according to the accuracy reported in Table I.

Table 1: Accuracies of Classifiers

	Training	Cross Validation
Random Forest	0.963036	0.942487
Logistic Regression	0.960624	0.960418
KNN	0.960115	0.960431
Neural Network	0.960115	0.960365
Gradient Boosting	0.954380	0.960431
Decision Tree	0.925852	0.820362
Ada-Boost	0.924440	0.820073

VI. RECOMMENDATIONS

We could improve the results for giving testing data if we apply eXtreme Gradient Boosting classifier

REFERENCES

- [1] <https://www.kaggle.com/c/santander-customer-satisfaction>
- [2] <https://www.santanderbank.com/us/about/about-us>
- [3] <http://scikitlearn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>
- [4] <http://scikitlearn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
- [5] <https://csns.calstatela.edu/site/f17/cs4661-1/item/6165525>
- [6] <http://ogrisel.github.io/scikitlearn.org/sklearntutorial/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- [7] <https://pdfs.semanticscholar.org/486e/88644a0acd9256eadbf3b757a53817e86581.pdf> http://scikitlearn.org/stable/modules/neural_networks_supervised.html
- [8] http://scikitlearn.org/stable/modules/feature_selection.html
- [9] <http://scikitlearn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
- [10] <http://scikitlearn.org/stable/modules/generated/sklearn.svm.SVC.html>
- [11] <https://www.youtube.com/watch?v=ufHo8ybk6g4>
- [12] <http://scikitlearn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>
- [13] <http://scikitlearn.org/stable/modules/neighbors.html>

Health Expert

Vinayak Sharma

*Vivekanand Education Society's Institute Of Technology
Mumbai, India
vinayak.sharma@ves.ac.in*

Shivani Subnani

*Vivekanand Education Society's Institute Of Technology
Mumbai, India
shivani.subnani@ves.ac.in*

Gayatri Mulchandani

*Vivekanand Education Society's Institute Of Technology
Mumbai, India
gayatri.mulchandani@ves.ac.in*

Gaurav Gianani

*Vivekanand Education Society's Institute Of Technology
Mumbai, India
gaurav.gianani@ves.ac.in*

Anjali Yeole

*Vivekanand Education Society's Institute Of Technology
Mumbai, India
anjali.yeole@ves.ac.in*

Abstract- The main purpose of data mining application in health care system is to develop automated tool for identifying and disseminating relevant healthcare information. The objective of our system to provide prediction of disease depending on symptoms so as to take proactive treatment against disease. The system would reduce the human effort, reduce cost and time constraint in terms of human resources and expertise, and increase the diagnostic accuracy. In most developing countries, insufficiency of medical specialist has increased the mortality of patients suffering from various diseases. Insufficiency of medical specialists will never be overcome in short period of time. The main idea of project is to assist doctors, which fail to detect fatal diseases. The intelligent doctor will accept symptoms of patient. The patient's own past history of diseases, maternal diseases, and paternal diseases will be stored. The symptoms, and the databases are matched to produce list of diseases and sufferings with their probabilities. We have used Apriority and FP Growth algorithm for predicting the disease for a given set of symptoms. The whole process can be termed as KDD.

Keywords: Disease Prediction, KDD (Knowledge Discovery in Databases and Data Mining), Apriori Algorithm, FP Growth Algorithm

I. INTRODUCTION

Every human being faces health related issues each & every day where one cannot come to know what kind of disease/health problems are been faced. To identify what is the cause depending on the symptoms you can come to know what kind of disease you have & what treatment is required. Certain issues related to resources, lack of knowledge in medical science such kind of process may help doctors as well as patient to cure their disease by prediction analysis. Detecting diseases at early stages can enable to overcome dangers.

However, waiting for students to become doctors and doctors to become specialists, many patients may already die. Current practice for medical treatment required patients to consult specialist for further diagnosis and treatment. Other medical practitioner may not have enough expertise or experience to deal with high risk diseases. As most of the high risk diseases could only be cured at the earlier stage, the patients may have to suffer for the rest of life. The AI is a study to evaluate human intelligence into computer technology. It starts with asking about symptoms to the patient, performs computations and provides the possible diseases. The system which is fully trained with knowledge & can display disease depending of what kind of symptom an individual posses. Also with such system one can keep track of their health.

II. KNOWLEDGE DISCOVERY IN DATABASES AND DATA MINING

KDD is the process of changing the low- level data into high-level knowledge. KDD includes statistics, database systems, computer programming, machine learning, and artificial intelligence. The Knowledge Discovery in Databases process comprise of a few steps leading from raw data collections to some form of new information. The iterative process consists of the following steps: [1]

- (1) *Data cleaning:* Also known as data cleansing it is a phase in which noise data and unrelated data are removed from the collection.
- (2) *Data integration:* At this stage, several data sources, often heterogeneous, may be shared in a common source.
- (3) *Data selection:* At this step, the data related to the analysis is decided on and retrieve from the data collection.
- (4) *Data transformation:* Also known as data consolidation, it is a phase in which the chosen data is

transformed into forms appropriate for the mining procedure.

(5) *Data mining*: It is the essential step in which clever techniques are applied to extract patterns potentially useful.

(6) *Pattern evaluation*: This step, firmly interesting patterns representing knowledge are known based on given measures

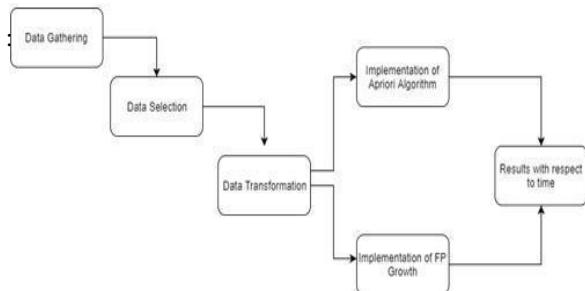


Fig. 1: Steps used carried out for prediction

PREPROCESSING:

Data gathering and selection: The raw data has been collected from the world wide web and data relevant to our purpose has been selected for further processing.

Data Transformation: After the first step the data is transformed into xls file so as to form a standard database. This xls file is extracted or read using file handling concept and stored in a database using MySQL. From the xls file only those information are read which are associated with the basic objective of our intended application. For example various stops words, verbs and adjective irrelevant to the application are kept behind and only the key meanings are Read from the xls file so that it becomes easier for the application to implement the algorithms i.e Apriori and FP Growth over the disease-symptom database. from the xls file so that it becomes easier for the application to implement the algorithms i.e Apriori and FP Growth over the disease-symptom database.

We have used Apriori and FP Growth algorithm for predicting the disease for a given set symptoms. These symptoms are provides by an user as inputs. On accepting these inputs the application executes these algorithms over them by accessing the database created using MySQL in step 2 during preprocessing stage.

A) APRIORI ALGORITHM

The Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules.

Apriori is a “bottom up” approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and group of candidates are tested against the data). Apriori is designed to operate on

database containing transactions, (for example: collection of items bought by customers).

Key Concepts are:

1. Frequent Itemset Search:
Obtain item occurrence:
Items that occur more than one times in the entire dataset.
Get frequent item sets.
Generate items that occur frequently.
2. Apriori property: Any subset of frequent item set must be frequent.
3. Obtain rules that have greater confidence :
Rules which satisfy minimum confidence are listed.

Apriori

```

// Mn: Item set of size N
// Fn : frequent item set of size N Step
1: F1 = {frequent items}; Step 2: for
(N = 1; N<=Fn; N++) {
Step 3: Mn+1 = Medical symptoms derived from Fn
after entering first symptom
Step 4: each t transaction in the database do {
Step 5: Increment count of all the symptoms in Mn+1
Step 6: Fn+1= min_support medical data in Mn+1
}
Step 7: end}
Step 8: return union NFn;
  
```

Here we have implemented the Apriori algorithm by generating only one candidate set. This is because here our motive is to predict only one disease for a set of inputted symptoms.

B) FP GROWTH ALGORITHM

FP Growth stands for frequent pattern growth. It is a scalable technique for mining frequent patterns in a database. FP Growth is a two step procedure.

Step 1: Build a compact data structure called the FP-Tree. (Build using two passes over the data set).

Step 2: Extracts frequent itemsets directly from the FP-Tree
FP tree Generation: Itemsets are considered in order of their descending value of support count. To facilitate tree traversal, an item header table is built so that each item points to its occurrences in the tree via a chain of node links.
Frequent item generation: Frequent items are directly extracted from the FP tree on the basis of maximum frequency.

III. METHODOLOGY

Fig. 2 shows the flowchart of the proposed system.

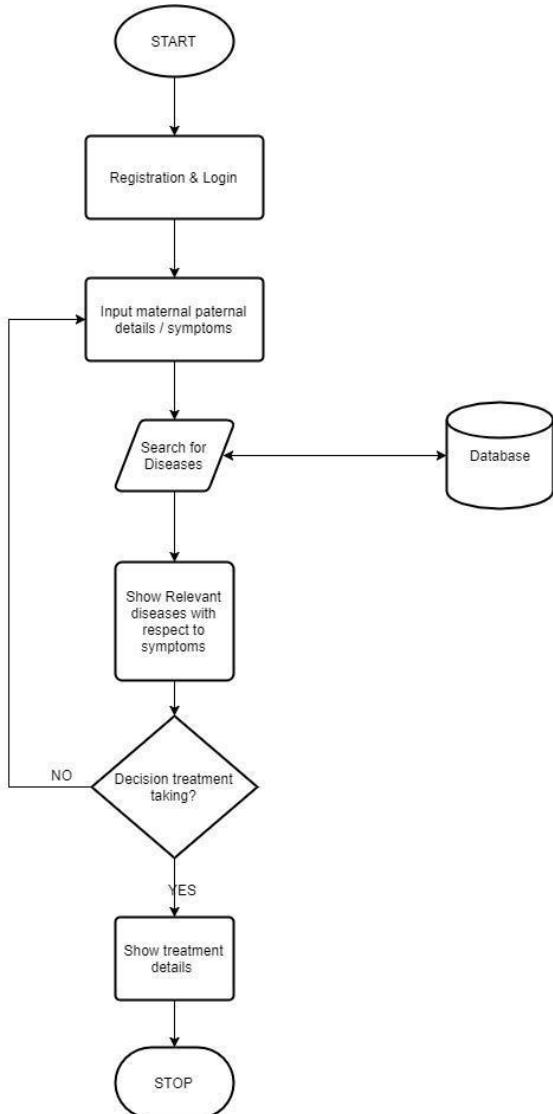


Fig. 2: Flowchart

REFERENCES

- [1] B.Srinivasan, K. Pavya, "A Study on Data Mining Prediction Techniques in HealthCare Sector", International Research Journal of Engineering and Technology (IRJET), vol. 3, no.3, 2016.
- [2] S.Vijayarani, M. Divya, " An Efficient Algorithm for Generating Classification Rules", IJCST ,vol. 2, Issue 4, 2011.
- [3] Muhamad Hariz Muhamad Adnan,Wahidah Husain, Nur'Aini Abdul Rashid(2012), "Data
- [4] "Mining for Medical Systems: A Review ,International conferences on advances in computer and information technology.
- [5] K. Rajalakshmi & Dr. S. S. Dhenakaran, "Analysis of Data mining Prediction Techniques in Healthcare Management System", International Journal of Advanced Research in Computer Science and Software Engineering,vol.5, no. 4, 2015.
- [6] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp 487-499, Santiago, Chile, September 1994.
- [7] Ruban. D.Canlas, "Data mining in Healthcare: Current applications and issues", IJAR, 2008.
- [8] HianChyeKoh and Gerald Tan, "Data Mining Applications in Healthcare, journal of Healthcare Information Management, vol. 19, no. 2, 2012.

IV. CONCLUSION

In this paper, we've proposed a system that aims at reducing mortality of patients suffering from various diseases. This system will benefit the doctors by making the diagnosis more reasonable. Here, we have presented a model with an efficient approach considering multiple parameters. Data sets are provided so that system can learn and map the symptoms. Data mining algorithms are used for machine learning implementation.

Automatic Human Detection System

Bhushan Nemade

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
bnemade@gmail.com*

Shreysijay Singh

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
jayshreysi@gmail.com*

Pooja Yadav

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
rpoojayadav218@gmail.com*

Kirti Vishwakarma

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
kirtyvish21@gmail.com*

Abstract—Nowadays Object Tracking is an important task in video processing because of its variety of applications in visual surveillance, human activity monitoring and recognition, traffic flow management etc. Multiple object detection and tracking in outdoor environment is a challenging task because of the problems raised by poor lighting conditions, variation in poses of human object, shape, size, clothing, etc. Proposed System proposes a novel technique for detection and tracking of multiple human objects in a video. Human objects are detected with help of this trained detector and are tracked using particle filter. The experimental results show that the proposed technique can detect and track multiple humans in a video adequately fast in the presence of poor lighting conditions, variation in poses of human objects, shape, size, clothing etc. and the technique can handle varying number of human objects in a video at various points of time. The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room.

Keywords-Learning based method, false annotations, poor alignment, Surveillance system, abnormal activity detection

I INTRODUCTION

The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room. According to our research, the existing system shows localization errors, false annotations and poor alignment while capturing the images. Training data becomes very large due to complexity of articulated human poses and varying conditions. To overcome these drawbacks, Human Detection Algorithm and Motion Detection Algorithm with enhanced features are used in the proposed system, it gives better alignment and discards double detection and vertical structure problem. The human occurrence is detected in few seconds and alerts are sent to the control room immediately.

Human Detection is an important task in video processing because of its variety of applications in visual surveillance, human activity

monitoring and recognition, traffic flow management etc. Multiple object detection and tracking in outdoor environment is a challenging task because of the problems raised by poor lighting conditions, variation in poses

of human object, shape, size, clothing, etc. It poses a novel technique for detection and tracking of multiple human objects in a video. The experimental results show that the proposed technique can detect and track multiple humans in a video adequately fast in the presence of poor lighting conditions, variation in poses of human objects, shape, size, clothing etc. and the technique can handle varying number of human objects in a video at various points of time.

II. RELATED THEORY

The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room. According to our research, the existing system shows localization errors, false annotations and poor alignment while capturing the images. Training data becomes very large due to complexity of articulated human poses and varying conditions. Existing motion detection security system detects the motion by sensing a change in field view of the camera that is connected with system and as soon as change is detected the alarm starts ringing. The limitation of this system is that it is not intelligent enough. When the motion is detected it does not sense whether change in the field view of the camera is due to an object, animal or human. Hence there was a need of more intelligent system therefore we have designed AHDS keeping in mind the limitation of existing system. The aim is to detect the motion and perform some intelligent operation before ringing the alarm.

III. ALGORITHM FOR HUMAN DETECTION

1. Learning-based methods for human detection

The core discriminative learning-based approaches train different kind of classifier on a large number of negative and positive

images samples, where humans are well framed. Every method has to extract appropriate features and the main apprehended information from training data is the spatially recurring of local shape events. If the trained classifier does not detect an object (misses the object) or mistakenly detects the absent object (false alarm), it is easy to make an adjustment by adding the corresponding positive or negative samples to the training set. However, because of the complexity of articulated human poses and varying viewing conditions, the training data become very large (especially positive samples), consequently the generalization capability of the trained classifier can be compromised. The basic common process of training data is shown in figure below. The features and the used classifiers for each technique are given in Common learning processes.

2. Human activity analysis

In this section, we proceed to evaluate and analyze the human activities as the detected objects and classify them into two groups: normal activities and abnormal activities, based on the support vector machine (SVM). The flow diagram for this step is shown in. The basic idea of support vector machines (SVM) is to find the optimal HYPERPLANE that splits a dataset into different categories. Once the HYPERPLANE is chosen, the distance to the nearest data point of the classes is maximized. This gives an idea about a simple example with two classes in the plane.

3. Motion detection system

Motion detection is the first essential process in the extraction of information regarding moving objects and makes use of stabilization in functional areas, such as tracking, classification, recognition, and so on. The frequently-used algorithms for motion detection are studied, including frame difference method and background subtraction method, and an algorithm composing of those methods for motion detection is proposed. This paper presents a new algorithm to detect moving objects within a scene acquired by stationary camera. In this paper Difference of successive frames and background were calculated by taking mean of n consecutive frames and then compare it with current frame with the help sub block matching-based scheme. It increases the sensitivity of human motion detection.

This algorithm presents an image with white pixels will be greater than a predefined alarm level (threshold), an alarm is produced about a motion event. This estimated background is just the previous frame. It evidently works only in particular conditions of objects, speed and frame rate. It is very sensitive to the threshold so that a noisy image motion will be detected in such places compared to places where there is no motion at all. If the object is moving smoothly, a small change is obtain which is less than the predefined threshold, so, it is impossible to detect moving object. Things become worse, when the object is moving very slowly, then the algorithms will not give any result at all.

IV MOTIVATION

The goal of the proposed system is to create a surveillance system based on the characteristics of the human presence and motion. The proposed system will be able to detect human presence without any use of sensors on the basis of the human characteristics and motion detection. In critical climatic condition such as rain and storm the scaling of the image is mostly accurate. Human detection is an essential and significant task in any intelligent video surveillance system, as it provides the fundamental information for semantic understanding of the video footages. It has an obvious extension to automotive applications due to the potential for improving safety systems. Many car manufacturers The security and surveillance cost will be reduced by using this system.

V. PROBLEM DEFINITION

Detecting human beings in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in a dense crowd, person identification, gender classification, fall detection for elderly people, etc. The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays. Thus, human detection and motion analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition.

VI. METHODOLOGY

The proposed methodology is illustrated in Figure1. It contains the following steps:- Take Video Input, Video Cropping, Background Subtraction, Motion Detection, Multi-Scale Video Cropping, Check Human Characteristics(Above 80% Correctness), Machine Learning Algorithm, Human Detected. The original image is a combination of stationary part and motion part. The AHD system makes a decision at each frame for a particular object being tracked. After certain number of frames, object being tracked, the AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity. The aim is to automatically guess the motion of a human or human body from monocular or multi-view video images.

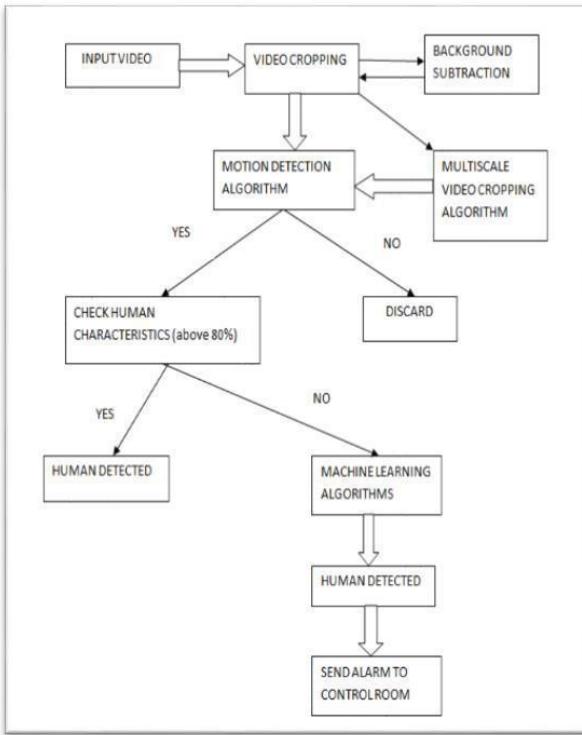


Fig:1: System Architecture

VII. SYSTEM ARCHITECTURE

Phases of our system are

1. Take Video Input
2. Video Cropping
3. Background Subtraction
4. Motion Detection,
5. Multi-Scale Video Cropping,
6. Check Human Characteristics(Above 80% Correctness)
7. Machine Learning Algorithm Human Detected.

1. Take Video Input

Video input of the surrounding will be taken from camera this video will consist of all the objects present in that surrounding frame and capture the video and use it as input image.

2. Video Cropping

Process the image for motion detection and if it fails then background subtraction is applied and in the next step we are using the motion detection algorithm to detect the human.

3. Background Subtraction

Use an image for human recognition using principal component analysis algorithm. The object appearing in front of the camera is subtracted from the background.

4. Motion Detection

The concerned object's motion is monitored and checked whether motion is present using training set to train the system for recognition.

5. Check Human Characteristics (Above 80% Correctness)

If 80% accuracy is not achieved machine learning algorithm is applied to it.

6. Machine Learning Algorithm Human Detected. At the end it generates the alarm in the control room when the human is detected.

VIII. EXPECTED RESULTS

1. Outputs

In Proposed system, the original image is a combination of stationary part and motion part. The AHD system makes a decision at each frame for a particular object being tracked. After certain number of frames, object being tracked, the AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity. The aim is to automatically guess the motion of a human or a human body from monocular or multi-view video images. The system provides functionality of human characteristics and motion. The system also provides search capability for the user can easily view the human characteristics records. The system also provides the functionality of sending notification and reminder via Alert.

Outcomes

This approach consists of hybrid features which allows a low-dimensional feature vector and higher recognition accuracy from image sequencing data for human activity recognition. These features include lighting condition recognition properties, invariant characteristics and human silhouettes normalization based on optical flow and distance parameters features. During experimental results, we used two challenging depth image datasets where proposed system has shown significant recognition accuracy performance over the state-of-the-art methods. The position of the human are detected at various distances using scaling method to provide the accuracy in recognizing the human.

Phase 1(Information Gathering)

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done using motion detection system to design.

Phase 2 (Analysis)

Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase. Training data becomes very large due to complexity of articulated human poses and varying conditions

Phase 3 (Design)

After certain number of frames, object being tracked. The AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity.

Phase 4 (Integration)

Constructing the various modules of the proposed solution to integrate them into a prototype.

Phase 5 (Testing)

The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stress testing.

Phase 6 (Deployment)

The benefits of Analysis on after deploying the proposed prototype include: Proper and more efficient way of monitoring of Human without them actually knowing it. Effective way of managing and storing image data. Reduce the waiting time of human detection. Faster Serving Process. More Secure location. Better access control by centralized user account through remote server. Better decision-making so as to detect more human.

REFERENCES

- [1] William Shackleford, Geraldine Cheok, Tsai Hong, "Performance Evaluation of Human Detection Systems for Robot Safety" (Conference of Springer Science+Business Media Dordrecht 2016)
- [2] Shanshan Zhang, Rodrigo Benenson, Mohamed, "How Far are We from Solving Pedestrian Detection" (IEEE Conference on Computer Vision and Pattern Recognition 2016)
- [3] Nagiouldtaleb, Adilchergui, Mohamed larbi ben maati, "Overview on Automatic Detection of Human Body" (IEEE conference on Advanced Video and Signal based Surveillance (AVSS) 978-1-5090-5146-5/16/\$31.00 ©2016 IEEE)
- [4] Malek Al-Nawashi, Obaida M. Al-Hazimeh, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments" (17 May 2016_ The Author(s) conference 2016).
- [5] Shahryar, Kamal Ahmad Jalal Daijin Kim, "A Hybrid feature Extraction Approach for Human Detection" (Conference of King Fahd University of Petroleum & Minerals 2015)

Opinion Mining of Customer's Feedback

Sanjana Puthran

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Shweta Mishra

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Abstract- In an e-commerce environment where millions of transactions take place between the providers and users, a need for establishment of validity of the service provided arises. A customer feedback system has been provided by the marketplace operators in order to fulfill such need. But the feedback generated may not be always relied upon. The feedback may positively or negatively affect its sales, instead of showcasing the actual genuineness of the product or service, in customer's point of view. This work proposes an enhancement to traditional feedback system by introducing an Trust Reputation System (TRS) which helps filtering out the valid consumer feedback using a set of algorithms, thereby creating a trust degree for the user.

Keywords: *Trust Reputation System, Opinion Mining, Sentiment Analysis.*

I. INTRODUCTION

The consumers in the online market face the problem of filtering out the best products from a list of variety of options. There are various marketplace operators who provide feedback system to help the customers to identify quality products, by reviewing the customer opinion and accordingly choose the product. Most of the consumers buy products based on product reviews. This either negative or positively affect the sale of the products. Also, this paves a way for spammers for decreasing the sale of the product. To eliminate this, the paper focuses on enhancing the feedback system by introducing the concept of trustworthiness .This can be done through Trust Reputation System. TRS are programs that allows customers to rate each other. Using such methods can help decrease the number of spammers, thereby potentially increasing the amount of genuine reviews. The advantage of such reviews is that it helps in determining the genuineness of the product.

II. RELATED WORK

Sentiment analysis has been studied in wide area of domain such as movie review, teaching review, product review, e-learning, hotel review and many more. Most scholars

Isheta Tilara

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Purvi Sankhe

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

focused to quantitative data analysis. However, some studies have been done on qualitative data using sentiment analysis, which helped finding six works that mentioned the idea of using opinion mining and sentiment analysis in education.

Algorithms such as Naive Bayes, k-means and Support Vector Machine are used in opinion classification. The paper also focuses on the truth reputation system. There exists several truth reputation system architectures having different algorithms to calculate the reputation score related to the product. Many authors [1][2][3][4][5][6] have proposed in their work several TRS architectures with different algorithms to calculate reputation score related to the product.

Also, a few academic work on Truth reputation system has been devoted to the inclusion of the semantic analysis of feedbacks in the calculation of the trust score of the product and specially the trust degree of the user. Even in studies attempting to provide more complex reputation methods, some issues are still not taken into consideration, such as the credibility of referees, the update of the trust degree of the user at any intervention, the age of the rating and the feedback or the concordance between the given rating which is a scalar value and the textual feedback associated to it. In contrast to the mentioned TRS, the proposed design overcomes these issues and makes use of an algorithm which includes analysis of textual feedbacks in order to calculate the trust degree of the user giving the feedback and a trustworthy reputation score for the product.

III. OVERVIEW

The consumers in the online market face the problem of filtering out the best reviews or feedback for purchase of the products. This work tries to eliminate the problem by listing out the best reviews so that it becomes easy for the customers to decide on a product by analysing other consumer experiences, by allowing them to post their reviews. Consumers dealing with the online market might sometimes buy substandard products.

Though the e-commerce company provide facilities like return and exchange of products, the process becomes a tedious task sometimes. The project aims to provide the customers an opportunity to select the desired products based on the rating of the item they wish or plan on to buy, which has been evaluated

on the basis of rating and reviews contributed by the consumers with the help of a Truth Reputation System (TRS).

IV. FEATURES OF THE PROJECT

The Opinion Mining of our project will be based on Sentiment analysis algorithms & methods and also on Truth Reputation System algorithm .Trust Reputation Systems (TRS) will provide the necessary information to support relying parties in taking the right decision in any electronic transaction. In fact, as security providers in e-services, TRS have to faithfully calculate the most trustworthy score for a targeted product or service. Thus, TRS must rely on a robust architecture and suitable algorithms that are able to select, store, generate and classify scores and feedbacks

V. PROPOSED WORK

In the proposed architecture, for each user who wants to leave a rating (appreciation) and a feedback (semantic review), we analyze the customers attitude towards a number of short and selected feedbacks and stored by product in the knowledge base. This user's review is going to be reached by any other user. Then, we suppose that we have a path relaying all the users (the nodes). As a result, we need to know the trust degree of the user and determine the trust degree of the feedback.[4]

A) Algorithm Description: Trust Reputation System Design

The customer starts by giving a rating and a textual feedback about a specific product. When they click on submit, in order to validate the given information, it will be directing the user to another interface showing this message for example: "please give us your opinion about the following feedbacks before validating the information you gave below:" In this interface customers will find chosen feedbacks from the database from different types. Those feedbacks can be fabricated in order to summarize numerous users feedbacks stored in the database. The generated feedbacks can be stored in another knowledge base. So as feedbacks are added in the ordinary database, it will fill the knowledge database with prefabricated feedbacks using text mining algorithms and tools. However, some users can give already summarized feedbacks that can directly be included in the knowledge database. Indeed, there are many text mining and data mining algorithms and tools that could search the most appropriate feedbacks that are first of all related to the product and that can recapitulate and summarize most of each type of the customers feedbacks. Actually, before sending the customers feedback and appreciation about the product to the trust reputation system, it has to verify the concordance between them in order to avoid and eliminate contradiction or malicious programs attacking our system. In the redirected interface, it will display several feedbacks from different types. However, the user can specify the number of feedbacks to be liked or disliked. Of course, it can also specify the minimum and the maximum number of feedbacks to be displayed by the user.

In fact, this project is trying through this redirection to detect and analyse the user intention behind his intervention on the e-commerce application. Hence, it examine and evaluate his intention using other prefabricated feedbacks with different types. Of course, it is already a trustworthiness of each feedback. Consequently, it uses its reputation algorithm studied in section [4] in order to generate the user trust degree which plays the role of a coefficient and then rectify his appreciation according to his trust degree and generates the score of the feedback. Indeed, each feedback has trustworthiness in a threshold [-5,5]. The closest is the trustworthiness to 5, the most trustworthy the feedback is. The closest is the trustworthiness to -5, the very untrustworthy is the feedback. If the feedback is trustworthy its score would be included in [0,5] else it would be included in [-5,0].

B) TRS algorithm

Reputation algorithm used in this TRS is using semantic feedbacks analysis in order to generate a trustful reputation score for the product. Actually, it has 3 types of feedbacks:

Positive feedbacks: Represent opinions that expressing a positive point of view about the product. Those ameliorative opinions contain a positive content concerning the product. Then, the adjective positive is referring to the nature of the content of the feedbacks not its trustworthiness. However, each feedback whatever is its type can have either a positive trustworthiness or a negative trustworthiness. Either positive trustworthiness or negative one, it is gradual: it has degrees as float in a threshold [-5,5].

Negative feedbacks: represent opinions talking negatively about the product. Logically, the users giving such opinions are not satisfied of the commented product. This feedback could be telling the truth or a part from the truth or could be far from the truth. That's why, each feedback has its trustworthiness represented by a float number between -5 and 5.

****mitigated feedbacks:** represent feedbacks that are talking positively about some aspects of the product and negatively about other aspects. They are also characterised by trustworthiness included in [-5,5].

****Contradictitious feedbacks:** represent feedbacks with a contradictitious content for example a feedback where the user is not talking about the specified product but another one or he/she is affirming that the camera of a mobile phone is great and later in the same opinion is saying that the camera is very bad. In fact, to it has to start by detecting the contradictitious feedbacks. Then it is in need of a semantic analysis algorithm and tool that can detect the contradiction in a specific content related to a product. It can personalize the analysis according to the product. For instance, if the user says that "the swimming pool of the hotel which does not afford one is not clean", the algorithm must be able to detect this great contradiction. It can give to the algorithm for each product as an input the property of the algorithm; if there is no similarity we can consider it as a contradiction. But the agreement includes the meaning of course. Because if the customer writes that the negative thing about this hotel is that there is no swimming pool. He is telling the truth then obviously the presence of an absent property in a feedback doesn't mean that there is a contradiction.

Actually, before sending the customers feedback and appreciation about the product to the trust reputation system, we have to verify the concordance and the alliance between them so we don't have contradiction.

After verifying the concordance between the appreciation and the textual feedback we are going to redirect the user to the selection of prefabricated feedbacks. Then the user is going to click on like or dislike according to each feedback. The event of click will be managed in order to get some information needed in the calculus of the trust degree of the user. The function uses as a parameter the id of the feedback in order to get from Knowledge base its trustworthiness. It needs to get also the previous trust degree of the user if he has been already engaged in a transaction or he has used the application for rating purpose. The user choices either "like" or "dislike" is an important parameter to determine his trustworthiness [4]. Initially, the user gives a rating and a textual feedback about the purchased product. Then it will validate the information provided through an interface. In fact, in this interface it will find chosen feedbacks from the database from different types. The feedbacks can be used to summarize numerous users feedbacks stored in the database. The generated feedbacks can be stored in other knowledge base. So as much as we add feedbacks in the ordinary database, it will fill the knowledge database with prefabricated feedbacks using text mining algorithms and tools. Some users can give summarized feedbacks that can directly be included in the knowledge base. Before sending the feedback to trust reputation system it has to go through concordance test, so that no contradiction of rating and feedback occurs.

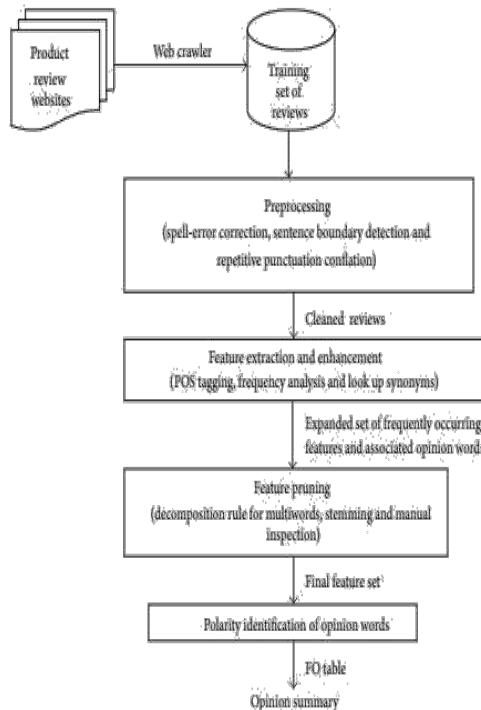


Fig. 1: System Design for Opinion Mining

VI. RESULT

```

To verify the concordance between the rating and the textual
feedback: Boolean concordance;
concordance =Test_ concordance (int appreciation, string
feedback) ;
If (concordance)
URL (url_feedbacks_interface); // redirection to the feedbacks

interface Else
URL (url_page);
  
```

After measuring the concordance the feedback is sent to Trust Reputation System for further processing. At the final stage we get only filtered feedback. Hence only genuine feedback about the product are generated.

VII. CONCLUSION

Lack of information regarding particular products leads to wrong selection of product which in turn leads to huge hole in the pockets of the customer. Thus, the project aims to provide accurate and true reviews about particular products which will help customers in picking up the right product. It attempts to calculate the trust degree of the user according to his subjective choice either "like" or "dislike" and according to the feedback. Those results such as trust weight and scores help users making a decision about purchasing or not a product from an e-commerce application. However those scores are not always truthful. Then, they can falsify the weight and the ratings. Semantic feedbacks are more meaningful than single scores.

VIII. FUTURE SCOPE

The consumers dealing with the website would be able to access precise data and reviews of the consumers feedback and use it intelligently for product selection and for buying of it as well. This software would be useful for any similar e-commerce business dealing with problems regarding the issues of trustworthiness of reviews. The provision of visual representation can be used by customers to buy genuine products. On some extent it would also help the marketplace operators and vendors to filter out their potential customers. In today's time data is said to be the biggest asset for any company or organization. Thus, it is of immense importance to analyses the data and get some results out of it.

IX. ACKNOWLEDGEMENT

We sincerely thank to our guide Mrs. Purvi Sankhe, our HOD Dr. Rajesh S. Bansode, our Dean Dr. Kamal Shah and our Principal Dr. B. K. Mishra for his/her guidance and support for carrying out our project work.

REFERENCES

- [1] Wararat Songpan, "The Analysis and Prediction of Customer Review Rating Using Opinion Mining" Department of Computer Science, Faculty of Science, Khon Kaen University, Mandalay, Myanmar

- [2] Khon Kaen, Thailand, A. Josang R. Hayward Simon Pope: Trust Network Analysis with Subjective Logic. Proceedings of the Second International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2008), Cap Esterel, France, August 2008
- [3] Fereshteh Ghazizadeh Ehsaei, Ab. Razak Che Hussin, "Acceptance of Feedbacks in Reputation Systems: The Role of Online Social Interactions", Information Management and Business Review Vol. 4, No. 7, pp. 391-401, July 2012 (ISSN 2220-3796).
- [1] Hasnae Rahimi, Hanan El Bakkali, "A New Reputation Algorithm for Evaluating Trustworthiness in E-Commerce Context Information Security Research Team (ISERT).
- [5] Kang Liu, Liheng Xu, and Jun Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", Journal of Information Technology, 2006.
- [6] A. Gutowska and A. Sloane: "Modelling the B2C Marketplace: Evaluation of a Reputation Metric for e-commerce", Proceedings of Web Information Systems and Technologies - WEBIST, pp. 212-226, 2009.

Price Trend Prediction Using Data Mining Algorithm

Vinesh Yadav

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
vineshyadav98@gmail.com*

Pramod Chauhan

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai India
pramodchauhan2912@gmail.com*

Sunil Chaurasia

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai India
sunilgl4matic@gmail.com*

Pranjali Kasture

*Department of Information Technology,
Thakur college of Engineering and Technology,
Mumbai, India
kasturepranjali@gmail.com*

Abstract: Data mining can be applied on past and present financial data to generate patterns and decision making algorithm. The system will analysis the historical stock data's of some companies. Based on different factors which affect the stock values of company like demand and supply of shares, popularity of company, profit earned. The proposed system will predict future price, buy and sell possibility of the share by representing it in graphical manner. Here user can select the company in which user is interested or who is interested to know about stock behaviour of that company. Going through many surveys, the conclusion we come across is that it is not possible to consistently predict moments with better than average result as -prices are affected by companies growth, revenue.

Keywords: Data mining; Data preprocessing; Stock value; Stock exchange

I. INTRODUCTION

Price trend prediction of stock market application provides a way of predicting the prices of stock. The objective of the object analytics system is to generate efficient result of prediction for the user. With the advancement of storage techniques and digitization of work in every field, the amount of stored data is tremendously increasing influence in information Technology has caused a sizeable change constantly.

The stock predictor is an application which provides the prediction of stocks shares' price of listed companies. In stock exchange, there are many losers and gainers in company in everyday before the closing of Bombay stock exchange. This is a very helpful application for prediction of price of share of Bombay stock exchange by which user will get to know price of different shares directly in the application.

The stock market process is full of uncertainty as well as is affected by many factors. Hence the Stock market prediction is

one of the important exertions in finance and business. Price trend prediction of stock market provides the efficient solution for the user to predict future prices of stock. The successful prediction of stock's future price will yield significant profit for the user.

II. LITRETURE SURVEY

Multi kernel learning method has higher degree of accuracy and lower degree of false prediction, compared to existing single kernel method. Integrating news and trading volume data with historical stock price information can significantly improve the significantly effectiveness of stock market volatility prediction compared to many existing prediction methods [1].

The prediction of price of stock market is nearly efficient as compared to previous prices of stock market [2]. The traditional time series of anomaly time detection is to find out one of the biggest point of outliers from a series of randomly generated numbers. In proposed system, it uses the time sequence of time series and subsequences features efficiently to discover anomalies [3]. Outlier algorithm is consistent with k-mean clustering algorithm, average return and successful rate is tested against algorithm as well as the prediction about two quantities is correct and satisfactory. Various outlier detection algorithm is used which is used efficiently [4].

III. PROPOSED SYSTEM

This proposed system, analyzes previous stock data for certain companies, with the help of certain parameters that affect stock value. The proposed system will implement these values in data mining algorithms. This will also help us to determine the values that particular stock will have in near future. Proposed system will determine the month's high stock prize and low stock prize with help of outlier data mining algorithms. The software maintains administrator level and user level. It generates report for prediction on previous stock. However, according to efficient market hypothesis, the market price will follow a random walk and permanent prediction strategy.

A traditional way to predict stock trend is using data mining technique on the basis of stock prices unfortunately, the data of stock price have many noises and for noisy data people always build stochastic volatility models to make prediction. Anomalies have long term predictability on the stock trend. In our method system will utilize anomalies to remove noise so that prediction can be done easily. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument added on an exchange. The successful prediction of a stock's future price could yield significant profit. The main purpose of this project is to achieve faster and efficient solution for stock prediction of Bombay Stock exchange which will meet user's expectation and fulfil specified requirement.

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument added on an exchange. The successful prediction of a stock's future price could yield significant profit. The main purpose of this project is to achieve faster and efficient solution for stock prediction of Bombay Stock exchange which will meet user's expectation and fulfil specified requirement. Many banks, financial institutions, large-scale investors and stockbrokers have to buy and sell stock in shortest possible time

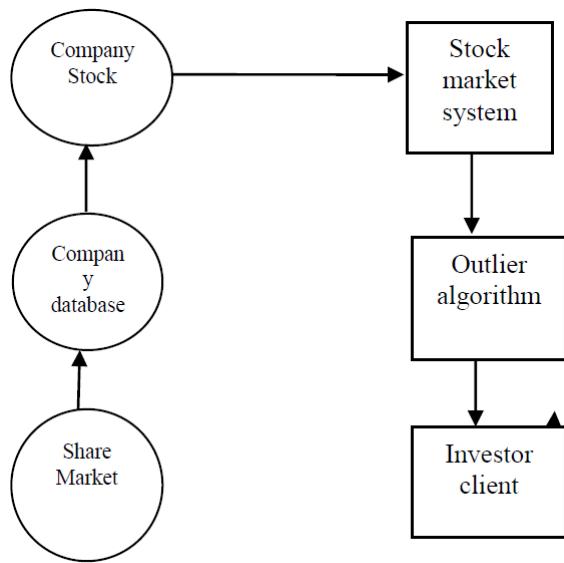


Fig. 1: DFD diagram for price trend prediction

IV. METHODOLOGY

Past stock prices data of listed companies will be taken for prediction. Data will be selected for preprocessing to remove null values and holidays in data. Normalization is applied to reprocessed data for better comparisons of stock data. Data mining algorithm is used to find the pattern recognition to find whether user should buy or sell stock. Figure 1 shows the following data flow diagram for price trend prediction and figure 2 shows flow diagram of price trend prediction

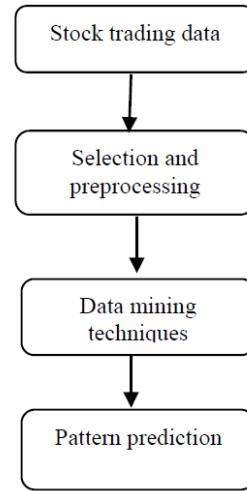


Fig. 2: Flow diagram of price trend prediction

V. EXPECTED OUTPUT

The expected outcome of proposed system shows the following figure 3 and figure 4 which describe the input interface for user as well as admin for mode of operation in system. Admin level has login which handles the database for stock values of listed companies. In user level, login as user consist of username and password as well as have to enter the number of company to see the stock values. User have to enter the start date and end date for corresponding companies in specified date format. Graph is generated which shows the flow of the stock values of companies for prediction of stock price. Stock values are normalized and preprocessed to remove the null values and holidays as well as plotted the stock price of specified companies.

The screenshot shows a terminal window titled 'Price Trend Prediction of Stock Market'. The user enters '3' for the mode selection, indicating they want to log in as an existing user. They then list the available companies: WIPRO.BO, SBIN.BO, OFSI.BO, BOMBEING.BO, ICICIBANK.BO, KOTAKBANK.BO, LITFH.BO, and RCOM.BO. The user then enters the count of companies to view, followed by the names of two companies: WIPRO.BO and SBIN.BO. Finally, the user specifies the start date as '2017-10-08' and the end date as '2017-10-18'.

Fig. 3: Input values for companies

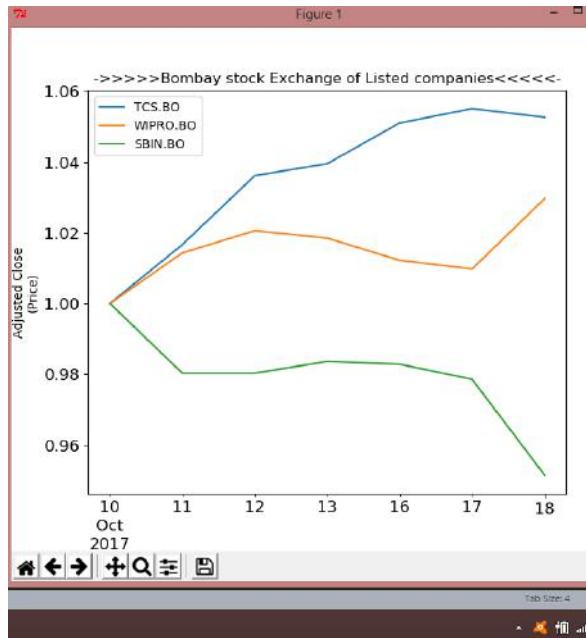


Fig. 4: Stock values of companies

VI. FUTURE SCOPE

Data mining automates the process of finding the way of predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical example of the predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, as well as identifying segments of a population likely to respond similarly to given events.

Although there has been some study regarding trend analysis and trading of stock market indices; a detailed study is, however, required for developing an intelligence system for making decisions regarding buying and selling of stocks and portfolios. In this regard it is suggested to develop effective learning procedures for taking accurate decisions in stock market trading.

Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

VII. CONCLUSION

The main idea of prediction of stock is based on outlier detection algorithm. In this clustering based approach is followed so that efficient prediction of future stock prices of some companies can be generated. The price trend prediction

of stock market is based on machine learning which analyzes previous data. Therefore it is important to detect outlier from extracted data. There are so clustering algorithm as one of efficient technique. As future lies, further advancement is going on in outlier detection methods. It helps in detection of outlier for imprecise and incomplete data set.

REFERENCES

- [1] Feng wang, Ling Liu and Chenxiao Dou " Stock Market Volatility Predict ion: A service-oriented Multi-kernel learning approach", IEEE Ninth International Conference on services Computing, June 2012.
- [2] Chen Ming Guo, Ling-Yu Xu, Hui-Fang Liu, Lie Wang, Bo Han "The Financial data Anomaly Detection Research based on Time Series", International Conference on Computer Science and Application, January 2015.
- [3] Zhao Lei, Wang Lin "price trend prediction of stock market using Outlier Data mining Algorithm" , IEEE Fifth International Conference on Big Data and Cloud Computing, October 2015
- [4] Rashi Bansal, Nishant Gaur, Dr. shailendra Narayan Singh, "Outlier Detection: Application and Techniques in Data Mining", International Conference on Computer Science and Application, January 2015.
- [5] http://www.investopedia.com/articles/07/mean_reversion_martingale.asp
- [6] <http://www.nseindia.com/content/research>
- [7] http://en.wikipedia.org/wiki/stock_market_prediction

A Survey on Sentiment Analysis Algorithms for Opinion Mining

Dhiraj Patil

Department of Information Technology
Thakur College of Engineering & Technology
dhirajpatil46@yahoo.com

Zaheer Alam

Department of Information Technology
Thakur College of Engineering & Technology
zaheer.alam@thakureducation.org

ABSTRACT - Conclusion mining and assessment examination is quickly developing zone. There are various online business locales accessible on web which gives choices to clients to give criticism about particular item. These criticisms are particularly useful to both the people, who will purchase that item and the associations. An exact technique for anticipating assessments could empower us, to remove feelings from the web and foresee customer's inclinations. There are different calculations accessible for sentiment mining. Before applying any calculation for extremity recognition, pre-preparing on input is completed. From these pre-prepared audits feeling words and protest on which conclusion is produced are removed and any sentiment mining strategy is connected to discover the extremity of the survey. Feeling mining has three levels of granularities: Document level, Sentence level and Aspect level. In this paper different calculations for notion examination are contemplated and difficulties and applications show up in this field are talked about.

Keywords - Sentiment Analysis, Opinion Mining, Web Content, Machine Learning

I. INTRODUCTION

Slant Analysis (SA) or Opinion Mining (OM) is the computational investigation of people's assessments, mentalities and feelings toward an element [3]. As a rule, feeling mining gathers data about the positive and negative parts of a specific subject. At last, the positive and exceptionally scored assessments acquired about a specific item are prescribed to the client. So as to advance showcasing, extensive organizations and representatives are making utilization of sentiment mining [4] .

Much research exists on feeling examination of client conclusion information, which for the most part judges the polarities of client audits. In these examinations, assessment investigation is regularly directed at one of the three levels: the record level, sentence level, or quality level. In connection to conclusion investigation, the writing overview done demonstrates two kinds of systems including machine learning and semantic introduction are imperative [3].These procedures are appeared in figure 1.

There are a few difficulties in Sentiment examination. The first is a feeling word that is thought to be certain in one circumstance might be viewed as negative in another circumstance. A moment challenge is that individuals don't generally express assessments samy. Most customary content preparing depends on the way that little contrasts between two bits of content don't change the significance in particular. In Sentiment investigation, in any case, "the photo was extraordinary" is altogether different from "the

photo was not awesome". Individuals can be opposing in their proclamations [4].

Most surveys will have both positive and negative remarks, which is to some degree reasonable by examining sentences each one in turn.

Clients express their sentiments about items or administrations they expend in blog entries, shopping locales, or audit destinations. It is helpful for both the customers and also for the makers to realize what overall population think about a specific item or service[6] . In the casual medium like twitter or web journals, the more probable individuals are to consolidate distinctive sentiments in a similar sentence which is simple for a human to see, however more troublesome for a PC to parse. Some of the time even other individuals experience issues understanding what somebody thought in view of a short bit of content since it needs setting [5] .

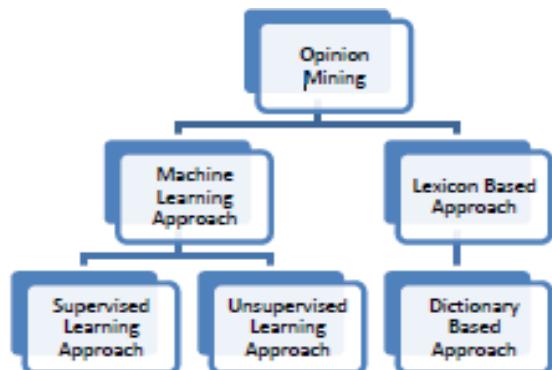


Figure 1. Assessment Mining Techniques

There are 3 levels of sentiment mining:

1.Document Level

In this methodologies entire record is considers as a solitary substance and the examination approaches in connected all in all record. The outcome produced in report level at times not suitable [5,6].

2.Sentence Level

In the sentence level methodologies each sentence is considered as an element and investigation approaches is connected on person sentence then their outcome is condensed to give the general consequence of the report [5,6].

3.Aspect Level

Expression level supposition mining is otherwise called viewpoint based conclusion mining. It performs fine grained investigation and straightforwardly takes a gander at the sentiment. The objective of this level of examination is to find feelings on parts of things [5,6].

- Aspects that are expressly said as things or thing phrases in a sentence are called as unequivocal perspectives.

- Implicit perspectives are not expressly said in a sentence but rather are inferred

There are distinctive sorts of calculations to break down assumptions. In this paper every one of the systems utilized for sentiment mining is overviewed. Issues and uses of conclusion mining are additionally examined.

Whatever remains of the paper is composed in the accompanying way. Area 2 gives the diagram of the systems utilized for sentiment mining with its related work. Segment 3 contains issues in sentiment mining. Segment 4 contains utilizations of sentiment mining . Area 4 contains conclusion.

II. RELATED WORK

2.1 Supervised Learning Approach

This strategy contains two arrangements of records which are preparing and a test set. To find out about the report, preparing set is utilized by classifier. For approval reason test set is utilized. For audit order numerous systems can be utilized. Kinds of administered learning strategies:

2.1.1 Decision tree classifier

Choice tree classifier gives a progressive deterioration of the preparation information space in which a condition on the credit esteem is utilized to isolate the information. The condition or predicate is the nearness or nonattendance of at least one words. The division of the information space is done recursively until the point when the leaf hubs contain certain base quantities of records which are utilized with the end goal of arrangement. In [7] Movie audit highlights got from IMDb was removed utilizing opposite report recurrence and the significance of the word found. Chief part examination and CART were utilized for highlight choice in light of the significance of the work as for the whole record. The order exactness acquired by LVQ was 75%. Investigating passionate variety in pre-adult age and purposes for these progressions utilizing information mining strategies is proposed in [11]. By characterizing feelings and utilizing choice tree distinctive enthusiastic varieties are broke down. On the off chance that at that point rules are additionally produced from choice tree. Exception examination is utilized to distinguish feeling variety in kid having any sort of incapacity.

2.1.2 Linear classifier

a. Support vector machine:

Content information are in a perfect world suited for SVM grouping as a result of the meager idea of content, in which few highlights are unessential, yet they have a tendency to be connected with each other and by and large sorted out into straightly distinguishable classes.

In [10], machine learning (SVM) joined with space particular dictionaries is actualized for angle arrangement and extremity recognizable proof of item survey. SVM is prepared to display

viewpoint characterization and this prepared SVM is utilized for extremity order per angle. The exploratory outcomes show that the proposed strategies have accomplished around 78% precision. Electronic information are connected to feeling cause extraction sub framework and correlative component determination strategy, in view of the yield of these highlights are blended. In preparing process, web post with obscure feelings are given to SVM and SVR grouping model and the yield gives data about the sort of feeling [13].

b. Neural arrange

Neural Network comprises of numerous neurons where the neuron is its essential unit. The contributions to the neurons are signified by the vector over line X_i which is the word frequencies in the i th record. There are an arrangement of weights A which are related with every neuron utilized as a part of request to register a component of its sources of info. In view of sources of info and weights yield is produced.

2.1.3 Rule based classifier

In manage based classifiers, the information space is demonstrated with an arrangement of standards. The left hand side speaks to a condition on the list of capabilities communicated in disjunctive typical frame while the correct hand side is the class name. The conditions are on the term nearness. Term nonattendance is infrequently utilized in light of the fact that it isn't useful in meager information.[8] proposes a control based way to deal with feeling cause part location for Chinese smaller scale web journals. It introduces the feeling model and concentrates the relating cause segments in fine-grained feelings. The enthusiastic dictionary can be built physically and consequently from the corpus. In the interim, the extents of cause parts can be ascertained in the impact of the multi-dialect highlights in light of Bayesian likelihood. The examination comes about demonstrate the attainability of the approach.

2.1.4 Probabilistic classifier

a. Naïve bayes

The Naïve Bayes classifier is the least difficult and most usually utilized classifier. Gullible Bayes characterization show figures the back likelihood of a class, in light of the dissemination of the words in the report. The model works with the BOWs highlight extraction which disregards the position of the word in the record. It utilizes Bayes Theorem to anticipate the likelihood that a given list of capabilities has a place with a specific mark.

The framework which is proposed in [6] extricates angles in item client surveys. The things and thing phrases are extricated from each audit sentence. Least help edge is utilized to locate every regular viewpoint for the given survey sentences. Gullible Bayesian calculation utilizing regulated term checking based approach is utilized to recognize whether sentence is certain or negative supposition and furthermore distinguishes its quantity.

The paper [12] presents a technique for conclusion investigation, on the survey made by clients to motion pictures. Order of audits in both positive and negative classes is done in view of a guileless Bayes calculation. As preparing information we utilized a gathering (pre-characterized in positive and negative) of sentences taken from the film surveys. To enhance arrangement we evacuated irrelevant words and presented in characterization gatherings of words (n-grams). For $n = 2$ bunches we accomplished a considerable change in order.

b.Bayesian organize

The principle suspicion of the NB classifier is the freedom of the highlights. The other outrageous presumption is to expect that every one of the highlights are completely reliant. This prompts the Bayesian Network display which is a coordinated non-cyclic diagram whose hubs speak to arbitrary factors, and edges speak to contingent conditions. BN is viewed as an entire model for the factors and their connections. In Text mining, the calculation multifaceted nature of BN is extremely costly; that is the reason, it isn't often utilized.

c.Maximum entropy

The Maximum entropy Classifier (known as a restrictive exponential classifier) changes over named include sets to vectors utilizing encoding. This encoded vector is then used to compute weights for each component that would then be able to be consolidated to decide the undoubtedly mark for a list of capabilities.

In [11], a novel strategy is utilized to gather different students twitter messages On this dataset preprocessing for opinion examination is performed It includes different middle of the road operations evacuate equivocalness. The pre-processed dataset is utilized to constructed user's enthusiastic state arrangement and SVM, ME and innocent bayes classifiers are connected and the outcomes are exceptionally effective.

Table I. Accuracy of Various Supervised Algorithm

Title, Author, Publication	Method	Accuracy
<p>Title: Opinion Mining Using Decision Tree Based Feature Selection Through Manhattan Hierarchical Cluster Measure [7]</p> <p>Author: Jeevanandam Jotheeswaran, Dr. Y. S. Kumaraswamy</p> <p>Publication: Journal of Theoretical and Applied Information Technology, 2013</p>	Naive bayes	Precision: 0.50 Recall: 0.79
<p>Title: Sentiment Analysis: Measuring Opinions [10]</p> <p>Author: Chetashri Bhadane, Hardi Dalal, Heenal Doshi</p> <p>Publication: Science Direct, 2015</p>	SVM	78%
<p>Title: Analysis And Identifying Variation In Human Emotion Through Data Mining [11]</p> <p>Author: Jasakaran Kaur, Sheveta Vashisht</p>	Decision Tree	NA

Publication: Int.J Computer Technology & Applications, 2012			
Title: Twitter Sentiment Mining (Tsm) Framework Based Learners Emotional State Classification And Visualization For E-Learning System [9] Author: M.Ravichandran, G.Kulanthaivel	SVM, Maximum Entropy	95%, 95%	
Publication: Journal of Theoretical and Applied Information Technology, 2014			
Title: A Rule-Based Approach To Emotion Cause Detection For Chinese Micro-Blogs ^[8] Author: Kai Gao, Hua Xu, Jiushuo Wang Publication: ELSEVIER, 2015	Assiciation rule	75%	
Title: Extracting Aspects And Mining Opinions In Product Reviews Using Supervised Learning Algorithm ^[6] Author: A.Jeyapriya, C.S.Kanimozhi Selvi Publication: IEEE, 2015	Frequent itemset mining, Naive bayes	92%	
Title: Applying Supervised Opinion Mining Techniques On Online User Reviews ^[12] Author: Ion Smeureanu, Cristian Bucur Publication: Informatica Economică, 2012	Naive bayes, n-gram	80%	

2.2 Dictionary Based Approach

In this approach first of all a small set of sentiment words which are known as seed words are collected manually with their known positive or negative orientations. Then this set is grown by searching their synonyms and antonyms in WordNet or another online dictionary. The new words are added to the existing seed list. Then next iteration is started. The iteration should be stopped when no new words are found. Manual inspection set is used at last to clean up the list.

In [18], Wordnet is used as dictionary Author uses mobile phone reviews from amazon website. It is input to the system. Polarity is calculated on the basis of majority of opinion words. Experimental results of „AIRC Sentiment analyzer system” is compared with proposed system and proposed system provides better accuracy. In future, some enhancements in this technique will be carried out. It would deal with the sentences contain relative clauses like not only-but also and the sentences contain clauses neither-nor, either-or etc.

In paper [19] an Aspect based Opinion Mining system named as “Aspect based Sentiment Orientation System” is proposed which extracts the feature and opinions from sentences and determines whether the given sentences are positive, negative or neutral for each feature. Negation is also handled by the system. To determine the semantic orientation of the sentences a dictionary based technique of the unsupervised approach is adopted. To determine the opinion words and their synonyms and antonyms WordNet is used as a dictionary. All the features of the product on which reviews are given would be identified and the orientation of the sentence for each feature would be determined. The polarity of the given sentence is determined on the basis of the majority of opinion words. In the end the system will generate the feature wise summary of positive, negative and neutral sentences which will be easier for users to read, analyse and help them in taking the decision whether the product is to be purchased or not.

Table 2. Accuracy of Various Dictionary-based Algorithms

Title, Author, Publication	Method	Accuracy
Title: Mining Of Product Reviews At Aspect Level [19] Author: Richa Sharma, Shweta Nigam and Rekha Jain Publication: International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.3, May 2014	Dictionary based unsupervised learning	74%
Title: Polarity Detection at Sentence Level [18] Author: Richa Sharma, Shweta Nigam, Rekha Jain Publication: International Journal of Computer Applications, Volume 86- No 11, 2014	Lexicon dictionary based approach	67%

III. ISSUES OF OPINION MINING

A positive or negative notion word may have their contrary significance in a specific space so it is difficult to anticipate by its watchword meaning [10].

Interrogative Sentence An interrogative sentence might not have neither positive nor negative supposition but rather the catchphrase utilized as a part of the feeling might be sure or negative[6].

Sarcastic Sentences Few sentences as muscle heads may damage the significance of the entire sentences such sort of sentence require a power complete consideration toward the catchphrases and sentences. These entertaining sentences not just violet the sentence of a specific sentence yet additionally demolish the estimation of the entire document[7].

Sentiment without estimation words once in a while suppositions does not utilize any notion words like great , better , best , most exceedingly bad ,awful and so on yet the sentences may have its positive or negative criticism about the item , administrations and strategies.

Conditional sentences contingent sentences are additionally an issue in Sentiment mining restrictive sentences is likewise making a similar issue like interrogative sentences[7].

Author and Reader understanding point (individual to individual fluctuating) Dollar cost is expanding concerning Indian rupee. This record have both the constructive and antagonistic importance and its esteem is shifting from individual to individual. This sentence has the positive notion for the Exporter while this same sentence has the negative supposition for the shippers [9].

Spam Reviews Spam suppositions are those notions which are posted by the inverse or contender association for expanding their item esteem or their association esteem among the clients. Some government official may utilize a similar spam survey to only for their reputation.

IV. APPLICATIONS OF OPINION MINING

- Business and internet business applications, for example, item surveys and film appraisals [3]
- Opinions in the social and geopolitical setting
- Predicting stock costs in view of assessments that individuals have about the organizations and assets [3]
- Determine territories of an item that should be enhanced by condensing item surveys to perceive what parts of the item are by and large thought to be great or awful by clients [4]
- Customer inclination

V. CONCLUSION

Feeling examination has turned out to be exceptionally well known field of research. A great deal has been examined in this field yet at the same time there are numerous issues as assumption examination forms content based unstructured information. Word reference based approach takes less handling time than regulated learning approach however exactness isn't up to the check. Regulated learning approach gives better exactness. From this study, it can be reasoned that directed systems give better exactness contrasted with word reference based approach.

In future, different feeling synopsis calculations ought to be connected to create outline of all surveys gave by clients.

REFERENCES

- [1] Jiawei Han, Micheline Kamber and Jian Pei, "Data mining Concepts and Techniques", Third Edition, Morgan Kaufmann Series in Data management Systems
- [2] Charu C. Aggarwal, "Data Mining: The Textbook", Springer, 2015
- [3] Walaa Medhat, Ahmed Hassan, HodaKorashy, "Sentiment Analysis Algorithms And Applications: A Survey", Ain Shams Engineering Journal, 2014
- [4] Blessy Selvam1 , S.Abirami2, "A Survey On Opinion Mining Framework", International Journal of AdvancedResearch in Computer and Communication Engineering, 2013
- [5] G.Vinodhini, R.M.Chandrasekaran, "Sentiment Analysis And Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, June 2012
- [6] A.Jeyapriya, C.S.KanimozhiSelvi, "Extracting Aspects And Mining Opinions In Product Reviews Using Supervised Learning Algorithm", IEEE, 2015
- [7] JeevanandamJotheeswaran, Dr. Y. S. Kumaraswamy, "Opinion Mining Using Decision Tree Based Feature Selection Through Manhattan Hierarchical Cluster Measure", Journal of Theoretical and Applied Information Technology, 2013
- [8] Kai Gao, Hua Xu, JiushuoWanga, "A Rule-Based Approach To Emotion Cause Detection For Chinese Micro-Blogs", ELSEVIER, 2015
- [9] M.Ravichandran, G.Kulanthaivel, "Twitter Sentiment Mining (Tsm) Framework Based Learners Emotional State Classification And Visualization For E-Learning System", Journal of Theoretical and Applied Information Technology, 2014
- [10] Chetashri Bhadane,Hardi Dalal, Heenal Doshi, "Sentiment Analysis: Measuring Opinions", Science Direct, 2015
- [11] Jasakaran Kaur, ShevetaVashisht, "Analysis And Identifying Variation In Human Emotion Through Data Mining", Int.J.Computer Technology & Applications, 2012
- [12] Ion Smeureanu, Cristian Bucur, "Applying Supervised Opinion Mining Techniques On Online User Reviews", Informatica Economică, 2012
- [13] Weiyuan Li, Hua Xu, "Text-based emotion classification using emotion cause extraction", ELSEVIER, 2013
- [14] Rizvaan Irfan, Christine K. King , "A Survey on Text Mining in Social Networks", The Knowledge Engineering Review, 2004
- [15] Diksha Sahni, Gaurav Aggarwal, "Recognizing Emotions and Sentiments in Text: A Survey ", International Journal
- [16] Faiza Belbachir, B'en'edite Le Grand, "Opinion Detection: Influence Factors", IEEE, 2015.
- [17] Reshma Bhonde, Binita Bhagwat, SayaliIngulkar, ApekshaPande, "Sentiment Analysis Based on Dictionary Approach", International Journal of Emerging Engineering Research and Technology Volume 3, Issue 1,2015
- [18] Richa Sharma, Shweta Nigam, Rekha Jain, "Polarity Detection at Sentence Level", International Journal of Computer Applications, Volume 86- No 11, 2014
- [19] Richa Sharma,Shweta Nigam and Rekha Jain, "Mining Of Product Reviews At Aspect Level", International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.3, 2000.

An empirical study of the impact of Knowledge Management on E-Supply Chain Management in Indian public sector oil companies

Aaditya Ajit Desai

Department of Information Technology

TCET, Mumbai, Maharashtra, India

aaditya1982@gmail.com

Sunil Kumar Rai

Department of Information Technology

NMIMS, Mumbai, Maharashtra, India

Email- sunilrai582003@yahoo.com

Abstract: The Indian public sector Oil Industry is vastly dependent on imports. The major oil producers in India are government public sector organizations like BPCL, HPCL and IOCL. It is said that the usage of technology, especially the software and related hardware is limited. Knowledge Management (KM) can help to improve the Supply Chain Management (SCM) by improving its most important processes through tools, procedures and practices.

Keywords – Knowledge Management (KM), Supply Chain Management (SCM), Downstream, Public Sector Oil companies, Bharat Petroleum Corporation Limited (BPCL), Hindustan Petroleum Corporation Limited (HPCL), Indian Oil Corporation Limited (IOCL).

I. INTRODUCTION

The choice of public sector oil industry was because the processes followed, the tools used and practices followed particularly here are complicated. There are four levels of functions followed in the industry like customers, retailers, depot managers and general managers.

India exports about 70% of its oil from other countries and the GDP of India is also affected by the variations in the oil prices to a large extent. Downstream supply chain management has a greater impact on the overall profitability of the company compared to upstream supply chain management which has certain limitations. For the survey, Mumbai city was the best choice as Mumbai has a greater impact on the *consumer and industrial usage of oil* as a resource. Also, Mumbai region has a *dense population*, having a representation from almost all parts of India. Also the major corporate offices of the Indian public sector oil industry like BPCL, HPCL and IOCL are located in Mumbai.

II. RESEARCH MOTIVATION AND OBJECTIVES

2.1 Motivation of the research: Due to the complex nature of the oil industry, it was chosen in the research where decision making in short term (like operational and transaction processing) and long term (like strategic level planning) is difficult. The focus of research is on looking out for processes in the organization which can be redesigned or improved with the help of Knowledge Management.

2.2 Objectives of the research: The objective of this research is to analyze the impact of KM on operational and strategic planning for various stakeholders in Indian oil sector companies.

2.2.1 Following are the research objectives identified:

1. To understand the use of Knowledge Management in downstream Supply Chain Management in oil sector around the world.
2. To identify the critical issues in collaborative decision making at strategic, middle and operations level.
3. To investigate those issues that are present (or not) and to what extent in the oil companies in India.
4. To ascertain the level of Knowledge Management implementation in downstream Supply Chain Management in Indian oil companies.
5. To ascertain the parameters of KM implementation in downstream SCM in Indian oil companies.

III. PROBLEM DEFINITION

3.1 Problem Definition: The Literature Survey and initial sample survey has led to the identification of the following problems:

1. The public sector oil companies in India use E-Supply Chain Management but the common issues in E-supply chain management remain unsolved.
2. The international oil companies use Knowledge Management extensively at all three levels: operations management, middle management and strategic management.

Some common problems faced by E-Supply Chain Management of oil companies in India and outside India are resolved using Knowledge Management this is evident from Table 1. The research carried out will focus on whether the Indian oil companies use KM in the downstream operations of E-SCM.

If yes, then what impact will KM have on the downstream operations of E-SCM and at which levels of the management will Knowledge Management be used.

IV. EXPERIMENT AND RESULTS

4.1 Literature Survey: The Literature Survey about E-SCM and KM helped us to get the facts, issues and parameters related to E-SCM and KM.

4.1.1 Issues and Parameters: 50 Issues and parameters were identified as a part of the literature survey. The issues and parameters were used as an input for the formation of the questionnaire.

4.1.2 Initial Survey: Pilot survey was carried out over a sample of customers, retailers, depot managers and general managers. The results were tested with reliability test like Cronbach's alpha test. Adjustments like change in the language of the questionnaire, change in the scale of the questionnaire were carried out.

4.1.3 Pilot Survey Outcome:

Refer Table 2 for Pilot Survey

4.1.4 Questionnaire Survey I:

4.1.4.1 Formed out of interview:

Interviews were carried out at various levels like general managers, retailers, depot managers and customers. The findings of the interview are placed in the annexure placed at the end.

4.1.4.2 Formation of questionnaire:

Questionnaire formation was based on the interviews taken of the experts and the literature survey done.

4.1.5 Validation of Questionnaire:

Validation of Questionnaire was done using Cronbach's factor for checking the consistency of data collected. Following are the results observed in Table 3.

4.2.1 Formation of questionnaire:

Questionnaire formation was based on the interviews taken of the experts and the literature survey done.

4.3 Validation of Questionnaire:

Validation of Questionnaire was done using Cronbach's factor for checking the consistency of data collected.

Following were the results observed in Table 3.

Table 1: Literature Survey Outcome:

S No	Issue No.	Issue	S No	Issue No.	Issue
International Experiences					
Operational			Strategic / Practice		
FO1	S1	Information sharing	FS1	S2	Strategic Planning
FO2	S4	Building Trust	FS2	S3	Career management
FO3	S5	Time to market, Speed , response time, reliability, security	FS3	S6	Information overload
FO4	C1	Lateral co-ordination	FS4	E1	Transfer of best practices
FO5	BP1	Transportation Cost	FS5	E2	Capturing expertise
FO6	BP2	Customer Service	FS6	E3	Train field representatives
FO7	BP6	Cost of reverse logistics	FS7	BP3	Outsourcing
FO8	E5	Minimize paper work	FS8	BP4	Many supplier issue
			FS9	BP5	Vendor selection
Indian Experiences					
Operational			Strategic / Practice		
IO1	I1	Time taken-transportation time	IS1	I2	Demand forecasting
IO2	I3	Depot to retail transport	IS2	I5	Global issues
IO3	I4	Idle stock	IS3	H3	Dynamic pricing
IO4	H1	Inventory holding	IS4	B1	KM on daily basis
IO5	H2	Sharing of products	IS5	B2	Sharing experiences
IO6	H4	Trust and collaboration	IS6	B3	Knowledge value

Summary: Total International = 17, Total Indian = 12, TOTAL= 29

Table 2: Pilot Survey Outcome

S No	Issue No.	Issue	Parameters (Measure)	Depot Manager	Retailer	Customer
International Experiences from Literature survey relevant as per survey						
FO1	S1	Information sharing	Level of sharing of information (LV=1, V=5)	Y	Y	Y
FO8	E5	Minimize paper work	Level of usage of ERP (LV=1, HV=5)	Y	Y	
FO8	E5	Minimize paper work	Use of electronic medium to record (LV=1, HV=5)	Y	Y	
FO7	BP6	Cost of reverse logistics	Goods returned(LV=1, HV=5)	Y	Y	
FS6	E3	Train field representatives	Quality of trainers (LV=1, HV=5)	Y	Y	
Indian Experiences from Literature survey relevant to India						
IS1	I2	Demand variability	Effect of demand variability on sales of goods (LV=1, HV=5)	Y	Y	Y

IO1	I1	GPS	Use of GPS (LV=1, HV=5)	Y	Y	Y
IO3	I4	Stocking of goods	Idle stock percentage (LV=1, HV=5)	Y	Y	
Additional Issues as per Pilot Survey relevant to India						
PSO 1	PS1	Availability of Goods	Rate the availability of goods in the stock (LV=1, HV=5)	Y	Y	Y
PSO 2	PS2	Customer Service	Frequency with which goods are provided to the customer (LV=1, HV=5)		Y	Y
PSO 3	PS3	Co-ordination between retailers and customers	Level of co-ordination between retailers and customers (LV=1, HV=5)		Y	Y
PSO 4	PS4	Co-ordination between retailers and depot managers	Level of co-ordination between retailers and depot managers (LV=1, HV=5)	Y	Y	
PSO 5	PS5	Demand Forecasting	Frequency of Goods ordered per week (LV=1, HV=5)	Y	Y	Y
PSO 6	PS6	Smartphone applications	Frequency for the use of smartphone applications (LV=1, HV=5)		Y	Y
PSO 7	PS7	Handheld devices	Frequency for the use of handheld devices (LV=1, HV=5)	Y	Y	
PSS1	PS8	Web portals	Frequency for the use of web portals (LV=1, HV=5)		Y	Y
PSO 8	PS9	Customer service	Frequency for the customer complaints resolved (LV=1, HV=5)		Y	Y
PSO 9	PS10	RFID	Use of RFID (LV=1, HV=5)	Y	Y	
PSO 10	PS11	Wireless sensors	Use of electronic wireless sensors (LV=1, HV=5)	Y	Y	
PSS2	PS12	On-time delivery	Speed of order fulfillment (LV=1, HV=5)	Y	Y	Y

Legend: P=Parameter, m =measure, lv = lower value, vh = higher value (m,v where applicable)

Table 4: Pilot Survey Summary

Issue category	Total Issues	Total Parameters	Depot Managers	Retail Managers	Customers
Operational	16	16	12	15	8
Strategic	3	3	3	4	3
TOTAL	19	19	15	19	11

4.5 Formation of Hypothesis:

Stating the Hypothesis

Null Hypothesis (H0): “The use of Knowledge Management in operationalizing collaborative decision making in the downstream Supply Chain Management of oil companies in western region of Indian operations in respect of critical factors is low” Alternate Hypothesis (H1): “The use of Knowledge Management in operationalizing collaborative decision making in the downstream Supply Chain Management of oil companies

in western region of Indian operations in respect of critical factors is adequate.”

4.6 Survey I Outcome:

Following is the table of issues and facts that will help us to identify the critical issues in SCM from an Indian perspective:

Table 4: Issues and facts to identify critical issues in SCM from Indian perspective

Sno	Issue	GMs		Depot Managers		Retailers		Customers		Correlation
		Criticality	Parameter Value	Criticality	Parameter Value	Criticality	Parameter Value	Criticality	Parameter Value	
FO 7	Goods Returned			H	1 good per day	H	1 good per day	H	1 good per day	0.85
FO 8	Level of Usage of ERP			H	Level: Low to Very Low	H	Level: Low to Very Low	H		0.87
FO 1	Level of sharing of information	H	Level: Low to Very Low	H	Level: Low to Very Low	H	Level: Low to Very Low	H	Level: Low to Very Low	0.90
FO 8	Use of electronic	H	Level: Low to	H	Level: Low to	H	Level: Low to	H	Level: Low to Very Low	0.93

	medium to record		Very Low		Very Low		Very Low			
IS1	Effect of Demand variability on sales of goods	H	Level: Low to Very Low	H	Level: Low to Very Low	H	Level: Low to Very Low			0.87
IO3	Stocking of goods			H	Level: High to Very High	H	Level: High to Very High			0.82
PS O1	Availability of Goods			H	Level: High to Very High	H	Level: High to Very High	H	Level: High to Very High	0.87
PS O5	Frequency with which goods are provided to the customer			H	Level: High to Very High	H	Level: High to Very High	H	Level: High to Very High	0.89
PS O3	Level of co-ordination between retailers and customers					H	Level: Low to Very Low	H	Level: Low to Very Low	0.8
PS O4	Level of Co-ordination between retailers and depot managers			H	Level: Low to Very Low	H	Level: Low to Very Low			0.83
PSS 1	Frequency for the use of web portals	H	Level: High to Very High	0.91						
PS O8	Frequency for the customer complaints resolved	H	Level: High to Very High			L	Level: Low to Very Low	L	Level: Low to Very Low	0.85
PS O9	Use of RFID	H	Level: High to Very High	H	Level: High to Very High	H	Level: High to Very High			0.83
PS O10	Use of Electronic wireless sensors	H	Level: High to Very High	H	Level: High to Very High	H	Level: High to Very High			0.82
PS O2	Speed of order fulfillment	H	Level: High to Very High	H	Level: High to Very High	H	Level: High to Very High			0.85
FS6	Quality of trainers	H	Level: High to Very High	H	Level: Low to Very Low	H	Level: Low to Very Low			0.77
PS O5	Frequency of goods ordered per week			M	Level: Low to Very Low	M	Level: Low to Very Low	M	Level: Low to Very Low	0.75
PS O6	Frequency for the use of smart phone applications	H	Level: High to Very High	M	Level: High to Very High	H	Level: High to Very High	M	Level: High to Very High	0.78
PS O7	Frequency for the use of handheld devices	H	Level: High to Very High	M	Level: High to Very High	M	Level: High to Very High	H	Level: High to Very High	0.67
IO1	Use of GPS			M	Level: Low to Very Low	M	Level: Low to Very Low			0.53

Table 6: Survey II Impact of KM on categorized issues and implementation levels:

S. No.	Issue / Parameter	Impact of KM on SCM					Implementation Levels			
		VH	H	M	L	VL	Tool	Proc	Practice	Overall *
FO7	Goods returned	26	40	5	24	5	N	Y	N	L
PS01	Availability of Goods	8	84	4	2	2	N	Y	N	L
PS02	Frequency with which the goods are provided to the customer	10	82	4	3	1	N	N	N	VL
PS03	Co-ordination between customers and retailers	14	64	4	12	6	Y	N	N	L
FO1	Sharing information	24	65	7	2	2	Y	Y	N	M
FO8	Use of electronic medium to record	15	62	15	5	3	Y	N	N	L
PS07	Frequency for the use of handheld systems	14	65	18	1	2	Y	Y	N	M
PS08	Frequency of the customer complaints resolved	4	78	14	2	2	Y	N	N	L
FS6	Quality of trainers	17	68	10	4	1	N	Y	Y	H
PS09	Use of RFID	1	76	18	4	1	Y	Y	N	M
IS1	Effect of Demand variability on sales of goods	14	67	16	2	1	Y	N	N	L
IO1	Use of GPS	2	81	11	4	2	Y	Y	N	M
PSO10	Use of electronic wireless sensors	1	93	3	2	1	Y	Y	N	M
IO3	Stocking of goods is costly	22	61	10	5	2	N	Y	N	L
PSS2	Speed of order fulfillment	11	71	11	5	2	N	N	N	VL
FO8	Use of ERP	1	59	7	30	3	Y	Y	Y	VH
PS04	Co-ordination between retailers and depot managers	24	61	7	7	0	N	Y	N	L
PS06	Frequency for the use of smartphones	32	51	3	4	10	Y	N	N	L
PSS1	Frequency for the use of web portals	1	67	25	4	2	N	Y	N	L

Table 5. Overall Implementation level

N	N	N	VL
---	---	---	----

Tool	Procedure	Practice	Overall *
30 %	30 %	40 %	100%
Y	Y	Y	VH
Y	N	Y	H
N	Y	Y	H
Y	Y	N	M
Y	N	N	L
N	Y	N	L

V.CONCLUSION

5.1 Conclusion from KM impact study:

Sno	Issue / Parameter (Category wise) Critical, Essential, Very Important, Important	Impact of KM on SCM(Average)	Implementation Levels (Average)	State of Impact / Implementation (type)
FO7	Goods returned	VH to H	L	Inadequate
PS01	Availability of Goods	VH to H	L	Inadequate
PS02	Frequency with which the goods are provided to the customer	VH to H	VL	Inadequate
PS03	Co-ordination between customers and retailers	VH to H	L	Inadequate
FO1	Sharing information	VH to H	M	Inadequate
FO8	Use of electronic medium to record	VH to H	L	Inadequate
PS07	Frequency for the use of handheld systems	VH to H	M	Inadequate
PS08	Frequency of the customer complaints resolved	VH to H	L	Inadequate
FS6	Quality of trainers	VH to H	H	Inadequate
PS09	Use of RFID	VH to H	M	Inadequate
IS1	Effect of Demand variability on sales of goods	VH to H	L	Inadequate
IO1	Use of GPS	VH to H	M	Inadequate
PSO 10	Use of electronic wireless sensors	VH to H	M	Inadequate

IO3	Stocking of goods is costly	VH to H	L	Inadequate
PSS2	Speed of order fulfillment	VH to H	VL	Inadequate
FO8	Use of ERP	VH to H	VH	Adequate
PS04	Co-ordination between retailers and depot managers	VH to H	L	Inadequate
PS06	Frequency for the use of smartphones	VH to H	L	Inadequate
PSS1	Frequency for the use of web portals	VH to H	L	Inadequate

Impact Implementation State Types: Table 6 (b)

Impact	Procedure	Type
VH, H	VH, H	Adequate
M	VH, H, M	Adequate
VH	M, L, VL	Inadequate
H	L, VL	Inadequate
M	VL	Inadequate
All other Combinations		Average

State of implementation KM in SCM in Downstream Side of Oil PSUs in Mumbai Region:

5.1.1 Overall Impact – Implementation:

$$\text{Inadequate Percentage} = \{(CI * 4) + (EI * 3) + (VI * 2) + (II)\} / (TT * 10) * 100 = 63.02\%$$

$$\text{Adequate Percentage} = \{(CA * 4) + (EA * 3) + (VA * 2) + (IA)\} / (TT * 10) * 100 = 1.579\%$$

$$\text{Satisfactory Percentage} = \{(CS * 4) + (ES * 3) + (VS * 2) + (IS)\} / (TT * 10) * 100 = 18.42\%$$

From the tables above, it is evident that KM usage in the industry for various factors considered as high impact KM factors, the implementation of KM in the organizations is not widely observed. Thus we can say that the KM implementation in Indian oil companies is low in terms of implementation tools, procedures and tools and procedures together.

REFERENCES

- [1] Efraim Turban, Dorothy Leidner, Ephraim McLean, James Wetherbe "Information Technology for Management: Transforming Organizations in the Digital Economy", 6th Edition, pgs.303-307.
- [2] Joines J.A., &Thoney, K, Kay M.G, "Supply chain multi-objective simulation optimization", Proceedings of the 4th International Industrial Simulation Conference., Palermo, pp. 125-132, 2008.
- [3] Ganeshan, R., Jack, E., Magazine, M.J., Stephens, P., "A taxonomic review of supply chain management research", Quantitative Models for Supply Chain Management. Kluwer Academic Publishers, Massachusetts, pp. 841-879, 1999.
- [4] "Optimization Engine for Inventory Control", white paper from Golden Embryo Technologies pvt. ltd., Maharastra, India, 2004.
- [5] Jinmei Liu, HuiGao, Jun Wang, "Air material inventory optimization model based on genetic algorithm, "Proceedings of the 3rd World Congress on Intelligent Control and Automation, vol.3, pp: 1903 - 1904, 2000.
- [6] C.M. Adams, "Inventory optimization techniques, system vs. item level inventory analysis," 2004 Annual Symposium RAMS - Reliability and Maintainability, pp: 55 - 60, 26-29, Jan, 2004.
- [7] E.W.T. Ngai, Li Xiu and D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification." *International journal of*

Table 7.

S n	Issue Category	Total No	Adeq uate	Inadeq uate	Average (Satisfac tory)
1	Critical	TC=15	CA=0	CI=15	CS=7.5
2	Essential	TE=2	EA=1	EI=1	ES=1
3	Very Important	TV=2	VA=0	VI=2	VS=1
4	Important	TI=0	IA=0	II=0	IS=0
	TOTAL	TT=19			

Expert Systems with Applications 36 (2009) 2592–2602 /ELSEVIER.

- [8] K. Balasubramanian, "Supply Chain Management in Oil Downstream Distribution Business: A perspective on IT Alternatives and Issues", White paper by Infosys.
- [9] David Evans, Shaun Bretstein, "Revolutionizing the downstream Supply Chain: Getting product to market more profitably using best-fit technology and processes", Wipro Council for Industry Research, Wipro.
- [10] Bruton, G., and M. White, "The Management of Technology and Innovation: A Strategic Approach", Vancouver, Canada: Thomson South-Western, 2007.
- [11] Debowski, S., "Knowledge Management", Sydney, Australia: Wiley, 2006.
- [12] Hartley, R., and J. Rowley, "Organizing Knowledge: An Introduction to Managing Access to Information", Hampshire, England: Ashgate Publishing, 2008.
- [13] Lambert, D.M., "Supply Chain Management: Processes, Partnerships, Performance", Florida, USA: Supply Chain Institute, 2008.
- [14] Rubenstein, A., "Managing technology in the decentralized firm", New York, USA: Wiley, 1989.
- [15] Chima, C.M., "Supply-Chain Management Issues in the Oil and Gas Industry", Journal of Business and Economics Research, (2007)
- [16] Crooks, E., "BP overtakes Shell in market capitalization", Financial Times, 11 January, 2010, A1.
- [17] Paige Leavitt, with contributions from Cynthia Raybourn and Cindy Hubert, Released October 2002
- [18] Sapient Corporation. "Brain Drain: Retaining Intellectual Capital in the Energy Industry", 2001.

- [19] Robert M. Grant, Department of Management Bocconi School of Management Bocconi University, “The Development of Knowledge Management in the Oil and Gas Industry”.
- [20] Shatina Saad, Zulkifli Mohamed Udin, Norlena Hasnan, “Dynamic Supply Chain Capabilities: A Case Study in Oil and Gas Industry”, Int. J Sup. Chain. Mgt, Vol. 3, No. 2, June 2014.
- [21] Christopher M. Chima, “Supply-Chain Management Issues In The Oil And Gas Industry”, Journal of Business & Economics Research – June 2007, Volume 5, Number 6.
- [22] Nancy C. Shaw, Mary J. Meixell, Francis D. Tugge, “A Case Study of Integrating Knowledge Management into the Supply Chain Management Process”, Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS’03) 0-7695-1874-5/03 \$17.00 © 2002 IEEE.
- [23] Arne Wiig, “Supply chain management in the oil industry: The Angolan case”
- [24] Elisa Kusrini, Subagyo and Nur Aini Masruroh, “Good Criteria for Supply Chain Performance Measurement”, Int J Eng Bus Manag, 2014, 6:9 | doi: 10.5772/58435
- [25] Adrian Drone, “Supply Chain Knowledge Management: A Conceptual Framework”, IESE Business School.
- [26] G.P. Kurien and M.N. Qureshi, “Study of performance measurement practices in supply chain management”, International Journal of Business, Management and Social Sciences, Vol. 2, No. 4, 2011, pp. 19-34.
- [27] Hamidreza Panjehfouladgaran, Rosnah Yusuff, Tang Sai Hong, Seyed Mahdi, Homayouni “Qualitative Performance Measurement of Supply Chain Management using Fuzzy Logic Controller”, The 11th Asia Pacific Industrial Engineering and Management Systems Conference, Melaka, 7 – 10 December 2010.
- [28] Ondieki John Nyamasege and Oteki Evans Biraori, “Effect Of Supplier Relationship Management On The Effectiveness Of Supply Chain Management In The Kenya Public Sector”, International Journal of Managing Value and Supply Chains (IJMVSC) Vol. 6, No. 1, March 2015.

Adaptive Automated Attendance Analysis for Education System

Kameshwar Gupta

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Akanksha Das

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Abstract— Adaptive Automated Attendance System is an advancement that has taken place by replacing the traditional marking attendance system. The proposed system aims at simplifying the attendance system and its management by face detection and recognition method. This system includes detection of faces through a high definition camera, the camera captures several faces of students within the class further these images are recognized by comparing them with the existing features in database thereby conducting the attendance. Existing algorithms for multiple object recognition like the Support vector machine, Trace transform, Independent component analysis and Naïve Bayes classification have multiple drawbacks such as extensive memory is required, lower speed, order of components is difficult to determine and loss of accuracy. The proposed system uses deep learning algorithm such as Convolutional neural network cascade which is used for face detection and Linear Discriminant Analysis is used for face recognition. This system is capable of detecting face in various angles, lighting conditions and has the ability to detect face even if there is an occlusion in the face up to some extent. Most of the existing systems are time consuming and require some kind of semi manual work from students or teachers. The proposed system aims at removing such work and providing better accuracy.

The benchmark parameter is the number of successfully detected faces, previous system could achieve this parameter up to 70% whereas proposed system achieves it up to 80%. The proposed system provides a detailed reporting system with more accurate results and also records the monthly attendance and semester attendance which simplifies the task of identifying the defaulters.

Keywords: LDA, Ada-Boost, cascading classifier, background regions.

I. INTRODUCTION

Face recognition is an essential field in many applications, one of which is Attendance Management System .The process of taking attendance of the student in the classroom is a tedious job for teachers like calling out the names waiting for response and also maintaining the attendance till the month to generate the report. Analyzing them manually can lead to wastage of time and thus in order to save time and improve the accuracy, we have used the concept of face recognition based attendance system. Thus face detection and recognition module detect the faces from the image captured by the camera, and the image

Pratikkumar Mishra

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

Bhushan Nemade

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai*

of the face is stored.

In Adaptive automated attendance analysis for education system, database of students with their corresponding faces along with different angles and variation is stored and trained over a period of time. All this can be done through various face detection techniques such as Viola-Jones and Convolutional Neural Network Cascade. Image of a class containing multiple faces is taken with the face being detected and recognized by comparing their extracted features in the database, thereby making the attendance of the student. The main objective of this project is to design and implement a large scale attendance system of educational institute or organizations. In this project we plan to design and implement a software application that allows college to detect and recognize upon database history. Further we try to scale our application to achieve high throughput and scalability.

II.RELATED THEORY

The proposed system aims at automating and simplifying attendance management system by using face recognition. The objective of the project is to design and implement a large scale attendance system of educational institutes or organization. the plan is to design. and implement a software application that allows the college to detect and recognize upon database history.

The system consists of a camera that captures the images of the classroom and sends it to the image enhancement module. After enhancement the image comes in the Face Detection and Recognition modules and then the attendance is marked on the database server. At the time of enrolment templates of face images of individual students are stored in the Face database, Here all the faces are detected from the input image and the algorithm compares them one by one with the face database. If any face is recognized the attendance is marked on the server from where anyone can access and use it for different purposes.

Algorithms for face detection

1. Viola-Jones face Detector

In Viola-Jones system a simple feature is used, with relation to the feature sets. Viola and Jones make note that the fact the choice of features instead of a statistical pixel based system is important due to the benefit of ad-hoc domain encoding. In the case of face detection this is particularly important. Features can be used to represent both the statistically close facial information and sparsely related background data in a sample image.

In its simplest form the features can be thought of as pixel intensity set evaluations. This is where the sum of the luminance of the pixels in the white region of the feature is subtracted from the sum of the luminance in the remaining gray section. This difference value is used as the feature value, and can be combined to form a weak hypothesis on regions of the image. Within the implementation four of the Haar-like features are chosen, the first with horizontal division, the second a vertical, the third containing two vertical divisions and the last containing both the horizontal and vertical division. The features are called Haar-like because of their resemblance to Haar-basis functions.

2. Neural Network-Based Face Detection

In neural network-based face detection system. A regionally connected neural network examines small windows of an image and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network.

It simplifies the difficult task of manually selecting nonface training examples, which must be chosen to span the entire space of non-face images. Simple heuristics, such as using the fact that faces rarely overlap in images, can further improve the accuracy.

3. Boxflow algorithm

The algorithms aims to further enhance the discriminative capability of face detectors in the deployment domain through unsupervised detector adaptation. BoxFlow – a new unsupervised detector adaptation method that can effectively adapt a face detector pre-trained static images to videos. BoxFlow unsupervisedly adapts face detectors through fully exploiting the motion contexts across video frames. In particular, BoxFlow introduces three novel components: Generalized heat map representation of face locations with augmented shape flexibility; Motion based temporal contextual regularization among adjacent frames for unsupervised face detection refinement.

A self-paced learning strategy that adapts face detectors from easy data samples to challenging ones progressively. The viola-jones face detector algorithm which is the best one is only giving 60% precision. The experiment conducted can be estimated on the basis of the values of precision and recall hence we are using convolutional neural network to improve the efficiency by 10%. And it will further make the system more efficient.

III. MOTIVATION

Attendance is very important for a student, a single absent adds a big difference in the performance of the college. Mostly students of high school are prone to absences, it is because of the laziness to attend the class, some students prefer going to computer shops playing games rather than entering the class and some student cannot refuse the influence of a friend inviting to go with them during class period. Some of this reasons are not reported to the parents or guardians because the way of informing them is the traditional way which is inviting the parents through students and the parent need come to college and communicate with the teacher about the absenteeism of the

student. Proposed system takes care of these flaws present in the manual based attendance system.

IV. PROBLEM DEFINITION

Most of the time of lecture or laboratory starts the lecturer or teaching assistant delays the lecture to record student's attendance. This is a lengthy process and takes lots of time and effort, especially if it is a lecture with huge number of students. It also causes a lot of disturbance and interruption when an exam is held. In this project, we intend to find the best model and set of features to design and implement a large scale attendance system of educational institutes using face recognition. In these project we plan to design and implement a software application that allows college to detect and recognize upon database history.

V. PROPOSED METHODOLOGY

The Proposed methodology is illustrated in Figure. It contains following steps: - Data Extraction (Image input of class), Data Preprocessing (Face detection), Data Integration and Transformation (Recognition and Comparison), Attendance Database

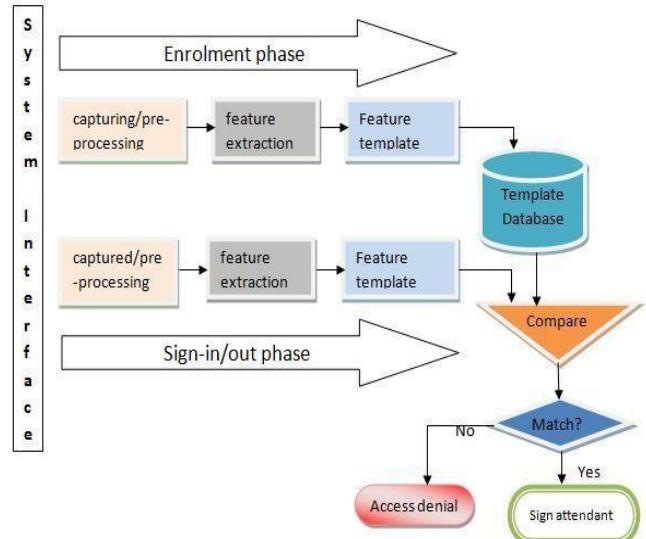


Fig. 1: Implementation Method Adaptive Automated Attendance Analysis for Education System

VI. SYSTEM ARCHITECTURE

Phases of our system are:

1. Image input of class
2. Face Detection using Viola Jones
3. Face Recognition using LDA
4. Feature Comparison
5. Face recognition using CNN
6. Attendance Database
7. Monthly Report
1. Image input of class

Image input of class will be taken from camera. This

image will consist of all the students present in class for that particular lecture.

2. Face Detection using Viola Jones algorithm

This algorithm is used to detect the face of each student. This algorithm consists of four phases integral image, Haar-like Features, AdaBoost, Cascading Classifier. In integral image we assign each pixel a value. And these values are converted by summing up all the pixel values present above and at the left side. Haar-like features consists of special pattern which is compared with image pixels and accordingly 1 or 0 is assigned. In AdaBoost only some necessary features are selected removing redundant ones. In Cascading classifiers weak classifiers are cascaded to make a strong classifier. The detailed method is explained in later section.

3. Face Recognition using Linear Discriminant Analysis

In this algorithm image is divided into several parts and on each part LDA is applied. Initially 3×3 top left corner window is taken and each pixel is converted into equivalent grey level. After this centre pixel value is compared with all the neighborhood values and if centre pixel value is greater than surrounding pixel then it is 1 else 0. After this, the 8-bit code is converted into decimal equivalent and this procedure is repeated for the whole image. After this all the decimal values are plotted on histogram.

4. Feature Comparison

In this phase the calculated features are compared with stored features. In this for testing we have used SQL database. This includes variation in gestures, luminosity, expression, surrounding, race, ethnicity, age, camera quality, color saturation, focus, and other parameters.

5. Face Recognition Using Convolutional Neural Network

If the results are still not matched we then use the Convolutional neural networks. Convolutional neural networks (CNNs) have very powerful discriminative capability, while maintaining high performance. The proposed CNN cascade operates at multiple resolutions, quickly rejects the background regions in the fast low resolution stages, and carefully evaluates a small number of challenging candidates in the last high resolution stage.

6. Attendance Database

If features match in comparison phase, then attendance for particular student is registered in the database. This database is stored in the backend of the system and the results drawn are all entered in the same.

7. Monthly Report

At the end of each month monthly report will be generated and will be mailed to HOD of department. The report can be generated as per the institute norms. This reduces or finishes off the entire work of the faculties of the institute. Any alterations required to be done in the final report can be done by the respective authorities.

In this way, the proposed system does the work of attendance marking using face detection and face recognition respectively and generates the final report as per the required institute norms.

VII. EXPECTED RESULTS

Outputs

The system will be able to detect faces through a high definition camera, the camera captures several faces of students within the class further these images are recognized by comparing them with the existing features in database thereby conducting the attendance. The system detects and recognizes up to 80% of faces of students present in the classroom thereby recording the attendance.

Outcomes

The application will be able to predict the identity of the student. The application will also reduce the time required for the recognition. By using this application will easily able to take attendance in schools, colleges, or in any organization.

Phase 1 (Information Gathering)

The identity of students will be fetched from the system that stores 400 images in the database and the image inserted for processing using comparison algorithms.

Phase 2 (Analysis & Design)

Using algorithms such as Deep Learning Neural Network the accuracy of precision and recalling can be achieve up to 80%.

Phase 3 (Implementation & Testing)

GUI for the project will be made successfully and basic requirement of graphical design like graphs, pie chart will be implemented using the attendance of the students.

Phase 4 (Deployment)

Finally phase will be integration of the preprocessed data with the GUI and the ranking of the software. Along with this the testing of the whole application will be done and try to get the accuracy up to 90% with the graphical representation.

Percentage improvement in results

Existing algorithms for multiple object recognition like the Support vector machine, Trace transform, Independent component analysis and Naïve Bayes classification have multiple drawbacks such as extensive memory is required, lower speed, order of components is difficult to determine and loss of accuracy. In the proposed system deep learning algorithm such as convolutional neural network cascade is used for face detection and Linear Discriminant Analysis is used for face recognition. This system is capable of detecting face in various angles, lighting conditions and has the ability to detect face even if there is an occlusion in the face up to some extent. Previous system could achieve accuracy up to 70% whereas the current system increases its accuracy up to 80%.

VIII. CONCLUSION

In order to reduce the faculty effort and to manage the time effectively the authors proposed automated attendance system base on face recognition in schools/colleges. In this system we have implemented an attendance system for a lecture, section or laboratory by which lecturer or teaching assistant an record student's attendance. It saves time and effort,

especially if it is a lecture with huge number of students. Most of the existing systems are time consuming and require for a semi manual work from the teacher or students. This approach aims to solve the issues by integrating face recognition in the process. Even though this system still lacks the ability to identify each student present on class, there is still much more room for improvement. Since we implement a modular approach we can improve different modules until we reach an acceptable detection and identification rate. Another issue that has to be taken in consideration in the future is a method to ensure users privacy. Whenever an image is stored on our servers, it must be impossible for a person to use that image. Computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

REFERENCES

- [1] B.-C. Chen, Y.-Y. Chen, Y.-H. Kuo, T. D. Ngo, D.-D. Le, S. Satoh, and W. H. Hsu. Scalable face track retrieval in video archives using bag-of-faces sparse representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016
- [2] F. Comaschi, S. Stuijk, T. Basten, and H. Corporaal. Online multiface detection and tracking using detector confidence and structured SVMs. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, 2015.
- [3] S. Xiao, S. Yan, and A. A. Kassim, “Facial landmark detection via progressive initialization,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 33–40

College Management System and Data Analysis with Machine Learning

Mary Margar

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai
e.marymargret@gmail.com*

Raj Asha

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai
rajasaha1711@gmail.com*

Abstract—In the process of college management, there are lot of challenges that are faced by the teachers and staff such as informing all students about some important event or taking attendance accurately. Also, there is no central storage where all student data is stored and organized. Also, there is no way to analyze the student data and take important decision based on the analysis. With the help of our College Management System, lot of benefits can be achieved by the college. First of all, student information will be stored in a single location and organized in a way which helps to access data easily. The process of taking attendance will be quick as well as done accurately and the students could also view their attendance at any time. Also, time of the teacher will be saved by easily posting notices to students as and when needed directly from his/her laptop or PC. Therefore, with the help of the college management system lot of time of teachers will be saved as well as data stored will be more accurate than before. We have created the user interface for the user to easily interact with the system. The system provides web pages with very simple forms for students and teachers. The database has been created with the tables linked together for easy access of data from the tables. Attendance module for taking and viewing attendance is also created along with the notice module for posting notices. Also, the profile page for the students and teachers has been created which provides general information about students and teachers.

I. INTRODUCTION

College Management System (CMS) provides almost all of the important features required in the college management. It helps in carrying out major functions for day-to-day college life. It provides a different teacher, student and staff login. Each of them would have different access rights. It also provides a facility to take and view attendance as well as provide a facility of notice board. Students and teachers can communicate with a messaging module. A module is been used to analyze the student's performance based on overall attendance, pointers, co-curricular and extra-curricular activities. Any person can store and edit the profile. Notice can either be college related or specific department or class related.

Meet Gadoya

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai
meetgadoya996@gmail.com*

Saurabh Maydeo

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai
saurabhmaydeo@gmail.com*

The traditional college management is a very tiring and time-consuming process. So, the basic motivation was to insert all of the college management functionalities into a single portal with better usability and efficiency.

In traditional management, teachers have to carry large book for attendance and manually find out the appropriate page for the class and the subject which could be a hectic task.

Here the notices are being displayed on notice board which has limited space. If there are multiple notice than some recent notice is been removed and student could be unaware of it.

In traditional management, there is no specific and proper mechanism to analyze a particular student performance. There is also no message functionality available.

All the college functions which are distributed in traditional approach are been combined into a single portal for better functionality.

II. LITERATURE SURVEY

We went through many previously done works related to this topic.

Paper published in International Journal of Advanced Research in Computer Engineering & Technology on April 2016 with title "Android College Management System" by author Vishwakarma Ganesh was useful. The major key findings were this system use data mining for remedial students and deals with IMEI number for better security.

Paper published in International Research Journal of Engineering & Technology on May 2016 with title "College Management System" by authors ShrikantPatnaik, KhushbooKumari Singh, RashmiRanjan, NikiKumari. The major key findings were this system also has hostel management.

Paper published in International Journal of Advanced Research in Computer & Communication Engineering on June 2013 with title "Student Information Management System" by authors S. R. Bharamagoudar, Geeta R. B, S. G. Totad was useful. The major key findings were this system can represent progress of students. Paper published in International Journal of Computer Applications on July 2015 with title "College

Management System" by author Lalit Joshi was also helpful. The major key findings were, this system provide facilities such as attendance of teachers, college feedback.

III. PROPOSED SYSTEM

College Management System would help to carry out college functionalities properly and all activities under single portal will make it easier to use. It helps is better attendance viewing function and various functions can be applied on the attendance database to find defaulters. It has proper module to analyze a student performance and provide some help to improve the overall performance. For analyzing, classification algorithms are being used. It requires active internet connection for working and the information about teachers, students and staff is been stored in the database. Users can communicate with others using a message module provided.

IV. METHODOLOGY

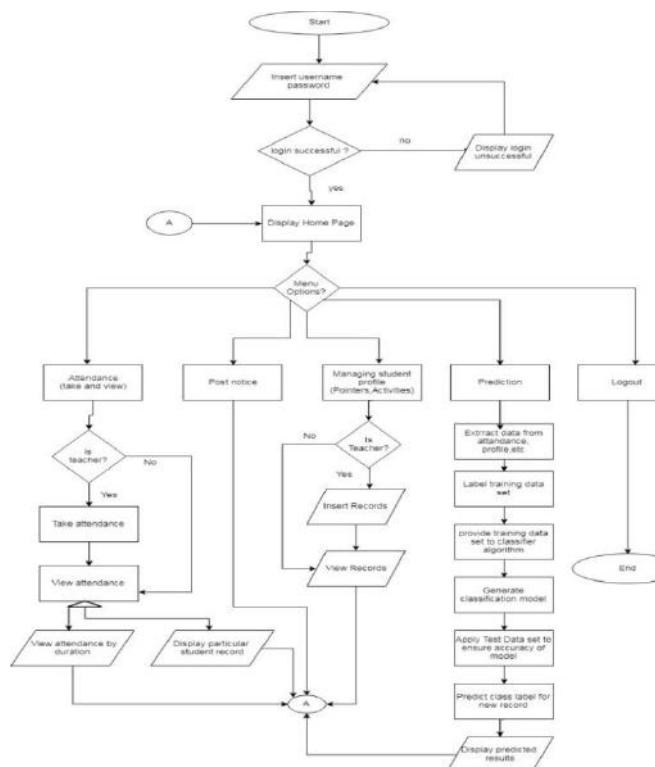


Fig 1. Flow Chart for the proposed system

Proposed system's complete work flow is shown in the fig 1 for College management System and Data analysis using machine learning algorithm will allow college to analyze the data of students and make various inferences from the data to take better decisions. Classify data using classification algorithms in order to classify new data to make predictions which has better probability

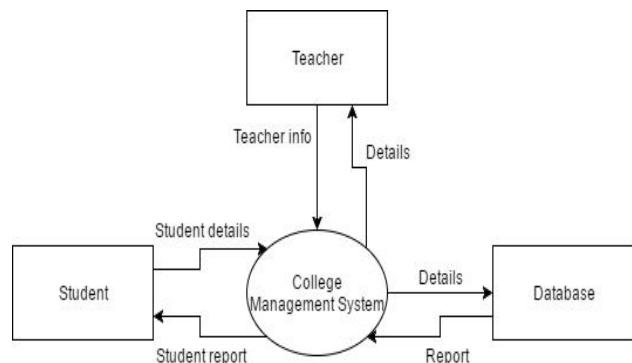


Fig.2. Block Diagram of the proposed system

As shown in the fig 2, Teacher can post notices to a class or a group of classes to inform about some important event or task that is to be done by the students. Teacher simply has to write a title, message and select the classes to which he/she wants to send notice. Only the selected class students will receive the notice.

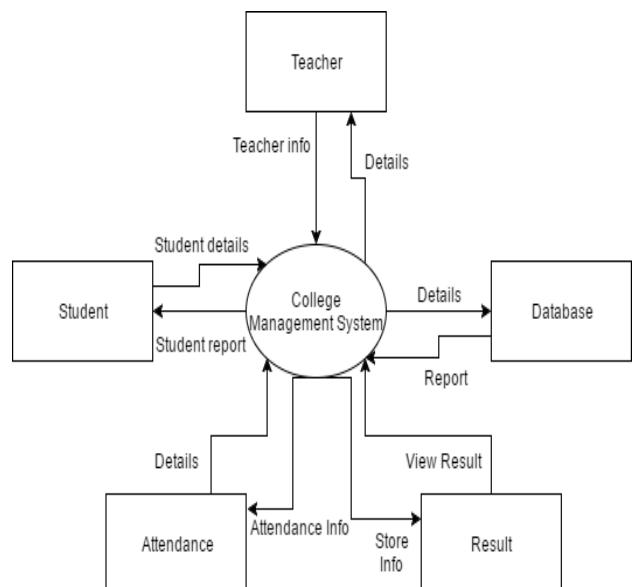


Fig 3-Block Diagram

Student's result can be stored and viewed by teachers and students. Term test 1 and 2 as well as the semester results of each student will be stored in the student's profile page in textual as well as visual form such as bar graph.

FEATURES OF THE PROJECT:

1. Authentication for students and teachers through a login system. So, users have to enter their username and password that is stored in the database.
2. Teachers can take attendance of students through a very simple user interface. Teacher have to select from multiple drop-down lists, the class, date, time, subject to display list of

students to which he/she can give attendance by selecting a checkbox.

3. Teachers and students can view attendance through a simple interface by simply selecting roll number of students or name of student. Attendance of student can be displayed in two views, day-to-day attendance and aggregate attendance of a selected date range.

4. Teacher can post notices to a class or a group of classes to inform about some important event or task that is to be done by the students. Teacher simply has to write a title, message and select the classes to which he/she wants to send notice. Only the selected class students will receive the notice.

5. Student's result can be stored and viewed by teachers and students. Term test 1 and 2 as well as the semester results of each student will be stored in the student's profile page in textual as well as visual form such as bar graph.

6. Data analysis using machine learning algorithm will allow college to analyze the data of students and make various inferences from the data to take better decisions.

7. Classify data using classification algorithms in order to classify new data to make predictions which has better probability

V.DESIGN/RESULTS

The screenshot shows a login form titled "Login". It contains fields for "Username" and "Password", both with placeholder text "Enter Username" and "Enter Password" respectively. Below these is a "Type" section with radio buttons for "Student", "Teacher", and "Staff". At the bottom is a "Submit" button.

Fig: 4: Login page

Authentication for students and teachers through a login system as shown in Fig: 4. So, users have to enter their username and password that is stored in the database.

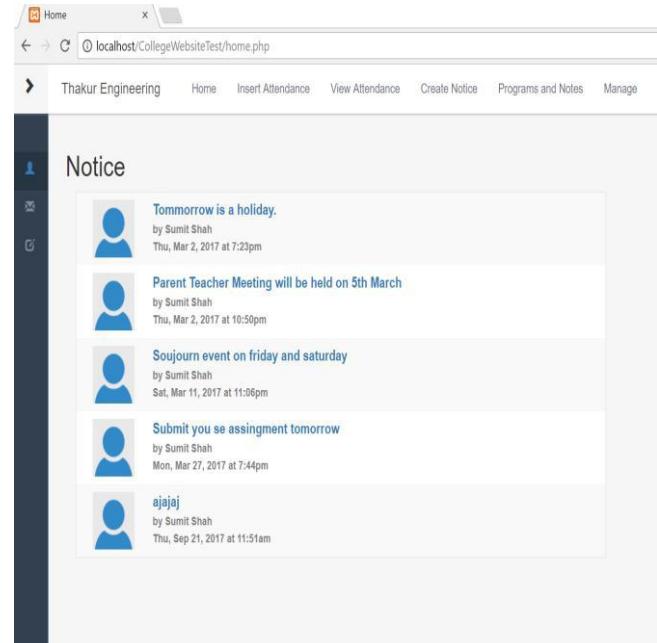


Fig: 5: Home page

Teacher can post notices to a class or a group of classes to inform about some important event or task that is to be done by the students. Teacher simply has to write a title, message and select the classes to which he/she wants to send notice. Only the selected class students will receive the notice.

The screenshot shows an "Attendance" page for the class "TEIT-B". It includes a search bar for "Class: TEIT-B", "Roll No.", "Name: Enter Name", "Date: 01-09-2017", and "To Date: 26-10-2017". Below this is a navigation bar with "Daily" and "Overall" tabs, where "Overall" is selected. A table displays student attendance data:

Class	Roll No.	Name	AIT			DMBI		
			P	T	%	P	T	%
TEIT-B	1	Salman Khan	0	2	0%	0	2	0%
TEIT-B	2	Sharukh Khan	1	2	50%	1	2	50%
TEIT-B	3	Raj Ashaa	0	2	0%	0	2	0%
TEIT-B	4	Amir Khan	0	2	0%	0	2	0%
TEIT-B	68	Saurabh Maydeo	1	2	50%	1	2	50%
TEIT-B	69	Raj Asha	2	2	100%	1	2	50%
TEIT-B	72	Tejas Mahajan	1	2	50%	1	2	50%
TEIT-B	88	Randy	1	2	50%	1	2	50%
TEIT-B	90	Rakesh Roshan	0	2	0%	0	2	0%

Fig: 6: Attendance page

Teachers and students can view attendance through a simple interface by simply selecting roll number of students or name of student. Attendance of student can be displayed in two views, day-to-day attendance and aggregate attendance of a selected date range.

VI.CONCLUSION

Colleges Management is a very crucial component for a college and it plays a very important role for the smooth functioning of the college and allows the college to manage things such as attendance of the students, informing students about important events and notices as well as communication between students and teachers etc. Therefore, for a smooth management of the college as well as ease of work and communication between students and teachers, college management system will provide a wide range of features and functionalities such as taking of attendance by teachers, viewing attendance by teachers and students, posting notices by teachers, sending of private messages between students and teachers, viewing results of students etc. Data Analysis component which includes prediction of pointer of students, classification of students into various grades as well as clustering of data allows the college to analyze the data and make various decisions for the students

VII. FUTURE SCOPE

The college management system can be further enhanced for carrying out almost all activities related to college. It can be

used to keep track of all submissions of the students such as assignment, experiments and remedial submission.

A dashboard can be used by the teacher to get the overall idea of the progress of each of the student from the student's attendance, marks, pointer, and submission status along with the student's participation in co-curricular and extra-curricular activities.

The entire website can have a mobile version in the form of an application which can be used by the students and teachers. This will provide mobility and ease of access to information. Students can also get notification on their mobile based on events like when a notice is posted or his/her attendance is taken.

VII.REFERENCES

- [1] Tom M. Mitchell "Machine Learning", McGraw-Hill, 1997 ISBN 0071154671, 9780071154673
- [2] W. Jason Gilmore "Beginning PHP and MySQL: From Novice to Professional" PublisherDreamtech Press; Fourth edition ISBN-10: 8184897456 ISBN-13: 978-8184897456
- [3] Sadeg BafandehImandoust And Mohammad Bolandraftar "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background"ijeraVol. 3, Issue 5, Sep-Oct 2013, pp.605-61.
- [4] SankarRajagopal "CUSTOMER DATA CLUSTERING USING DATA MINING TECHNIQUE" IJDMS, vol.3, no.4, 2011.

English To Hindi Machine Translation

Ankita Gaud

Department of Information Technology,
Thakur College of Engineering &Technology,
Mumbai

Akash Naik

Department of Information Technology,
Thakur College of Engineering &Technology,
Mumbai

Chanda Jha

Department of Information Technology,
Thakur College of Engineering &Technology,
Mumbai

Shruti Mathur

Department of Information Technology,
Thakur College of Engineering &Technology,
Mumbai

Abstract- The idea of a language translator is to foster communication between people coming from various linguistic backgrounds. In order to understand the culture or say any message from any person, we need to know what language he is speaking. Having found out the language, we then look for the meaning of the speech he has used. Moreover, one needs to understand the grammatical structure of a language, and its sentence formation. Having an idea about this once can use, says a "Translator" to convert the speech/text into a language he is familiar with. This obviously may sound a tedious process, but having said that, it has numerous applications and benefits.

Key Words – Natural Language Processing, Rule Based Statistical Learning, Cross validation.

I. INTRODUCTION

This Software Requirements Specification provides a complete description of all the functions and specifications of the language translation application. The expected audience of this document is the employees of a company or users who will use this system. It will also serve as an automated tool to generate translation results. The Language Translation Application is designed to generate translation and detection results of languages to professionals as well as an average user. Apart from translation results various other services is provided so as to give ease of usage to the user of this application.

II. LITERATURE SURVEY

In this section we now look at English-Hindi language MT project. The parameters we look at are: language pair, approaches used for handling problems in MT system. The scope of this paper is restricted to Hindi and English language as source/target language.

The approach for the current system to be developed is studied through the [4] paper and the project will have 2 target and source languages i.e. Hindi and English. The basic idea of the project is defined by reading paper [4].

The translation process will contain folding technique which is greater than the reference paper [1], where the sentences with more than 60 words can be translated easily and efficiently. The rare and unknown words will be easily translated using 10-fold cross validation approach as reference to paper [2].

The system will provide accuracy up to 66% taking reference [3] as the base improvement is to be done to get targeted accuracy in the system.

The current system will contain speech detection the reference is taken through paper.

III. PROBLEM DEFINITION

The project is based on the motive of translating complex sentences from English into targeted language i.e. Hindi. The application build will be using neural networking. The most important distinguishing feature of this approach from the basic encoder-decoder is that it does not attempt to encode a whole input sentence into a single fixed-length vector. Instead, it encodes the input sentence into a sequence of vectors and chooses a subset of these vectors adaptively while decoding the translation. This frees a neural translation model from having to squash all the information of a source sentence, regardless of its length, into a fixed-length vector. We show this allows a model to cope better with long sentences.

PHASE 1: (Planning, Analysis, Design, Coding)

- a. Planning – This phase involves the analyzing of different strategies and defines the suitable strategy for the application.
- b. Analysis – Critical study, analysis and review of feasibility for proposed solution and an estimate of time for each phase.
- c. Design – The application will be able to access with the dataset to provide efficient translation.
- d. Coding – Feed-forward application will be used and coded in java language.

PHASE 2: (Integration, Testing, Deployment)

- a. Integration – Integration of dataset with the application will be done for Human Computer interaction.
- b. Testing – Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.
- c. Deployment – Give the completed prototype to available server for security purpose in the active servers.

IV. PROPOSED WORK

On basic level, MT performs simple substitution of words in one language for words in another, but that alone usually cannot

produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed. Solving this problem with corpus statistical, and neural techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies.

Importance of the project:

The world is no longer the huge far-stretched globe it was once thought to be. Communication has broken all distance barriers, which once prohibited cultural and linguistic diversification. A person can now explore unknown lands, talk to various people from various parts of the world; understand their dialect and their culture. All this with just the idea of a something called the language translator.

Scope of the project:

The translation of rare and unknown words by increasing the accuracy and number of targeted languages by enhancing the dictionary. Use algorithm efficiently to reduce the compilation time and increases the efficiency.

V. FEASIBILITY STUDY

1. Technical Feasibility - The English to Hindi Machine Translation would be implemented using Python and MySQL which are proven and tested mature technologies to deliver the proposed solution.

2. Economic Feasibility - Development of this application is highly economically feasible. The organization needed not spend much one for the development of the system already available. The tools to be used are open source and free of cost. The only thing is to be done is making an environment for the development with an effective supervision. If we are doing so, we can attain the maximum usability of the corresponding resources. Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

3. Legal Feasibility-The project is will be developed under legal license of python and MySQL as both are considered as open source. The data will not be used for some illegal purpose and will be encrypted using some algorithms.

4. Operational Feasibility-An overview of the process portrayed that the proposed Machine Translation project will be beneficial in faster translation of sentences and enhance time complexity, space complexity and implementation of code.

5. Social Feasibility- Easy in translation between two languages and faster translation within less time with translation of rare and unknown words.

VI. METHODOLOGY USED

Methodology to be used for building the project will be Spiral Software Methodology as the project will be needed as it focuses on the past work of the project and decreases the risk rate of the project. As cost and risk evaluation is important for the project. And in future multiple and new additional features can be added. Additional to it have strong approval and documentation control.

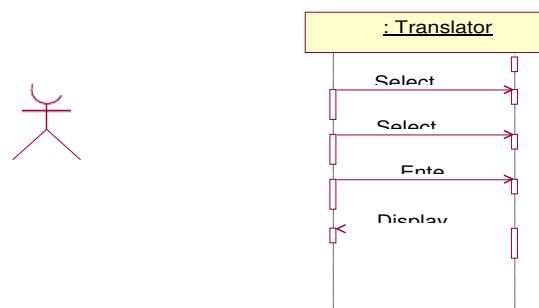


Figure 1: Sequence Diagram

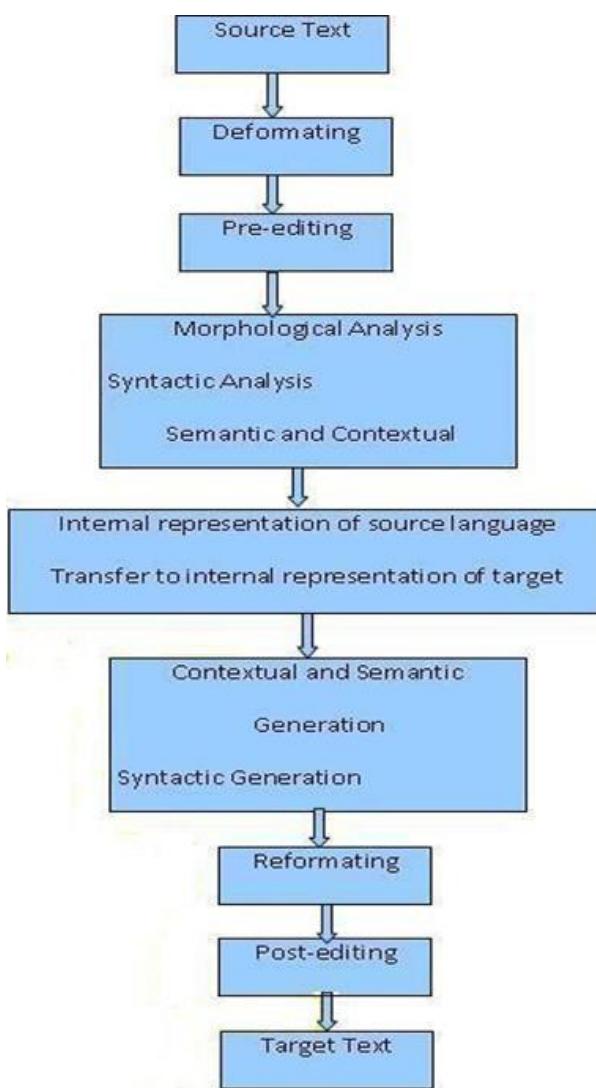


Fig: 2 Sequence Diagram

VII. OUTCOMES

Phase 1.1 Planning

This phase the outcome is to study the problem definition and define the scope of the project.

Phase 1.2 Analysis

The outcome is the analysis done by studying the resource paper and other resources and the work to be done phases wise is decided. The time estimation and feasibility study with respect to project is done.

Phase 1.3 Design

The graphical user interface is designed which is a html page in which the user will write the sentence to be translated.

REFERENCES:

- [1] Nadeem Jadoon Khan, Waqas Anwar, Nadir Durrani; “Machine Translation Approaches and Survey for Indian Languages”, COMSAT University,University of Edinburgh, 2014.
- [2] Dzmitry Bahdanau, KyungHyun Cho Yashua Bengo, “Neural Machine Translation by Jointly Learning to Align and Translate”, “Conference Paper at ICLR 2015”, 2016.
- [3] Yonghui Wu; Mike Schuster; Zhifeng Chen, Quoc V. Le; Mohammad Norouzi, “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”, 2016.
- [4] Amruta Godase and Sharvari Govilkar, “Machine Translation Development for Indian Languages and its Approaches” ,“International Journal on Natural Language Computing (IJNLC) Vol. 4”, No.2, 2015.

Smart Warehouse – Reduction in Post-harvest Losses

Aishwarya Nair

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India*

Vikram Deshmukh

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India*

Anand Ranka

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India*

Kamal Shah

*Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India*

Abstract -Storage systems and smart warehouses are the need of the hour as various technologies have helped us increase the production but these goods tend to spoil due to lack of proper infrastructure and management for their storage. This leads to a lot of wastage of food that can be provided to the needy and even the farmers do not receive the right compensation for the same. Thus, this project will help minimize loss of food grains and help farmers in their development which will eventually lead to development of the society as there is a dependency between agricultural growth and economic prosperity. The proposed solution is to make a traditional warehouse a smart warehouse with IOT enable sensors and intelligent monitoring to avoid quality losses and replacing the current LIFO order of storage to FIFO to reduce the total shelf time of the grains. In the process of grain storage, temperature, humidity and carbon dioxide concentration (CO₂) are major ecological factors that can influence directly on the quality of the Grain. Hence, there is a necessity to monitor the vital parameters continuously during storage and communicate the status to the manager in real time which becomes a challenging. This project implements the system which will enable us to record data and use the data to dispatch grains in a FIFO manner and along with it, implement a structure which will deal with space confinement issues by following a vertical architecture of storage, keep the relevant farmer and grain records, and keep the farmers updated about the progress status of their yield.

Keywords –

Smart Warehouse, LIFO, Grain, Farmers, storage

I. INTRODUCTION

India is an agricultural country and agriculture contributes to about 16% of total GDP and 10% of total export. Thus, it is significant that proper advancements be made in this field but one phase that lags behind is storage of agricultural products

Storage systems and smart warehouses are the need of the hour as various technologies have helped us increase the production but these goods tend to spoil due to lack of proper infrastructure and management for their storage. India's grain production has steadily increased due to advances in technology, but post-harvest loss is constant at

10%. Losses during storage, accounts for around 6% of the total losses as proper storage facilities are not available. In India, food grains are stored using traditional structures by small farmers. The surplus grains are stored with government agencies like: Food Corporation of India (FCI), Central and State warehousing Corporations. We propose a system which has a touch of IOT to automate the monitoring, storage and record maintenance.

II. OVERVIEW

A new wave of connected devices, also known as The Internet of Things, is poised to grant warehouses a heightened level of visibility into every facet of the warehouse. IoT technologies offer the promise of facilitating real-time workforce interaction to elevate productivity, while adding a new level of precision. When it comes to the application of IoT technology ,once cost prohibitive and now significantly more affordable, will play a critical role in modernizing warehouses.

III. MOTIVATION

The real motivation of our project is to help the farmers and the Food Corporation of India (FCI) by providing them with a storage structure which fulfils their needs and demands on today's advanced Technologies. The goal is to minimize the loss of food grains and other agricultural products incurred by the farmer and provide them with a sense of security for their efforts.This system will play a vital role in the development of the storage of smart and automated warehouse.

IV. LITERATURE SURVEY

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they can move forward.

1. The solution found is Distribution chain follows First-in-first-out and first-expired-first-out . The key findings are Distribution chain follows First-in-first-out and first-expired-first-out, Loss of quality taking into account the combined effects of temperature, O₂, CO₂ and relative humidity by equation :

$$= - \sum_{i=1}^{n-1} K_i$$

And the Research gap found are High implementation complexity, therefore not feasible for implementation of bulk storage of grains.[1]

2. The solution found is Surface space required:

3 cars System can accommodate : more than 9 cars. Chain and sprocket mechanism for rotation to minimize vibration and noise. Operation : Key pressing. Thus, flexible and does not require care taker .

The key findings are Surface space required: 3 cars System can accommodate: more than 9 cars. Chain and sprocket mechanism for rotation to minimize vibration and noise Operation : Key pressing. Thus, flexible and does not require care taker. and the Research Gap found is . It can be fully automated by integrating it with a panel Board..More secured by providing each platform a specific password, so that only whenever a particular password is typed the platform is retrieved.[2]

3. The solution found is .NODE 1: GPS mobile robot for weeding, spraying, keep vigilance

NODE 2: Motion detector, light sensor, humidity sensor, temperature sensor, room heater, cooling fan altogether interfaced with AVR microcontroller.

NODE 3 :Smart irrigation node

The Key findings are Setup for node1 consists of mobile robot with central server, GPS module, camera &other sensors interfaced with microcontroller which is interfaced with the raspberry pi. And the Research Gap found is Storage management, Retrieval management [3]

V. PROBLEM DEFINITION

The project focuses on making agriculture storage smart using automation and IOT technologies which can provide a probable solution for hunger. This we plan to achieve by fabrication, assembly and installation of a conveyorised system which will place the gunnies onto the mechanical lift based on vertical car parking system, count the gunnies and maintain two databases about the farmers and the jute bags for their proper management , storage , record maintenance and dispatch and also monitor the environmental effects on the grains and alert the respective authorities incase of need The further work was done as follows.

PHASE 1: (Planning, Analysis, Design, Coding)

- Planning – Conceptualizing and designing the architecture for proposed solution. Studying the feasibility and efficiency in the real market. To make use of MySQL Workbench, Amazon Web services, HSM20G humidity sensor, LM35 temperature sensor and MQ9 series CO2 concentration detector Sensor
- Analysis – Critical study, analysis and review of feasibility for proposed solution. If any smoke/Gas detected OR Temperature goes beyond threshold OR humidity goes beyond threshold or CO2 concentration goes beyond Threshold activates controller to take control actions
- Design – Hardware setup is designed using the components as described in planning phase and the proposed system will be developed

PHASE 2: (Integration, Testing, Deployment)

- Integration – Integrating of various proposed modules such as recognizing face module whether it detects face of the person, Android application modules.

- Testing –Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.
- Deployment – Give the completed prototype to Automobile industry for security purpose in the car.

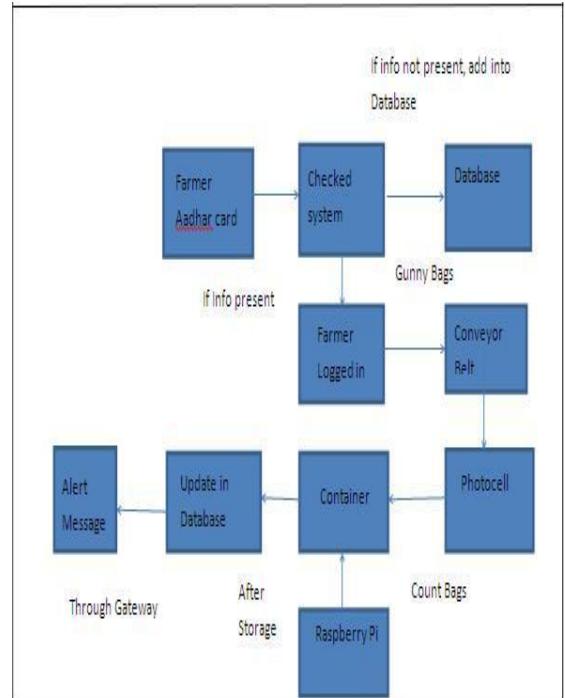


Fig 1: Block Diagram

VI. METHODOLOGY

The proposed solution is a structure wherein first the barcode scanner will read the card. After reading it will compare the values with the database. This will trigger two components – database and storage structure. The storage structure will be allotted to farmer. A conveyor belt system with a laser sensor will be attached to the storage containers which will count the number of sacks that have been delivered by keeping a count of the number of times the laser is cut and update the database. The database maintained will be in the ascending order of the date of expiry of the container and the dispatch of grains will be guided by this ,thus ensuring minimum shelf time and reducing quality loss.

Once the containers are filled with the sacks for storage the Temperature sensor, Humidity sensor and CO2 sensor connected to the Controller continuously sense the environment around the grain. The Sensors collect the information from environment, the collected signals through the analog to digital conversions since the signals sensed are analog in nature. Sensors also send the trigger pulse to corresponding pins of LPC2148. IF any smoke/Gas detected OR Temperature goes beyond threshold OR humidity goes beyond threshold or CO2 concentration goes beyond Threshold activates controller to take control actions. LPC2148 sends the AT commands to GSM/GPRS module to send alert message. Embedded Application developed in the administrator PC Stores parameter valves in the cloud and display graph for the period of monitoring which helps to take necessary action in Future.

Spiral model of development will be used since there are lots of risks to consider.

Hardware:**a) Temperature sensor :**

Temperature sensors LM35 is chosen since LM35 provides reliable accurate surrounding temperature directly in °C according output voltage. These sensors are analog in nature which senses surrounding temperature produces output voltage in mV proportional to surrounding temperature, so it is easy to take accurate readings by connecting to development boards. It can measure temperature ranges from -55°C to 150 °C.

b) Humidity sensor :

Humidity sensor HSM20G humidity sensor modules are chosen and they are configured with IC's circuitry to provide onchip signal conditioning. Absorption-based humidity sensors provide both Relative Humidity(%RH) and temperature outputs. Onchip signal processing ensures Relative Humidity output versus linear voltage. The Sensor laser trimming offering +5%RH accuracy and attains 2%RH accuracy with calibration. Packages are chemical resistant and operating in range of -40 °C to 85 °C [-40°F to 185°F] to accommodate erratic environment.

c) CO₂ concentration sensor

CO₂ concentration sensor or Gas sensor MQ series gas sensors are used to detect the CO₂ concentration level. These are Electrochemical innature having gas detector that can measures the concentration of a targeted gas by oxidization or reduction of the targeted gas at an electrode and measuring the resulting current. The various gas sensor modules are used now a days to detect different types of gases on such MQ-series gas sensor detects the presence Propane (LPG), Carbon Dioxide (CO₂), Methane (CH₄).

d) GSM/GPRS Module:

This GSM/GPRS Modem NeoWay M660 are simple to interface serial interface. It features with voice, SMS, and Data services. GPRS data related operations can be controlled with the help of ATtension commands from the PC and also from controllers. This module contains the highly popular SIM300 module inside it which holds sim card meant for all its data, voice related operations. This module come with familiar widely used standard RS232 interface, so these modules can be easily interfaced to PC and controllers.

VII. CONCLUSION

The expected outcomes as related to the survey objectives will include the following:

[1] The status of warehouses in terms of both capacity and quality assessed. Warehouses locations, owners and related market participants identified. General status and warehouse conditions assessed. Types of food and non-food materials stored/could be stored identified. Effective warehouse management procedures for food warehouses recommended.

[2] Process streamlining and increasing efficiency by integrated view of end processes, use of workflow, notification and alerts. Better coordination between warehouses and district offices of Area through integrated system.

[3] Maintain inventory of dead stock articles, handing taking over charge and charge report of every employee posted at the warehouse.

[4] Reduce redundancy in processes by capturing data at the source and in real time.

An Intelligent College Enquiry Voice Chatbot

Vivek Kumar

*Information Technology Department,
Thakur College of Engineering & Technology,
Mumbai , India*

Vedprakash Jha

*Information Technology Department,
Thakur College of Engineering & Technology,
Mumbai , India*

Bharat Paliwal

*Information Technology Department
Thakur College of Engineering & Technology
Mumbai , India*

Neha Patwari

*Information Technology Department,
Thakur College of Engineering & Technology,
Mumbai , India*

Abstract - There are hundreds of different bots, developed for a variety of reasons, they range from hardwired programs with simply coded patterns to systems build upon embedded learning algorithms that continuously expand their language knowledge base. Bots are created purely for fun or as a part of interactive games, internet information services; websites guides e-commerce agents or more. Traditional technologies just used to respond to queries fired by the user by retrieving the answers from their databases but we need to develop a system which would interact with the user in an interactive manner as any other human will interact. Also if the response to the query fired by the user is not present in the database of the system then the system should respond to the user's query using internet. Chatbots typically provide a text-based user interface, allowing the user to type commands and receive text response as well as speech response. Chatbots are usually a stateful services, remembering previous commands (and perhaps even conversation) in order to provide functionality. When chatbot technology is integrated with popular web services it can be utilized securely by an even larger audience. Chatbots are simple to use, It usually introduces itself and we respond with a statement or question. The bot consults its knowledge base or programming languages and replies. The conversation continues as long as it is interesting or useful for human. This system is for college purpose and will respond to user query in an interactive manner.

Keywords-chatbot, chatterbot, pattern matching, keyword matching.

I. INTRODUCTION

Chatbot is a conventional agent that is able to interact with users in a given subject by using natural language [1]. Normally, chatbot has the ability to answer questions from the user, provide comments, or bring a topic to be discussed with the user. Many chatbots have been deployed on the internet for the purpose of education, customer service, site guidance, or even

entertainment functions. Existing famous chatbot systems include ALICE, SimSimi and Cleverbot.

Chatbot can run on local computers and phones, though most of the time it is accessed through the internet. Chatbot is typically perceived as engaging software entity which humans can talk to. It can be interesting, inspiring and intriguing. It appears everywhere, from old

ancient HTML pages to modern advanced social networking websites, and from standard computers to fashionable smart mobile devices. Chatbots talk in almost every major language. Their language(Natural Language Processing, NLP) skills vary from extremely poor to very clever intelligent, helpful and funny. The same counts for their graphic design, sometimes it feels like a cartoonish character drawn by a child, and on the other hand there are photo-realistic 3D animated characters available, which are hard to distinguish from humans. And they are all referred to as "chat bots".

College Enquiry Chatbot project will be built using artificial intelligence algorithms that will analyze user's queries and understand user's message. This system will be a android application which will provide answers to the queries of the users. User's just have to go to the chat screen to ask any query related to the college.

The system replies using an effective Graphical user interface and a voice message which implies that as if a real person is talking to the user. The user don't have to register himself and login to the system since this is very overwhelming process. There is an Admin screen which will help admin to add new keywords and their appropriate response to the database. The user can query about the college related activities through online with the help of this application. The user can query college related activities such as date and timing of annual day, sports day, and other cultural activities, admission process and fees. This system helps the student to be updated about the college activities.

II. MOTIVATION

Artificial intelligence systems technology offers an alternative way to tackle complex ill-defined problems. They particularly useful in system modeling as in implementing complex mappings and system identification. Artificial intelligence (AI) systems comprise areas like, expert systems artificial neural networks, genetic algorithms, fuzzy and various logic and various hybrid systems, which combine two or more techniques. It gives us a touch of reality. Its advanced features like lip synchronization and audio-video-text response helps it to appear more real.

III. PROBLEM DEFINITION

Traditional technologies just used to respond to queries fired by the user by retrieving the answers from their databases but we need to develop a system which would interact with the user in an interactive manner as any other human will interact. Also if the response to the query fired by the user is not present in the AML database of the system then the system should respond to the user's query using internet.

Phase 1(Information Gathering):

Study different papers related to chatbot and getting more and more information about it. To know about the Artificial Intelligence and types of algorithms which is required in creating chatbot.

Phase 2(Analysis& Design) :

Develop the Graphical User Interface and analyze the internal structure and working model of the chatbot and how it will react to the user query. Making a list of different features that we can implement in our chatbot.

Phase 3(Implementation & Testing) :Developing an application in android and connecting it with a backend using volley library. Backend will be made using JSP and servlets. We will be creating Java Server Pages at the backend and different algorithms will be used for processing the user query and fetching the appropriate response for it from database. The code will be tested on real application and accuracy of responses will be measured. Then some methods and algorithms will be used for increasing the accuracy of response

Phase 4(Deployment) :

Once the bot is built, it will be deployed to hosting environment which is our college website. The hosted environment must be stable and needs its own monitoring and develops support.

IV. RELATED WORKS

The biggest challenge of developing a successful question answering system may be the ability to translate question from natural language into a formal representation or equality. Deep QA utilities named entity recognition and relation extraction to extract pieces of a formal representation, such as

a limited set of types of entities and a few types of relations expressed between them. For the most part, traditional question answering system does not get a complete meaning representation of a question and then try to answer it; rather, chat bot finds pieces of knowledge from questions and tries to exploit what sources it has as evidence for answers. Importantly, it does this in the presence of precision and recall errors that normally confound representation and reasoning systems [2].

Traditional question answering has three components such as question classification, information retrieval, and answer extraction. Question classification is used to classify the question based on the type of its entity. Information retrieval technique is employed to extract out relevant answers with query posted by their intelligent question answering system. Finally, answer extraction component is emerging topics in the question answering system where these systems are often requiring ranking and validating a candidate's answer[3]

DeepQA system goes further and discards the question classification operation, largely because it aims to deal with open-domain problems. Instead, it puts emphasis on the question analysis and answer filtering[4]. In our opinion, that is a solution which combines bottom-up and top-down processing strategy. To be more specific, constructing query and issuing it to retrieval document is a bottom-up process; filtering the candidate's answer out with the information extracted from question is a top-down process.

In the computer vision discipline, the appliances of bottom-up & top-down theory to some problems have achieved a higher level understanding [5]. Those researches inspire us to apply those theories to the research and implementation of a dialog system.

V. METHODOLOGY

Focusing on the interaction problem between chatbot and users, the basic challenge is how to understand commands that users issue to the chatbot. For a given user command, different chatbot systems have different solutions to parse. For example, some pretty basic systems can only distinguish "0" and "1" commands. To understand complex user commands expressed by natural language is still a terrible challenge for the natural language processing industry. But that doesn't prevent us from proposing our system framework to integrating more knowledge into chatbot. Our system architecture is showed in Fig. 1.

The basic algorithm that will be implemented for working of this proposed system is as follows:

- Step 1: Start the chatbot application.
- Step 2: Get the query from the user in the text form.
- Step 3: Send that user query to the application server.
- Step 4: Pre-processing of the query will be done on server side using different algorithms.
- Step 5: Fetch keywords from the query using algorithms.
- Step 6: Process the keywords using keyword matching algorithm.

Step 7: Get the appropriate data from database.

Step 8: Return the data as a response of query to the application.

Step 9: Either finish the conversation or go on with another query.

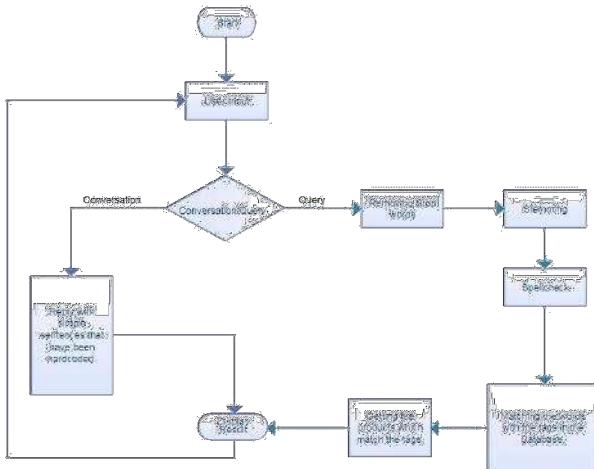


Fig. 1 Chatbot Architecture

VI. RESULTS AND CONCLUSION

The main objectives of the project is to develop an algorithm that will be used to identify answers related to user submitted questions. To develop a database where all the related data will be stored and to develop a good interface. The interface developed has two parts, one for simple users and one for the administrator.

A background research took place, which included an overview of the conversation procedure and any relevant Chatbots available. A database is developed, which stores information about questions, answers, keywords, logs and feedback messages. A usable system will be designed, developed and deployed to the college web server.

This chatbot design may be able to have another direction of interface between human and computer such as customer service. Another direction that can be made is to explore this design with additional tool such as web camera to make the agent can analyze the user's emotion and reaction.



Fig.2. Home screen

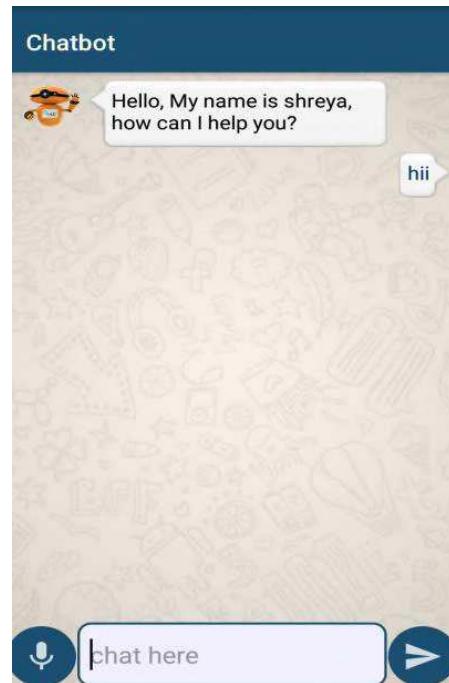


Fig.3. Chat screen

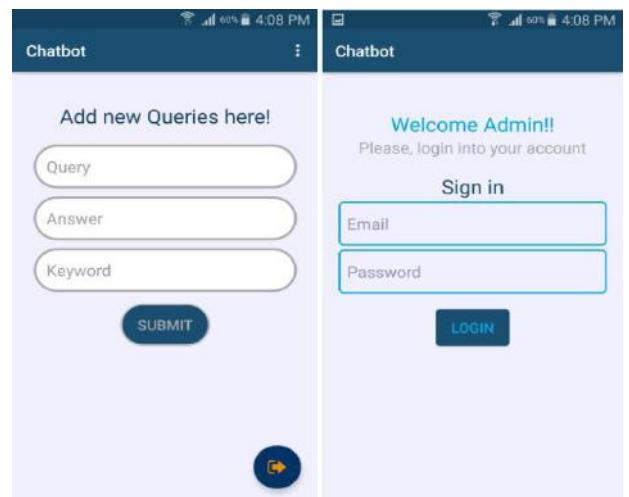


Fig.4 (a) Admin Screen (b) Admin login screen.

Fig. 2 shows the homescreen and Fig. 3 shows the chat box. Fig. 4 (a) shows the admin screen while admin login screen is shown in Fig. 4 (b).

VII. FUTURE SCOPE

Ranging from the fields of education to entertainment, from interesting gossips to work-related information, from admission procedure enquiry to comprehending quantum mechanics, Chat bots have considerable potential in various vital arenas involving communication and sharing of information in an efficient manner.

REFERENCES

- [1] Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL), pages 597– 604.
- [2] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. arXiv preprint arXiv:1506.02075.
- [3] Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 19(2):263–311.
- [4] Zongcheng Ji, Zhengdong Lu, and Hang Li. 2014. An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988.
- [5] Yishu Miao, Lei Yu, and Phil Blunsom. 2015. Neural variational inference for text processing. arXiv preprint arXiv:1511.06038.
- [6] Wenpeng Yin, Hinrich Schutze, Bing Xiang, and " Bowen Zhou, Abcnn: Attention-based convolutional neural network for modeling sentence pairs, 2015.

Automatic Human Detection System

Bhushan Nemade
Department of
Information
Technology,
Thakur college of
Engineering and
Technology
Mumbai,India
bnemade@gmail.com

Pooja Yadav
Department of Information
Technology,
Thakur college of Engineering
and Technology Mumbai,India
rpoojayadav218@gmail.com

Shreysijay Singh
Department of
Information Technology,
Thakur college of
Engineering and
Technology
Mumbai,India
jaysreysi@gmail.com

Kirti Vishwakarma
Department of Information
Technology,
Thakur college of
Engineering and
Technology Mumbai,India
kirtyvish21@gmail.com

Abstract—Nowadays Object Tracking is an important task in video processing because of its variety of applications in visual surveillance, human activity monitoring and recognition, traffic flow management etc. Multiple object detection and tracking in outdoor environment is a challenging task because of the problems raised by poor lighting conditions, variation in poses of human object, shape, size, clothing, etc. Proposed System proposes a novel technique for detection and tracking of multiple human objects in a video. Human objects are detected with help of this trained detector and are tracked using particle filter. The experimental results show that the proposed technique can detect and track multiple humans in a video adequately fast in the presence of poor lighting conditions, variation in poses of human objects, shape, size, clothing etc. and the technique can handle varying number of human objects in a video at various points of time. The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room.

Keywords-Learning based method, false annotations, poor alignment, Surveillance system, abnormal activity detection

I INTRODUCTION

The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room. According to our research, the existing system shows localization errors, false annotations and poor alignment while capturing the images. Training data becomes very large due to complexity of articulated human poses and varying conditions. To overcome these drawbacks, Human Detection Algorithm and Motion Detection Algorithm with enhanced features are used in the proposed system, it gives better alignment and discards double detection and vertical structure problem. The human occurrence is detected in few seconds and alerts are sent to the control room immediately.

Human Detection is an important task in video processing because of its variety of applications in visual surveillance, human activity monitoring and recognition, traffic flow management etc. Multiple object detection and tracking in outdoor environment is a

challenging task because of the problems raised by poor lighting conditions, variation in poses

of human object, shape, size, clothing, etc. It poses a novel technique for detection and tracking of multiple human objects in a video. The experimental results show that the proposed technique can detect and track multiple humans in a video adequately fast in the presence of poor lighting conditions, variation in poses of human objects, shape, size, clothing etc. and the technique can handle varying number of human objects in a video at various points of time.

II. RELATED THEORY

The proposed system presents human detection through visual surveillance at the critical geographic locations and under any atmospheric conditions. Any human presence will be detected and immediately notified to the control room. According to our research, the existing system shows localization errors, false annotations and poor alignment while capturing the images. Training data becomes very large due to complexity of articulated human poses and varying conditions. Existing motion detection security system detects the motion by sensing a change in field view of the camera that is connected with system and as soon as change is detected the alarm starts ringing. The limitation of this system is that it is not intelligent enough. When the motion is detected it does not sense whether change in the field view of the camera is due to an object, animal or human. Hence there was a need of more intelligent system therefore we have designed AHDS keeping in mind the limitation of existing system. The aim is to detect the motion and perform some intelligent operation before ringing the alarm.

III. ALGORITHM FOR HUMAN DETECTION

1. Learning-based methods for human detection

The core discriminative learning-based approaches train different kind of classifier on a large number of negative and positive images samples, where humans are well framed. Every method has to extract appropriate features and the main apprehended information from training data is the spatially recurring of local shape events. If the trained classifier does not detect an object

(misses the object) or mistakenly detects the absent object (false alarm), it is easy to make an adjustment by adding the corresponding positive or negative samples to the training set. However, because of the complexity of articulated human poses and varying viewing conditions, the training data become very large (especially positive samples), consequently the generalization capability of the trained classifier can be compromised. The basic common process of training data is shown in figure below. The features and the used classifiers for each technique are given in Common learning processes.

2. Human activity analysis

In this section, we proceed to evaluate and analyze the human activities as the detected objects and classify them into two groups: normal activities and abnormal activities, based on the support vector machine (SVM). The flow diagram for this step is shown in. The basic idea of support vector machines (SVM) is to find the optimal HYPERPLANE that splits a dataset into different categories. Once the HYPERPLANE is chosen, the distance to the nearest data point of the classes is maximized. This gives an idea about a simple example with two classes in the plane.

3. Motion detection system

Motion detection is the first essential process in the extraction of information regarding moving objects and makes use of stabilization in functional areas, such as tracking, classification, recognition, and so on. The frequently-used algorithms for motion detection are studied, including frame difference method and background subtraction method, and an algorithm composing of those methods for motion detection is proposed. This paper presents a new algorithm to detect moving objects within a scene acquired by stationary camera. In this paper Difference of successive frames and background were calculated by taking mean of n consecutive frames and then compare it with current frame with the help sub block matching-based scheme. It increases the sensitivity of human motion detection.

This algorithm presents an image with white pixels will be greater than a predefined alarm level (threshold), an alarm is produced about a motion event. This estimated background is just the previous frame. It evidently works only in particular conditions of objects, speed and frame rate. It is very sensitive to the threshold so that a noisy image motion will be detected in such places compared to places where there is no motion at all. If the object is moving smoothly, a small change is obtain which is less than the predefined threshold, so, it is impossible to detect moving object. Things become worse, when the object is moving very slowly, then the algorithms will not give any result at all.

IV MOTIVATION

The goal of the proposed system is to create a surveillance system based on the characteristics of the human presence and motion. The proposed system will be able to detect human presence without any use of sensors on the basis of the human characteristics and motion detection. In critical climatic condition such as rain and storm the scaling of the image is mostly accurate. Human detection is an essential and significant task in any intelligent video surveillance system, as it provides the fundamental information for semantic understanding of the video footages. It has an obvious extension to automotive applications due to the potential for improving safety systems. Many car manufacturers The security and surveillance cost will be reduced by using this system.

V. PROBLEM DEFINITION

Detecting human beings in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, human gait characterization, person counting in a dense crowd, person identification, gender classification, fall detection for elderly people, etc. The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays. Thus, human detection and motion analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition.

VI. METHODOLOGY

The proposed methodology is illustrated in Figure1. It contains the following steps:- Take Video Input, Video Cropping, Background Subtraction, Motion Detection, Multi-Scale Video Cropping, Check Human Characteristics(Above 80% Correctness), Machine Learning Algorithm, Human Detected. The original image is a combination of stationary part and motion part. The AHD system makes a decision at each frame for a particular object being tracked. After certain number of frames, object being tracked, the AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity. The aim is to automatically guess the motion of a human or human body from monocular or multi-view video images.

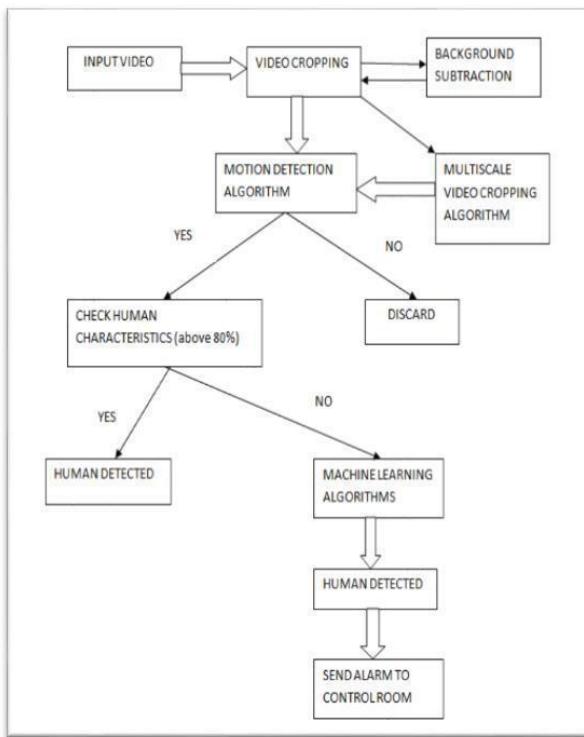


Fig:1: System Architecture

VII. SYSTEM ARCHITECTURE

Phases of our system are

1. Take Video Input
2. Video Cropping
3. Background Subtraction
4. Motion Detection,
5. Multi-Scale Video Cropping,
6. Check Human Characteristics(Above 80% Correctness)
7. Machine Learning Algorithm Human Detected.

1. Take Video Input

Video input of the surrounding will be taken from camera this video will consist of all the objects present in that surrounding frame and capture the video and use it as input image.

2. Video Cropping

Process the image for motion detection and if it fails then background subtraction is applied and in the next step we are using the motion detection algorithm to detect the human.

3. Background Subtraction

Use an image for human recognition using principal component analysis algorithm. The object appearing in front of the camera is subtracted from the background.

4. Motion Detection

The concerned object's motion is monitored and checked whether motion is present using training set to train the system for recognition.

5. Check Human Characteristics (Above 80% Correctness)

If 80% accuracy is not achieved machine learning algorithm is applied to it.

6. Machine Learning Algorithm Human Detected.

At the end it generates the alarm in the control room when the human is detected.

VIII. EXPECTED RESULTS

1. Outputs

In Proposed system, the original image is a combination of stationary part and motion part. The AHD system makes a decision at each frame for a particular object being tracked. After certain number of frames, object being tracked, the AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity. The aim is to automatically guess the motion of a human or a human body from monocular or multi-view video images. The system provides functionality of human characteristics and motion. The system also provides search capability for the user can easily view the human characteristics records. The system also provides the functionality of sending notification and reminder via Alert.

. Outcomes

This approach consists of hybrid features which allows a low-dimensional feature vector and higher recognition accuracy from image sequencing data for human activity recognition. These features include lighting condition recognition properties, invariant characteristics and human silhouettes normalization based on optical flow and distance parameters features. During experimental results, we used two challenging depth image datasets where proposed system has shown significant recognition accuracy performance over the state of-the-art methods. The position of the human are detected at various distances using scaling method to provide the accuracy in recognizing the human.

Phase 1(Information Gathering)

Proposing Statement of work, scope definition and scope boundary for planning the prototype from the problem definition to decide what needs to be done and what not to be done using motion detection system to design.

Phase 2 (Analysis)

Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

Training data becomes very large due to complexity of articulated human poses and varying conditions

Phase 3 (Design)

After certain number of frames, object being tracked. The AUD system computes majority decision taken by analyzing certain frames and then declare it as a human or non-human entity.

Phase 4 (Integration)

Constructing the various modules of the proposed solution to integrate them into a prototype.

Phase 5 (Testing)

The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stress testing.

Phase 6 (Deployment)

The benefits of Analysis on after deploying the proposed prototype include: Proper and more efficient way of monitoring of Human without them actually knowing it. Effective way of managing and storing image data. Reduce the waiting time of human detection. Faster Serving Process. More Secure location. Better access control by centralized user account through remote server Better decision-making so as to detect more human.

REFERENCES

- [1] William Shackleford, Geraldine Cheok, Tsai Hong, "Performance Evaluation of Human Detection Systems for Robot Safety" (Conference of Springer Science+Business Media Dordrecht 2016)
- [2] Shanshan Zhang, Rodrigo Benenson, Mohamed, "How Far are We from Solving Pedestrian Detection" (IEEE Conference on Computer Vision and Pattern Recognition 2016)
- [3] Nagiouldtaleb, Adilchergui, Mohamed larbi ben maati, "Overview on Automatic Detection of Human Body" (IEEE conference on Advanced Video and Signal based Surveillance (AVSS) 978-1-5090-5146-5/16/\$31.00 ©2016 IEEE)
- [4] Malek Al-Nawashi, Obaida M. Al-Hazimeh, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments" (17 May 2016_ The Author(s) conference 2016).
- [5] Shahryar Kamal Ahmad Jalal Daijin Kim, "A Hybrid feature Extraction Approach for Human Detection" (Conference of King Fahd University of Petroleum & Minerals 2015)

Webpage Ranking Search Engine with SEO Suggested Project

Tejas Sutone

*Department of Information Technology,
Thakur College of Engineering and
Technology,
Mumbai, India
tejassutone27@gmail.com*

Harsh Bharucha

*Department of Information
Technology,
Thakur College of Engineering
and Technology,
Mumbai, India
hrshbharucha@gmail.com*

Vikas Kaul

*Department of Information
Technology,
Thakur College of Engineering
and Technology,
Mumbai, India
kaul.vikas@thakureducation.org*

Abstract—Search Engine Optimization (SEO) Techniques often come under the scrutiny regarding their viability and ethical acceptance in the web community. Ever since internet marketing and web analytics became key components of marketing budget of a business, obtaining higher ranking in the Search Engine Result Pages (SERPs) has become highly competitive and consequently outsourced to professional firms. SEO is an evolving methodology which adapts itself to the latest algorithm used by the search engines. In this paper, we discuss the ever-changing SEO guidelines with reference to Google search engine. Current viability of old-school SEO techniques – keywords, backlinks and meta descriptions is examined with reference to regular search algorithm updates by Google. The system provides accurate results and suggestions as per users need. It is easy way to find relevant information about any subject at very low cost. Every user can login and then search for particular results .The user gets a list of URLs according to the query asked that have been generated by the system. : The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data.

Keywords: Keyword, web mushup, meta, google updates, page rank.

I. INTRODUCTION

A search engine is an information retrieval system that has various website data stored in it. It shows user appropriate results on search. It also consists of website data. The search engine records no of times the url was clicked on, its goes to to web url to fetch its meta data and gives points according to page errors. An seosuggester used with it allows user to enter a live url. Our system goes and scans the webpages extracts its meta features and provides seo suggestions. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data. The user can enter the data or text lines (query) to be searched. The system then uses this textand matches it with the content provided in the URLs fed in the database. The user query is matched with these contents and finally it generates a list of related URLs. The user can them click on the URLs thus generated and it will take them directly to the website. The search is based on users demand.

Once user clicks on the links he sees the web page and all the data on it. It also calculates and displays the web page rank based on website meta score and number of visits. The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their website's SEO. Search Engine Optimization is constantly changing as new aspects come into play, and others go. Fundamentally, businesses need to recognize the two crucial fundamentals for high-quality SEO, on-page and off-page. On-page SEO refers to what a publisher can control directly, whereas off-page SEO basically relies on user behavior, social engagements, visitors, and other publishers.

II. BACKGROUND

A search engine results page (SERP) is the page displayed by a search engine in response to a query by a searcher. The main component of the SERP is the listing of results that are returned by the search engine in response to a keyword query, although the page may also contain other results such as advertisements.

Now customers don't just use search engines to find websites to buy things – they use search engines throughout the whole consideration and purchase process:

For Awareness – with categorical & problem-based search queries like “Link Building Tools” or “Measure Social Media ROI”

For Research and Product Evaluation – with queries about specific products and problems, like “ERP for Manufacturing”, or “Facebook Analytics for Retailers”

For Final Product Selection and Reviews – with queries about specific brands, reviews, and comparisons.

III. IMPORTANCE

Search engine optimization (SEO) is the process of improving the visibility of a website or a web page in a search engines results page. Businesses operating in an online environment typically strive to promote their products and services on search engines. To gain more lead generation and ultimately customer's ongoing effort to improve a domain name, page rank and search engine performance is required.The system provides accurate results and suggestions as per users need. It is easy way to find relevant information about any subject at very low cost. Every user can login and then search for particular results .The user gets a list of URLs according to the query asked that have been generated by the system. : The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO. The

search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data.

For Awareness – with categorical & problem-based search queries like “Link Building Tools” or “Measure Social Media ROI”

For Research and Product Evaluation – with queries about specific products and problems, like “ERP for Manufacturing”, or “Facebook Analytics for Retailers”

For Final Product Selection and Reviews – with queries about specific brands, reviews, and comparisons.

Search engine optimization (SEO) is the process of improving the visibility of a website or a web page in a search engines results page. Businesses operating in an online environment typically strive to promote their products and services on search engines. To gain more lead generation and ultimately customer's ongoing effort to improve a domain name, page rank and search engine performance is required. The system provides accurate results and suggestions as per users need. It is easy way to find relevant information about any subject at very low cost. Every user can login and then search for particular results. The user gets a list of URLs according to the query asked that have been generated by the system. : The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data.

IV. MODULES OF SYSTEM

User Login: Every user can login and then search for particular results.

Admin Login: Admin has a separate login and password to enter the admin area.

Website URLs: Admin can enter the site URL and Meta data into the database.

Search Box: The system provides a search box for user to write down their search query. Real time Search List:

The user gets a list of URLs according to the query asked that have been generated by the system.

Page Rank Calculation: The system automatically scans pages and records number of visits on pages as per their usage in search engine and allocates page rank to them automatically. This data is shown to the admin.

SEO suggestions: The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO.

IV. PROBLEM DEFINITION

A search engine is an information retrieval system that has various website data stored in it. It shows user appropriate results on search. The search engine records no of times the url was clicked on, its goes to web url to fetch its meta data and

gives points according to page errors. An seo suggester used with it allows user to enter a live url. Our system goes and scans the webpages extracts its meta features and provides seo suggestions. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data. The user can enter the data or text lines (query) to be searched.. The system then uses this text and matches it with the content provided in the URLs fed in the database. The user query is matched with these contents and finally it generates a list of related URLs. The user can click on the URLs thus generated and it will take them directly to the website. The search is based on users demand. Once user clicks on the links he sees the web page and all the data on it. It also calculates and displays the web page rank based on website meta score and number of visits. The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their website's SEO.

V. METHODOLOGY

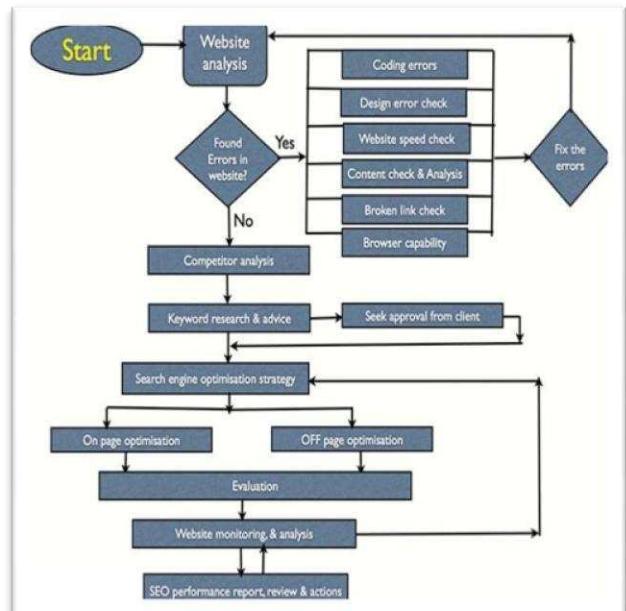


Fig1 Block diagram

Step-1: Website analysis

1.competitor analysis

2.keyword research and advice

Step-2: Installing visual studio 10 windows Based system

1.First double click on setup.exe

2.After clicking you can click on install Microsoft visual stdio2010

3.click next to start installationThen I accept the license term

4.Then setup will start to install its components

5. Reboot your computer and installation is complete

Step-3: create websiteseo optimization

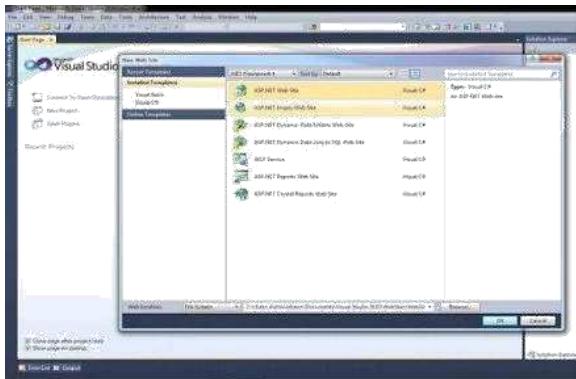


Fig 2: Website monitoring

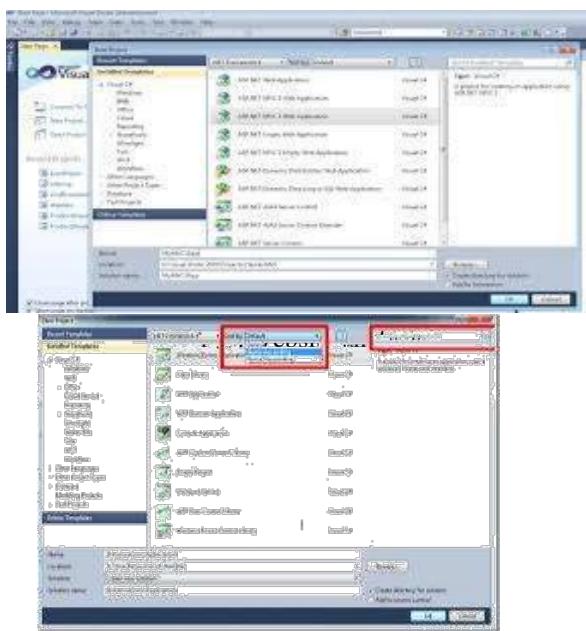


Fig4: Website Monitoring and Analysis

VI. CONCLUSIONS

Search Engine Optimization is constantly changing as new aspects come into play, and others go. Fundamentally, businesses need to recognize the two crucial fundamentals for high-quality SEO, on-page and off-page. On-page SEO refers to what a publisher can control directly, whereas off-page SEO basically relies on user behaviour, social engagements, visitors, and other publishers. The understanding the working of Crawling, Indexing, and Ranking is needed for Seo Practitioner for improve recognizing the Visibility of Website.

VII. FUTURE SCOPE

The project has a vast scope in future. The system can be used by any organization to provide user the convenience while accessing the website.

It can also be used by educational institutes, hospitals or companies in their software or applications to be used by employees. It can be very well used by site owners for SEO suggestions.

REFERENCES

- [1] Nirmala Shinge, Dr.NileshMahajanAn Analysis Studyof Web Page Ranking Algorithms 2016
- [2] Sanjay Kumar, Dharmender Kumar A Review Paper onPage Ranking Algorithms 2015
- [3] Seema Rani ,Upasna Garg A Review Paper On WebPage Ranking Algorithms 2014
- [4] Dilip Kumar, A.K.Sharma A Comparative Analysis of Web Page Ranking

Recommendation System Leveraging Heterogeneity of Trust in Social Networks

Vijay Varma

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Rohit Yadav

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Deepak Yadav

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Aaditya Desai

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Abstract— Recommendation system is a subclass of information filtering system that seeks to predict the rating or preference that a user would give to an item. Recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. There are also recommender systems for experts, collaborators, jokes, restaurants, garments, financial services, life insurance, romantic partners and Twitter pages.

Recommendation systems commonly use previous data of user. The problem in this is that it can be limited in perspective especially when much more data is available the main technological challenge in our project is to use the data of similar rating in connected user's to establish trust also trust between two unconnected users using common connections is one of the major challenges

The main parameter of the project is the percentage increase in view to buy conversion. If there is a substantial increase in this parameter the project can be said to be successful. The other important parameter is the reduction in less relevant recommendations. To find this out we use the feedback of customers on the recommendations

Keywords-components: *Recommendation System, Application, Trust Factor*

I. INTRODUCTION

The main limitation of current recommendation systems in use is that they largely rely on the users past data to give future recommendation. But a major assumption in this approach is that if the user views a particular product he may buy it. This approach can be less accurate as people who visit websites may not intend to buy it. For example, a normal user who visits website of an exotic car will most probably not buy the car. So taking this issue in consideration we thought of developing a system which will be useful to the customers as well as the merchant who intends to sell its products.

The main purpose of our project is to enhance the advertising capability of a merchant. Through the survey we found that there is need to develop such a system which will be of use to customers and merchants equally. And this motivated us to develop such an application which will fulfill their needs.

II. BACKGROUND

Recommendation System was and will always be an important part of E-businesses. There exists many ways to give recommendation to the user. These involve using the data from user's internet activity to predict their likes and dislikes and give recommendation based on it. The other method is using algorithms like Apriori for determining frequent item sets and give recommendation based on it. There is a lot of data available via the trusted connections of people on social media. This data can be used in order to predict the user's preferences in fields where the data about him is not sufficient to give reliable recommendations. In this project, the intention is to exploit this vast sum of data (from the trusted connections) and use it to enhance the experience of the customers by giving them better recommendations and also increases the sales of merchants via this new system. Thus this project intends to solve the problem of less relevant recommendations across various platforms and E-businesses and venture into a new type of system that is the one proposed below.

III. EXISTING SYSTEMS

History based Recommendation Systems:

These systems give recommendation based on history even after the user has bought the item/service. There is no use of data of linked users which can give better insights. Thus we propose the given system.

Problem Definition:

The currently used recommendation systems is based on the individual user data leveraging of data from trusted social media connections to give better recommendation is intended. The new system will implement a user friendly recommendation system which will be a Web application and mobile app (E-Life). This should be based on Trust Factor on social networking.

It should also determine trust between unconnected users based on Trust Propagation.

1. Cosine Trust

This is the simplest model for estimating the strength of trust between two users. We assume that two users trust each other in a particular category if they rate same products in that category similarly. For each category l, we define strength of trust between u and v as follows:

$$S_{u,v}^l = \frac{\sum_b R_{u,b} R_{v,b} C_{b,l}}{\sum_b R_{u,b} C_{b,l} \sum_b R_{v,b} C_{b,l}}$$

2. Recommendation Power

For each category l, the strength of trust between users u and v is defined as:

$$S_{u,v}^l = \sum_b \frac{R_{u,b} R_{v,b} C_{b,l}}{R_u R_b C_{b,l}}$$

A. Requirements and purpose of the study

Hardware Requirement:

- Processor: Standard processor with a speed of 1.6 GHz or more
- Hard Disk: 20 GB or more
- RAM: 1 GB RAM or more

Software Requirement:

- Java Eclipse IDE
- JSP, Servlet
- HTML, JavaScript and CSS for frontend
- Oracle for Database
- JDBC for Database Connectivity

IV. METHODOLOGY

It is proposed that two different approaches to model heterogeneous trust information in recommendation systems. In our first approach, we use k-means clustering to find communities of users with similar preferences in each product category. Thus, for each product category, we will have a different clustering of users into communities.

Clustering Methods:

One of the popular methods of clustering is given below and we intend to use it

K-means:

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Thus depending

on the trust values we can cluster the most trusted users by determining the differences in trust values calculated.

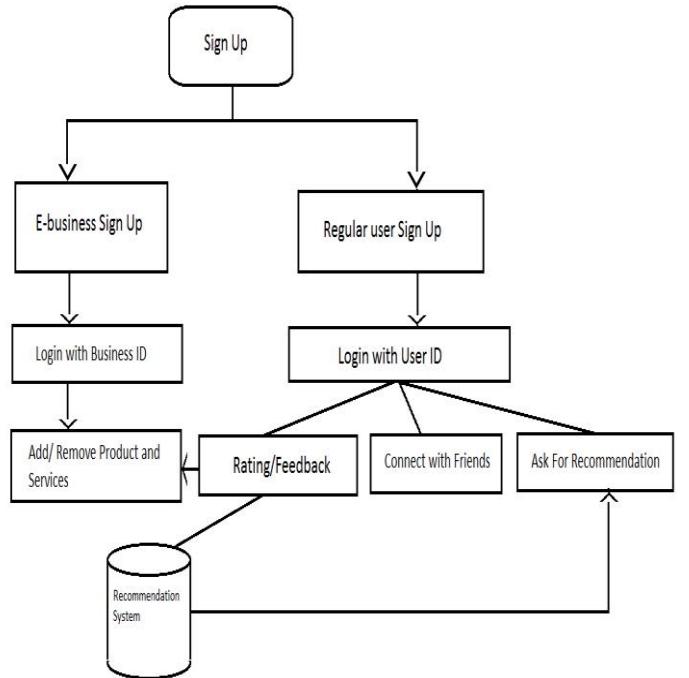


Fig. 1 Flow diagram

V. FUTURE SCOPE

The E-businesses listed on this platform will be able to leverage the vast data on social media and use it intelligently to give recommendations. This software would be useful for many types of businesses like Food joints, Hotels, Movies, Electronic products, etc. Here we have considered the people to people connectivity approach which was not previously explored the previous approaches was product to product approach that is the system previously used to give recommendation according to the similar products. For example if the user is searching for an Adidas shoes then the recommendation system use to show him the recommend products as different types of shoes that is completely based on the keyword, what we have explored is that there is a huge research left in the people to people recommendation system that is the recommendation system is based on the product ratings given by our friends to the products and the people s who have similar interest in the products respectively.

The future scope for the system is to explore the trust propagation technique to give recommendation system which could be useful for calculating trust between multiple users and could help to get the idea of the trust among two user and can help system to judge whether the user X would like the likes of user Y.

VI. CONCLUSION

As it is understood that a recommendation system is an important part of e-commerce, recommending relevant to product to the user is an important part of selling products online and it is important for the e-commerce websites to show products to these user according to their interest and/or according their social circles.

The existing recommendation systems just take in account the users previous search data or previous browsing history to recommend products to the user but through this project it is intended to take into account various other possibilities to reach an more productive, reliable and accurate

Recommendation system which can be used by the E-commerce and E-business to suggest products to their customer to reach greater customers satisfaction and can also increases their product sale to a good margin. Hence by deploying the Recommendation system, it will be easy for the E-businesses to recommend relevant products to the customer there by increasing the probability of sale and also provide better recommendations to the user.

Thus, it can be concluded that by the proposed system for recommendation system can be used effectively by E-business development and better customer relation.

REFERENCES

- [1] J. Ben Schafer, Joseph Konstan, John Riedl “Recommender Systems in E-commerce”, GroupLens Research Project ” 2016
- [2] Pulkit Agrawal, Jayanth Ramesh and Agrim Gupta,”Recommendation System Leveraging Heterogeneity of Trust in Social Networks-2015”.
- [3] Ziming zeng “An Intelligent E-commerce Recommender System Based on Web Mining-2009
- [4] Jianming He and Wesley W. Chu, “A Social Network-Based Recommender System(SNRS)- 2008”
- [5] J2EE: The Complete Reference – Jim Keogh
- [6] Recommender System Charu C. Aggarwal

Automation Using Raspberry Pi 3 Smart Mirror

Yash Kosambia

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Suresh Rajpurohit

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Harsh Singh

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Sandip Bankar

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Abstract- A smart mirror is a device that functions as a mirror with additional capability of displaying multimedia data, such as text, images, and videos. This device allows users to access and interact with contextual information, such as weather data, seamlessly as part of their daily routine. In this project, we developed Smart Reflect - a software platform for developing smart mirror applications. The main features of Smart Reflect are three fold. It is modular, lightweight, and extensible. It allows developers to sidestep the sandboxed environment created by web browsers and it supports plugins written in any programming languages. These improvements alleviate the hardware and software limitations inherent with the use of web browsers as a primary scriptable display method. In this paper, we describe the design and implementation of Smart Reflect and compare it with other similar platforms. We also discuss the potential uses and applications of smart mirrors with regard to the new capabilities that our platform provides.

Keywords – smart mirror, smart reflect, magic mirror.

I. INTRODUCTION

The world around us is constantly changing. With the advancement of science and technology, we are moving towards a more automated way of life. We have smart cities, smartphones, smart cars, and more. This fast way of life requires the development of Home Automation projects. For home automation has many real-world applications, for example, we can build a smart mirror which will be providing web feeds and many information's within the mirror.

Smart mirror is one of the growing priorities of modern day's shopping community. It is an alternative to the traditional mirror

being used in malls, automobiles, and homes. It is an interactive touch-free device that will address the need of accessing up-to-date information required by the user. The combination of mirror with Raspberry pi & display technology will provide useful information that will display and can be customized later thereby can provide much advancement.

II. MOTIVATION

Effective time management is one of the most important factors for success and productivity in a person's day-to-day life. With the increasing integration of technology in our lives, maintaining an efficient schedule has become both easier and more difficult. Keeping up to date with appointments, Twitter, news, social media, and other things is made easier through technology such as tablets, PCs, and smartphones yet also provide distractions that can interrupt anyone's routine. Technology has become another task in the day that time must be allotted for. In the finite time of the day, technology needs to be designed to work within our schedule and not be an extra piece to it [1, 2].

The key to effective time management involving technology is multitasking. Anyone in the business or academic world would agree that every second counts in the day.

III. PROBLEM DEFINITION

A smart mirror is one that is capable of displaying the date, time, weather and traffic conditions on its reflecting surface. These features will be scraped from the Internet and implemented using the raspberry pi board. The Pi board is programmed using Linux OS.

We will use the mean stack method to create the display page and JavaScript is used both at the client and server side. We will be having all the information that we need right in front of us. The smart mirror can also be upgradable to display browsers and social media websites. Adding a motion sensor to the mirror will further increase the speed [3]. A futuristic smart mirror that represents an unobtrusive interface for the ambient home environment.

The proposed system has various phases.

Phase 1: Requirements gathering (hardware and operating system)

Requirement gathering of hardware and software will be done and all the required materials will be gathered together.

Phase 2: designing phase (creating a reflecting display panel)

We will be attaching a reflective Mirror for the Display of the monitor's Output on to the mirror using the Raspberry pi.

Phase 3: coding the raspberry pi.

We will be using the Raspbian Software as the operating system in the Raspberry pi for the coding, which will be done using the java or python programming language.

Phase 4: Testing

We will be testing the developed Product by providing every possible situations, so that the product will be well tested for Quality Assurance.

Phase 5: Deployment

In the Deploying phase we will be deploying the product to the university for the Project Submission.

IV. RELATED WORK

The proposed smart mirror represents a natural interface that facilitates access to personalized services. This is an attempt to contribute to this design of a smart mirror-like interface as well as the smart environment in which the interface is used for interaction in the following, we briefly comment on some related research in this direction. Philips Home Lab [4] is a tested for creating perspective and context-aware home environments. Among several projects, their work on creating an intelligent personal care environment uses an Interactive Mirror [2] in the bathroom to provide personalized services according to the user's preferences. For example, children can watch their favorite cartoon while brushing their teeth. The mirror can provide live TV feeds, monitor the latest weather, and so on. The mirror is a combination of one or more LCD flat screen displays specifically combined with a mirrored surface and connected with a central processor to provide the intended services.

The Interactive Mirror serves as a motivation to provide ambient feelings in the home environment. The work in [4] proposes a Magical Mirror as an interface to provide basic services. The intended services to offer are interactive TV, specific weather data, and searches. Unlike our work, it promotes the use of ontology to personalize the services. However, conceptually, our work has similar objectivity to what the Magical Mirror intends to perform, except that we present a working prototype, whereas some of the functionalities in the Magical Mirror have been presented only by simulation. In addition, we use open standards like web services to communicate with the devices and customize various personalized services for the user, which is not present in the design of the Magical Mirror.

V. METHODOLOGY

The mirror is placed in different sections which include layers of glass mirror and display panel. The mirror will be having a raspberry pi connected to the display device in the frame.

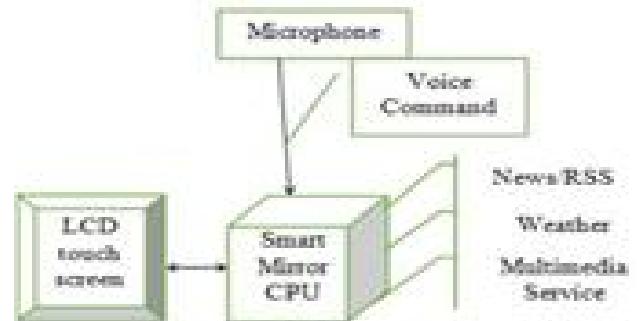


Fig 1: Circuit Diagram

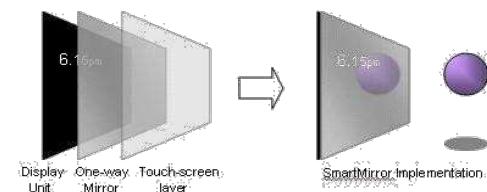


Fig 2: System layers

STEPS:

- Collection of all the hardware layers and frame.
- Raspberry pi configuration.
- Finally will be showing overall review to the end users

We plan to deliver a working prototype i.e. design and development of a futuristic Smart Mirror on Raspberry Pi 3 for the ambient home environment as well as for commercial uses in various industries. Most people have mirrors at home, so the concept of a smart mirror that you can interact with is attractive and can be fantasized by anyone.

VI. RESULT AND CONCULSION

We have designed a futuristic smart mirror that provides natural interaction between users and the ambient home services. The mirror display is provided by a flat LED display monitor which useful for displays all the necessary information which are the user. The mirror also provides a picture-in-picture sub-display to facilitate the display of services such as maps, videos via YouTube. We have developed a functional prototype to demonstrate our work. Overall, the prototype provides an easily extendable framework that can be utilized to provide even more functionality to the user. In our future work we will investigate how the surrounding context of the user and the environment can be utilized in order to provide optimal service experiences in the home environment. The system can be made much more useful to the users by adding more functionality like integrating light settings, speech processing, etc.

REFERENCES

- [1]. Derrick Gold, David Sollinger and indratmo, "Smart Reflect: Smart Mirror", Conference: 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), At Vancouver, BC, Canada. DOI: 10.1109/IEMCON.2016.7746277
- [2]. Anbu, Franklair, Edson,"Smart Mirror: Ambient Control System", DMI College of Engineering, Chennai, ISBN No.: - 1550/6250/10. DOI: 10.1115
- [3]. Jane Jose, Raghav Chakravarthy, Jait Jacob, "Home Automated Smart Mirror as an Internet of Things (IoT)", "IEEE Communications magazine", Vol. 6, Issue 2, February 2017", ISSN:2278-1021.
- [4]. Diego cassada masilla,"Magic Mirror", "ICMI 16.MA IIMI Workshop", Nov 12-16, 2014, Tokyo, Japan.



VII. FUTURE SCOPE

The result of this minimal design is a super clean and simple piece of tech that not only integrates quietly into the background of an otherwise normal household item, but also provides relevant information when it's needed. It perfectly augments the functionality of the bathroom, while not hindering its user with a whole bunch of cluttered buttons and annoying sound alerts. In other words, its technology from the movies brought to life. It's cool. It's stylish. It's simple. But don't look at what's behind it. That's where all the circuitry, wires and components are stored, and that part of the system isn't quite ready for its close-up.

Business Exchange

Rajat Nareshkumar Khemka

*Department of Computer Science, California State University,
Los Angeles, USA
rkhemka@calstatela.edu*

Abstract— In the competitive business environment guidance of business professional is necessary for gaining success in a lawful manner. In this paper we present an idea of project based on business professional interaction with client. The idea is to built a system through which business professional can manage their appointment on their mobile devices while it is also beneficial for clients as it saves their time of interaction which involves visiting the client, scheduling appointment, as everything can be done online through internet. Business professionals considered for this project are charted accountant, doctors, plumbers and mechanics. The idea is implemented using android and is available on google play store.

Keywords—*Business; payment exchange; meeting; professional; client;*

I. INTRODUCTION

In this project we present a business professional – client interaction system based on android operating system. The application's performance on mobile terminals makes it possible for client to access business professional details like (registration number, number of clients, experience, etc) and to obtain necessary responses to his/her queries on their mobile phones. While business professional can track clients details and decide whether to fix an appointment or not, by monitoring client's data and type of query client has. In this project, we solve this problem by developing an application based on android technology, through which the business professional can manage their appointments on mobile phone. In addition to this the client who is unable to go to visit office, can also book their appointment from a mobile phone within 2-3 minutes and then can interact with business professional through text messages. It is difficult to locate and book an appointment with an experienced professional. This consumes precious time of the client. The purpose of this project is to build a medium of communication that will ease the process of booking appointment with business expert. The client will book the appointment through his/her mobile phone. The business professional will be notified about the number of clients he/has has to attend whole day and queries regarding which they need to consult them. This application will save clients as well as business professional's time. The system will be useful for business professional as they can check their appointments whenever and from wherever he wants from their mobile phone.

II. LITERATURE SURVEY

The author states that to search the data online and get the relevant results, it is necessary to rank the search results. Learning to rank is a unique and important issue in the field of machine learning, is what the view of the author is. In this paper, the author has proposed an online ranking algorithm, referred to

as PMRank. In their approach, the perception with margins is employed to learn a ranking model. Since in the training process, margins are introduced, their approach can achieve nearly maximal margin or boundary based ranking. The degrees of freedom that have been set by the author in their approach, makes the training process focused on instances from particular ranks. This approach of their algorithm proves very favorable in the real-world application, such as information retrieval where the errors made on relevant documents are much more serious than the errors on irrelevant documents. The author defines that their synthetic and real world dataset shows that their approach is an effective and efficient ranking algorithm. [1]

SDD The author states that whenever a user comes across a query, search engine acts as a first solution to the query specified by the user. Search engines generally return a large number of pages in response to user queries. From the list of results obtained by the search engine, ranking methods are important to rank the search results. The ranking algorithms proposed are of two types, namely link or content oriented. In this paper the author proposed a page ranking mechanism called Page Ranking based on Visits of Links(VOL) is being devised for search engines, which works on the basic ranking algorithm of Google i.e. Page Rank and takes number of visits of inbound links of Web pages into account. He defines that this concept is very useful to display most valuable pages on the top of the result list on the basis of user browsing behavior, which reduces the search space to a large scale. The paper also presents a method to find link-visit counts of Web pages and a comparison between VOL with the Page Rank algorithm. [2]

The author talks about the ranking algorithm used by the google search engine, to rank the pages according to the user's input. He states that, it is one the most popular and famous search engines, but still its algorithm is still unknown. The author reversed engineered the algorithm in the recent times. He has done three studies on the ranking algorithm. The first study provided a broad overview and the second study focused on researching the impact of citation counts, the current study focused on analyzing the correlation of an article's age and its ranking in Google Scholar. The paper defines their current study as age and rankings of 1,099,749 articles retrieved via 2,100 search queries were analyzed. The analysis revealed that an article's age seems to play no significant role in Google Scholarships as ranking algorithm. It is also discussed why this might lead to a sub-optimal ranking. [3]

The author states that, after the results and the ranking of the searched data, efficiency is another important concept in database management. Database management functions are the fundamental problem underlying many computer applications. In such scenarios efficiency is difficult to achieve if we use the

traditional approach of Von Neumann. The author states that the recent advances in microelectronic technologies have prompted many new research activities in the design, implementation, and application of database machines which are tailored for processing database management functions. He defines that to build an efficient system, the software algorithm are needed to be designed to take the advantage of the characteristics of the search results. Special hardware units should be used, if they are cost-effective, to execute or to assist the execution of these software algorithms. [4]

III. IMPORTANCE OF THE PROJECT

The project provides ease of interaction between client and professional. Project plays a vital role in providing business between clients and business professional, with a list of experienced choices of experts. This project ensures that a client is communicating with a certified business professional and will be beneficial as it will avoid bogus peoples from registering and having access to client database.

IV. MOTIVATION

The motivation behind this project came from general people discussing problem about how to locate a business professional in their area. People often search on search engine regarding this issue; the interesting part here is the search algorithm used by search engine.

Users discussing their problems had while searching for doctor in their area: -

- (a) Where can I find a doctor?
- (b) How can I trust that doctor?
- (c) Should I visit the doctor located at a distance of 1 mile from Alhambra community school?

These are the requirements where in an individual need a doctor to whom he could consult. Since there were many doctors in Alhambra I decided to build a solution by applying ranking algorithm and database technologies to redirect users towards best professionals who can solve their query when the query is related to particular profession.

A. Problem Definition

To develop an application to support “Business Exchange” through which customers can communicate with business professional, to resolve their problem. Business exchange is very beneficial for common people. It is designed by focusing people who don't have knowledge regarding professional to trust in as there are many experts under a particular business professional profession. User can evaluate and provide feedback using feedback link. Involvement of user with an easy to use user interface. Interaction of user with professionals through text message, which would not involve cost from customer's end. To generate clients for professionals and company. To provide clients with a list of well experienced professional. To provide feedback facility with a view to evaluate a professional, this feedback will act as a parameter to rank the professional. Users can search business professional according to their preferred location. Users can fix their meetings through application. Users can view number of clients associated with a professional, they are willing to communicate with. Users can communicate with business professional using messaging facility. Users can rate

professional for the service they receive, this rating is used for rank calculation.

B. Objectives

1. To design a platform wherein we can bridge the gap between business professional and customers
2. Interaction with easy to interact user interface.
3. To provide professional for clients
4. Users can post their queries.
5. From the site of professional, they can directly access clients profile and accordingly filter the clients.
6. It provides clients stage, they can search different professional according to their requirement.
7. Provides feedback link for both client as well as professional

C. Android Oreo

Android Oreo or Android 8.0 (API level 26) allows activities to launch in picture-in-picture (PIP) mode. PIP is a special type of multi-window mode mostly used for video playback. It lets the user watch a video in a small window pinned to a corner of the screen while navigating between apps or browsing content on the main screen. PIP leverages the multi-window APIs available in Android 7.0 to provide the pinned video overlay window. PIP mode makes sure that UI elements are hidden and video playback continues when the activity is in PIP mode. The PIP window appears in the top-most layer of the screen, in a corner chosen by the system. One can drag the PIP window to another location. When user taps on the window two special controls appear: a full-screen toggle (in the center of the window) and a close button (an "X" in the upper right corner). When an activity is in PIP mode, it is in the paused state, but should continue showing content

Notifications have been redesigned notifications to provide an easier and more consistent way to manage notification behaviour and settings. These changes include

1. Notification channels: Android 8.0 introduces notification channels that allow developer to create a user-customizable channel for each type of notification developer wants to display. The user interface refers to notification channels as notification categories
2. Notification dots: Android 8.0 introduces support for displaying dots, or badges, on app launcher icons. Notification dots reflect the presence of notifications that the user has not yet dismissed or acted on
3. Snoozing: Users can snooze a notification, which causes them to disappear for a period of time before reappearing. Notifications reappear with the same level of importance they first appeared with. Apps can remove or update a snoozed notification, but updating a snoozed notification does not cause it to reappear
4. Notification timeouts: Developer can set a timeout when creating a notification using `setTimeOutAfter()`. Developers can use this method to specify a duration after which a notification should be canceled. If required, developer can cancel a notification before the specified timeout duration elapses.

5. Notification dismissal: Users can dismiss notifications themselves, and apps can remove them programmatically. Developer can determine when a notification is dismissed and why it's dismissed by implementing the `onNotificationRemoved()` method from the `NotificationListenerService` class.
6. Background colors: Developers can set and enable a background color for a notification. Developers should only use this feature in notifications for ongoing tasks which are critical for a user to see at a glance. For example, developer could set a background color for notifications related to driving directions, or a phone call in progress. Developers can also set the desired background color using `setColor()`. Doing so allows developer to use `setColorized()` to enable the use of a background color for a notification.
7. Messaging style: In Android 8.0, notifications that use the `MessagingStyle` class display more content in their collapsed form. Developer should use the `MessagingStyle` class for notifications that are messaging-related. Developers can also use the `addHistoricMessage()` method to provide context to a conversation by adding historic messages to messaging-related notifications [6].

D. Firebase Real-time Database

The Firebase Real-time Database is a cloud-hosted database. Data is stored as JSON and synchronized in realtime to every connected client. When developers build cross-platform apps with our iOS, Android, and JavaScript SDKs, all clients share one Real-time Database instance and automatically receive updates with the newest data. The Firebase Realtime Database lets developers build rich, collaborative applications by allowing secure access to the database directly from client-side code. Data is persisted locally, and even while offline, realtime events continue to fire, giving the end user a responsive experience. When the device regains connection, the Realtime Database synchronizes the local data changes with the remote updates that occurred while the client was offline, merging any conflicts automatically. The Realtime Database provides a flexible, expression-based rules language, called Firebase Realtime Database Security Rules, to define how data should be structured and when data can be read from or written to. When integrated with Firebase Authentication, developers can define who has access to what data, and how they can access it. The Realtime Database is a NoSQL database and as such has different optimizations and functionality compared to a relational database. The Realtime Database API is designed to only allow operations that can be executed quickly. This enables developer to build a great realtime experience that can serve millions of users without compromising on responsiveness. Because of this, it is important to think about how users need to access data and then structure it accordingly.

Capabilities are:-

1. Real-time: Instead of typical HTTP requests, the Firebase Real-time Database uses data synchronization—every time a data change, any connected device receives that update within milliseconds. Provide collaborative and immersive experiences without thinking about networking code
2. Offline: Firebase apps remain responsive even when offline because the Firebase Real-time Database SDK persists data to disk. Once connectivity is re-established, the client device receives any changes it missed, synchronizing it with the current server state.
3. Accessible from Client Devices: The Firebase Real-time Database can be accessed directly from a mobile device or web browser; there's no need for an application server. Security and data validation are available through the Firebase Real-time Database Security Rules, expression-based rules that are executed when data is read or written.
4. Scale across multiple databases: With Firebase Real-time Database on the Blaze pricing plan, developer can support app's data needs at scale by splitting data across multiple database instances in the same Firebase project. Streamline authentication with Firebase Authentication on project and authenticate users across database instances. Control access to the data in each database with custom Firebase Real-time Database Rules for each database instance [7].

E. Application Programming Interface

An API is a set of commands, functions, protocols, and objects that programmers can use to create software or interact with an external system. It provides developers with standard commands for performing common operations so they do not have to write the code from scratch. APIs are available for both desktop and mobile operating systems. The Windows API, for example, provides developers with user interface controls and elements, such as windows, scroll bars, and dialog boxes. It also provides commands for accessing the file system and performing file operations, such as creating and deleting files. Mobile APIs, such as the android API, provide commands for detecting touch screen input, such as tapping, swiping, and rotating. It also includes common user interface elements, such as a pop-up keyboard, a search bar, and a tab bar, which provides navigation, buttons the bottom of the screen. The android API also includes predefined functions for interacting with an android device's hardware, such as the camera, microphone, or speakers [8].

API used for this project are:

1. Google Maps: With the Google Maps Android API, you can add maps based on Google Maps data to your application. The API automatically handles access to Google Maps servers, data downloading, map display, and response to map gestures. You can also use API calls to add markers, polygons, and overlays to a basic map, and to change the user's view of a particular map area. These objects provide additional information for

- map locations, and allow user interaction with the map. The API allows you to add these graphics to a map:
- Icons anchored to specific positions on the map (Markers).
 - Sets of line segments (Polylines).
 - Enclosed segments (Polygons).
 - Bitmap graphics anchored to specific position on the map (Ground Overlays).
 - Sets of images which are displayed on top of base map tiles (Tile Overlays) [9].
2. Camera: The Android framework includes support for various cameras and camera features available on devices, allowing you to capture pictures and videos in your applications. Developer needs to use android.hardware.camera2 or camera intent. Developers also provide manifest permissions [10].

V. METHODOLOGY

Business Exchange consists of two modules:

Client Module

1. User first needs to login and fill his / her information
2. User needs to validate information entered by him
3. User searches for business professional by entering a suitable region and receives a list of professional
4. User can apply filter to view business professional information
5. User can enter his query/message, in query box which is provided below business professional information.
6. User receives response from business professional within 2 business days
7. User can decide a medium of communication (either online through text messages or offline by visiting to business professional's offices) by scheduling appointment.
8. A feedback link is provided after 4 business days which is used for ranking the professional.

Business professional Module

1. Business professional needs to login through business exchange login
2. Business professional needs to fills his information and provides scanned documents of degree certificate, social security number, driving license.
3. Business professional are only provided access to clients if they are verified by Business Exchange, if they are not verified their access is restricted and their account is blocked by Business Exchange.
4. After verification business professional can view their information which will be displayed to client, they can edit their information or can add new information.
5. Business professional are notified for query that they receive from users and also about number of meetings that are scheduled with clients
6. If a Business professional plan to cancel the meeting fixed in office, they should notify user through Business Exchange.

A. Data Flow Diagram



Fig. 1 Level 0 DFD

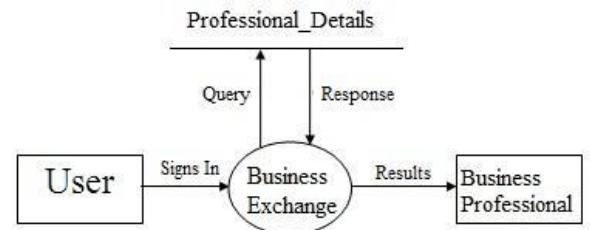


Fig. 2 Level 1 DFD

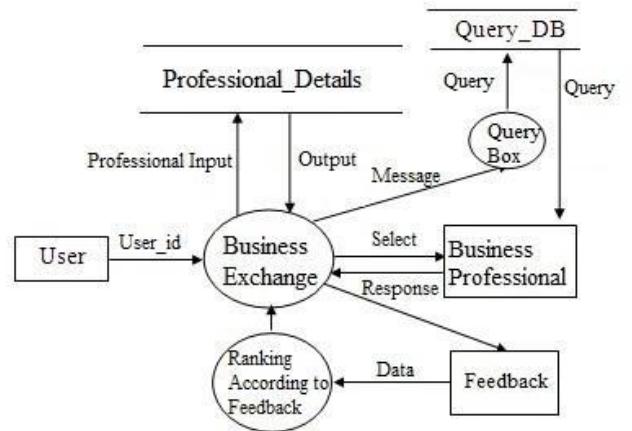


Fig. 3 Level 2 DFD

Project Planning

Project planning refers to everything one needs to set up for project to remain successful in future. It's the process a team needs to go through to establish the steps required to define project objectives, clarify the scope of what needs to be done and develop the task list to do it. The activities in project planning are varied because project team has to work out how to achieve their goals. Every project is going to be different as the objectives will be different. Most of the work of planning is thinking about what needs to do to get everything done and putting the structure in place to make that happen. The project planning phase comes at the start of the project: It's after the initiation phase where team members have really work hard to get approval, to go ahead and put the basics in place (like appointing one team member as the project manager) and before the delivery phase where team members actually do the work. We plan at the beginning to save time later. A good plan means that project manager don't have to worry about whether those people are going to be available on the right dates – because they have planned for tasks that are required to be completed on time to avoid last minute delays. Members don't have to worry about whether everyone agrees on what a quality outcome looks like.

like – because everything is already planned like what quality measures going to used. It sets out the processes that everyone is expected to follow so it avoids a lot of work load later during the duration of testing. For example, if you specify that estimates The system can be improved by implementing the project for United States and slowly expanding so that we have this product available throughout the world. To implement payment exchange platform so that clients can reserve the appointment by paying the appointment fees. To provide client feedback feature where in a client is blocked by Business Exchange if he receives 6 negative feedback from professionals

VI. RESULTS AND DISCUSSION

We expect that “Business Exchange” will provide following results:

1. An android based application for interaction between client and business professionals.
2. An application which consists of two modules:
3. User module which is designed to handle all task which a user might require to perform in order to communicate with business professional
4. Business professional module which is designed to handle all data required by business professional to deal with clients and to build business for themselves.
5. Business Exchange would increase competition among professionals, as more choices and reduced rates for customer enables customer to filter list of business professional according to their requirement.
6. Reduction in rates mainly occurs as customers have more number of options available.
7. Increase number of client's, reduce number of efforts an business professional has to put in, in order to built clients

are going to be worked out by subject matter experts based on their judgment and that's approved, later no one can complain that they wanted you to use a different estimating technique [5].

- as business professional can get details of clients through application.
8. An application which takes on 2-3 minutes to fix an appointment with charted accountant
 9. An application which allows charted accountants to access client's details and response to client's query from anywhere and anytime.

REFERENCES

- [1] Weijian Ni, Tianjin, Yalou Huang “Online ranking algorithm based on perceptron with margins”, IEEE Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress, 2008, pp. 814-819.
- [2] Kumar G, Duhan .N, Sharma, A.K, “Page ranking based on number of visits of links of Web page”, IEEE Computer and Communication Technology (ICCCT), 2011 2nd International Conference, 2011, pp. 11-14.
- [3] Beel J, Gipp B, “Google Scholar’s Ranking Algorithm The Impact of Articles Age (An Empirical Study)”, IEEE Information Technology: New Generations, 2009. ITNG ’09. Sixth International Conference, 2009, pp. 160-164
- [4] Raschid.L, Fei.T, Lam.H, Su.S.Y.W, “A Special-Function Unit for Sorting and Sort-Based Database Operations”, IEEE Transactions on Computers, vol. 35, no. 12), 2006, 1071-1077.
- [5] <https://www.projectmanager.com/project-planning>
- [6] <https://developer.android.com/about/versions/oreo/android-8.0.html>
- [7] <https://firebase.google.com/docs/database/>
- [8] <https://techterms.com/definition/api>
- [9] <https://developers.google.com/maps/documentation/android-api/intro>
- [10] <https://developer.android.com/guide/topics/media/camera.html>

Patient Data Analysis Using Structured Data Marts

Akanksha Chand

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, India*

Jyoti Vikram

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, India*

Padmaja Shukla

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, India*

Abstract—Acquiring the data of patients and services provided to them and then feeding it into the computer reduces the paperwork tremendously but it still doesn't provide any swiftness or accuracy in terms of analyzing that data. The amount and size of the data to be analyzed is gigantic and the hospital environment demands prompt processing of these data in order to efficiently make well informed decisions. This is where the concept of data analytics comes in handy.

Data Analytics is the science of analyzing data to convert information into useful knowledge. It supports the decision making process and also escalates the drawing of conclusions and predication processes as well. For easy access to the data, a data warehouse will be created from which the data will be extracted into the business intelligence tool being used for data analytics purpose. The expected result for the proposed system is to help analysts and actuaries of the hospital to take well-informed decisions on a daily basis, making their work simpler yet efficient to provide enhanced healthcare and for the smooth functioning of the hospital.

Keywords— Data Analytics; Healthcare; Data Marts; Decision Making; Dashboards; Qlikview

I. INTRODUCTION

Information explosion is a rapid increase in the amount of information being generated, especially as a result of the increased use, availability, and sophistication of information technology. Due to which, the problem of managing the information becomes more difficult and can lead to information overload. It is a major concern that needs to be addressed with utmost importance. The rapid technological advancements have enabled organizations to make use of the big data platforms and analytics tools available to handle the colossal volumes of data being generated from multiple sources. The knowledge obtained

from data analytics helps in understanding the current status and prospect better, and in many contexts enables them to make better, well-informed decisions that result in the benefit of the organization.

A hospital also generates plenty of data on a daily basis which needs to be stored and managed efficiently for both medical and business purposes. Deep analysis of these data can certainly reduce the costs of treatment, predict outbreaks of epidemics, avoid preventable diseases and improve the quality of life in general. Several other potential benefits from healthcare analytics are:

- 1) cutting down administrative costs
- 2) clinical decision support
- 3) 3) cutting down on fraud and abuse
- 4) better care coordination and
- 5) improving patient wellness.^[1]

The hospital for which this project is being developed focuses on more than just the physical health care of the patient. It believes in healing all the physical, mental and spiritual aspects of a person thus providing holistic care. Hence the analysis of data will not only help them facilitate better healthcare to the patients but also lay out various alternatives to lead a better life. With the help of the Qlikview dashboards, the doctors and administrators will be able to analyze the pattern in health issues of the patients or the trend in certain services provided which might take longer to discern without this tool. The data will also be updated periodically to accommodate the changes made in the data and regular back up will be taken so that the failure of any database doesn't cause havoc. In consequence the hospital will be able to provide better healthcare to people and also evolve along with it. The goal is to communicate information efficiently and clearly to users.

II. RELATED THEORY

Healthcare has come a long way from giving a reactive response to standard, isolated incidents through subjective decision making. Although hospitals have figured

operational and financial data in their working for decades, the use of universally usable clinical data to make more informed care choices is relatively new. Generally, this clinical data contains information about patients, such as diagnosis, treatments administered, medicines prescribed, procedures and lab tests conducted, and hospitalizations. This clinical data analytics moves the healthcare industry from a subjective, case by case approach to an objective, quantified approach that enables doctors to make better informed decisions. Through this objectivity, caregivers can reduce costs by reducing readmissions, emergency department visits, and wait times.

Needless to say, the detailed insight through clinical data breakdown not only drives business by improving clinical outcomes but also affects higher patient satisfaction and improved care. With analytics, patients can drive a personalized care through an acute analysis of their clinical data that can bring suggestions for preventive care and wellness measures. Through the available data, doctors can generate a rounded view of the patient's health and can drive better diagnosis and timely treatment. Other than reducing recurring costs of follow up care and hospital resources, caregivers can hope to provide a better care experience to patients by this personalization.[2]

A decision that is made by considering all the factors affecting the problem statement is a well-informed decision that in turn is beneficial to the organization. Apart from appealing the human eye, a single graph or chart can describe data that may be presented in multiple paragraphs thus saving a lot of time. One can easily detect efficiency of employees working in the hospital, details with respect to any specific patient and/or employee, upcoming problems and their solutions, areas of improvement that need more focus, areas that are over-focused and are not that important for growth of hospital, etc. Hence the visualization of data through dashboards will make the task of decision making for the analysts and administrators a lot more efficient and simple.

The main focus is on providing the hospital with data that is consistent, accurate and swift. For this we are using Qlikview tool that provides powerful consoles for measuring, monitoring and deploying the data on the server using Qlikview management console. Its automatic associations create endless possibilities for ad hoc querying without requiring tediously defined structures and hierarchies that are typical in other data analysis tools. It promotes unrestricted analysis of application data, helping users make time-saving and accurate decisions. It compresses data and holds it in memory, where it is available for immediate exploration by multiple users. For data sets too large to fit in memory, Qlikview connects

directly to the data source. It delivers an associative experience across all the data used for analysis, regardless of where it is stored. The users can start anywhere and go anywhere; and are not limited to pre-defined drill paths or pre-configured dashboards.

III. LITERATURE SURVEY

In the literature survey, the papers relevant to this project were studied. It included the title of the paper reviewed, its key findings and the solution proposed by the author after key findings. Then the research gaps were identified that can be implemented in this project and the based on those research gaps the problem definition was prepared.

In the paper "Spiritual care visual data analytics for patient healing [3]", dashboard and QVD were used to read the data and provide multidimensional view. The data loading technique used was full load which can be improvised by turning it to incremental load. The second paper, "A proposed star schema and extraction process to enhance the collection of contextual & semantic information for clinical research data warehouses[4]", comprised of CDA Statement, ACT dimension, Substance and Predicate Dimension as Table Column populated with 10 XML values. Populating the data warehouse using hybrid schema to extract data can reduce the redundant values by 65%.

In the third paper, "Using a business intelligence data analytics Solution in healthcare[5]", graphs were used to display data and only 1 operation could be performed. The use of Qlikview which shows dashboard with charts and in tabular form enables at least 8 operations to be performed on the data.

IV. PROBLEM DEFINITION

The hospital generates a lot of data on a daily basis with regards to the patients, their treatment, the doctors treating them, their financials, etc. These data are the lifeblood of the healthcare business, giving insight into an organization's current status and thus helping actuaries and other analysts make informed decisions and projections on a daily basis to provide better facilities and healthcare. But because of the complexity and size of healthcare data, working with it can be frustrating and challenging as each new data set has unique issues and problems. It might even be the case that an important observation is missed because of the amount and intricacy of data. Thus, the hospital needs a system designed in an easy way to analyze this data optimally for the benefit and smooth functioning of the organization at any given time.

V. FEATURES OF THE PROJECT

Data Visualization is the key feature of this project. It allows efficient communication of information to decision makers, which enables them to see the analytics presented visually, so they can grasp difficult concepts or identify new patterns and even forecast the data. Mobile-readiness will ensure all content and assets are accessible, legible, and usable to visitors across all devices. Users will also be able to explore data from any point in the analysis in a dynamic, interactive interface. They will have absolute control over data, allowing them to create personalized, data-driven discoveries with guided paths and analysis. Its flexibility allows unlimited dimensions and measures that can be modified within seconds. Also, data is compressed to 10% of its original size as only the essential bits of data in memory are required for any analysis and Qlikview Any type of data, including healthcare data, goes through three stages before an analyst can use it to achieve sustainable, meaningful analytics: Data capture, Data provisioning, Data analysis.^[6] Apex is the data entry software used by the hospital to capture the data. SQL Server 2008 is used as the database and Qlikview is used as the business intelligence tool. As shown in Figure 1, stored procedures are used to fetch the data collected and stored in the BI database.

From this database, the data is sent to the QVD maker to generate QVD files. These QVD files are used to create the dynamic dashboards that can include pie charts, line graphs, bar graphs, pivot tables that are personalized and are in turn useful for well-informed decision making.

heavily uses data dictionary. Hence, it compresses the original data to a very small size.

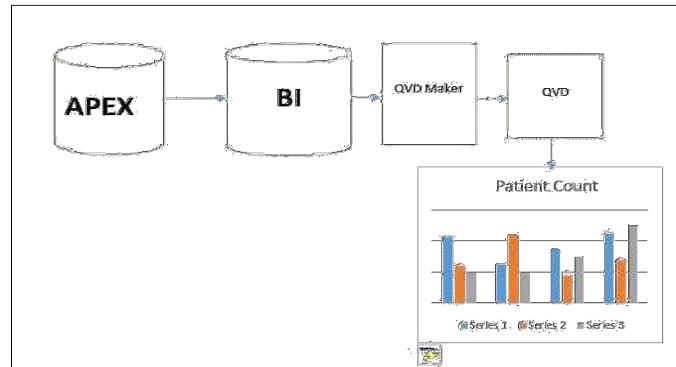


Fig. 1 Architecture

VI. RESULTS AND DISCUSSION

The dashboards are created by thoroughly observing and studying the working environment of the hospital, type of data being generated, its source and application. They are developed as per the requirements provided by the departments and follow the styles and format of the dashboards that are already on the hospital's server. They have been tested to validate the results displayed to ensure the accuracy of the data before being uploaded on the live server. They provide complete multidimensional view of the data required by the doctors, analysts and actuaries of the hospital in discovering insights and making meaningful decisions ergo, facilitating better healthcare and smooth functioning of the hospital.



Fig. 2 Dashboard (viewing all data)



Fig. 3 Dashboard (viewing selected data)

VII.CONCLUSION AND FUTURE SCOPE

The implementation of this project provides a single platform for the employees to access information that establishes the consistency of data across the hospital. The simplified dashboards developed using Qlikview convey visual analysis of the data that help in drawing conclusions and analyzing patterns, trends and correlations that might go undetected in data that are purely text based. These dashboards are deployed on the live server and a hyperlink for the same is given on their portal for employees to access them. It gives them clear and concise information swiftly. It enables the hospital to provide advanced healthcare to its patients with over the top performance at lower costs.

Implementation of SSIS (SQL Server Integration Services) which will allow incremental load approach i.e. it will only fetch the data which is recently added on that particular day and append it to the database instead of dropping the whole table and creating new table on daily basis that is currently being done. Use of SSIS instead of stored procedure, which is an ETL tool, will save a lot of time. Creation of a website which displays all the dashboards having details of the patients that can be accessed by doctors as well as department heads for analysis. Single window statistical reports in Google Sheets are to be created to embed in the website. SQL queries will be written using the Super Metrics Technology that will populate the data in Google Sheets.

VI. ACKNOWLEDGMENT

We would like to extend our gratitude to the people without whom this project would not have been possible.

We sincerely thank our guide Mr. Aaditya Desai for his guidance, support and endurance throughout our project work. We would also like to thank Mr. Vikas Kaul for providing us with the opportunity to work on this project. A special thanks to our external guide Mr. Ranganathan Balsubramaniam for providing us space, resources to complete our project along with his guidance and Mr. Bhushan Dukhande and Mr. Shridhar Dontula for helping us throughout our project. We also thank our project coordinator Ms. Neha Kapadia for arranging the necessary facilities to carry out the project work.

REFERENCES

- [1] Austin B. Frakt, and Steven D. Pizer, "The promise and perils of big data in healthcare", *The American Journal of Managed Care*, 2016.
- [2] <https://www.hcentive.com/category/health-care/page/2/>, hCentive, 2015.
- [3] Ronak Shah, Ritesh Mishra, Prashant Raina, Vikas Kaul, Ranganathan Balsubramaniam, Dr.S.K.Narayankhedkar and Vaishali Kamat, "Spiritual care visual data analytics for patient healing", *IOSR Journal of Computer Engineering (IOSR-JCE)*, 2016.
- [4] Michael Blechner, Rishi KanthSaripalle, and Steven A Demurjian, "A proposed star schema and extraction process to enhance the collection of contextual & semantic information for clinical research data warehouses", *IEEE*, 2012.
- [5] Osama Ali, Pete Crvenkovski, and Helen Johnson, "Using a business intelligence data analytics solutionin healthcare", *IEEE*, 2016.
- [6] <https://www.healthcatalyst.com/healthcare-analytics-best-practices>, Russ Staheli, 2014.

Intelligent Tourist System

Rohan Shetty

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, Maharashtra
shetty.rohan83@gmail.com*

Priyanka Verma

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, Maharashtra
priyanka96.tcet@gmail.com*

Dharti Singh

*Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai, Maharashtra
dhartisingh13@gmail.com*

Radhika Kotecha

*Department of Information Technology,
Thakur College of Engineering & Technology, Mumbai,
Maharashtra
kotecha.radhika7@gmail.com*

ABSTRACT - Tourism is one of the biggest sources of revenue of any country. But people find it difficult to plan their trip in foreign countries or even other states of their own country. In this technological era, people prefer to plan their trips themselves online rather than paying a huge amount to tour operators. The paper proposes a solution using of machine learning techniques and recommendation systems to simplify the trip planning process for the end user. Various methods that utilize finding patterns in given user's past behavior or to find similarities and correlations between known points of interest are included. These smart systems could be used to give personally tailored experience and recommendations for user. These recommendations could be dependent on their schedule, preferences and budget. Machine learning would also automate the travel planning process and make the finding of the new interesting places easier for wider audiences.

Keywords: Intelligent information, Travel Guide, Decision Support, Bayesian algorithm, Information filtering

I. INTRODUCTION

Normally tourists have to search a lot of websites to get information and routes about the destination and this takes a lot of time, thus making it frustrating. Many a times tourist don't have adequate information of the places they're visiting, thus having a tough time finding appropriate destination as per their taste and budget. Existing system will take a lot of time. The traveller might have to limit themselves into only few places to visit if they don't have enough time.

The proposed system is helpful for people that want to visit a city without having much information about it. The main idea is to design a system that is able to find a route using user criteria. It should also be able to estimate time needed to travel from one object to the next and if it is possible, advise which bus line or other public means of transport may be used. It also tracks the user's location using GPS. To build an application, that would-be context aware and assistive for the user, the concept of machine learning and predictive algorithms are needed. The application

would update its recommendations, when it detects that the user is in foreign country and connected to Wi-Fi-network. The current location and date would be used as a context, and automatically suggested activities would be given, based on the past interests. More the user travels, more attractive the recommendations should be for the user. The chosen machine-learning algorithm should recognize patterns in users' preferences and use these patterns, as base for the suggested points of interest.

II. LITERATURE SURVEY

In existing system, the traveller has to search a lot of websites in order to get the knowledge of the routes or information about the particular place. Existing system will take a lot of time. The traveller might have to limit themselves into only few places to visit if they don't have enough time. This system has no functionality to find a route using user criteria. We are unable to estimate the time needed to travel from one place to another and know which bus line or public transport to use in order to reach our destination. This system is not adapted to track the user's current location using Global positioning system to give better and accurate results. Features such as online forum which makes the system more user friendly and gives users a chance to communicate with each other for better feedback is absent. Also, personalization to cater the specific needs of the user is absent thus making it not very user friendly. Table I describes the existing systems.

Table I: Summary of Existing Systems

Ref. No.	Approach	Proposed Solution
[1]	An application, that would-be context aware and assistive for the user, the concept of machine learning and predictive	Categorical data found out to be difficult one to deal with Scientific Python

	<p>algorithms are needed. The application would update its recommendations, when it detects that the user is in foreign country and connected to Wi-Fi-network.</p>	<p>libraries, and after pre-processing those values to numerical representatives, so instead of using numerical values we use discrete values which is much simpler to be dealt with.</p>	[4]	<p>The system provides information about the locations in the form of notifications and also provides the shortest path to reach the destination location. The application also provides information about the weather conditions of a location the tourist wants to visit and recommends him/her to take the appropriate accessories according to the weather</p>	<p>This system does not use machine learning. Thus because of this new data is not updated, leading to the same stale results coming up and filtering is the same. Errors are also called due to multipath fading, which occurs when a signal bounces off a building or terrain before reaching the receiver's antenna, also can reduce accuracy.</p>
[2]	<p>Recommender systems (RS) are used to help users find new items or services, such as transportation or even people, based on information about the user, or the recommended item. These systems also play an important role in decision-making, helping users to maximize profits or minimize risks.</p>	<p>RS development lacks studies analysing early stages, such as requirements and design, and late stages, such as maintenance. Open questions in these stages must be investigated to improve the knowledge about the field.</p>	[6]	<p>Simplified service architecture as well as the employed technologies used for the development of the different system components. The system is mainly composed by a representational state transfer (REST) service with a data access layer which exposes a set of endpoints in order to be accessed by the client applications. Client applications are available as Web-based and Mobile</p>	<p>Cold start or first-rater problem - which means that a new item added to the system cannot be recommended to any user until that item is rated. Sparsity - users tend to rate only a few items over the entire set, resulting in a very sparse matrix, i.e., a matrix with a high percentage of empty cells. This compromises the ability of the systems at successfully locating neighbours, yielding poor recommendations . Good performances require that a very large number of users have rated a very large number of items. Grey sheep problem - individuals whose opinions are unusual, rarely receive accurate recommendations</p>
[3]	<p>An Intelligent Tourist Attractions Decision Support System (ITAS) that has an intuitive and interactive user interface, combines CF-based data mining with the Google Maps API, and uses an ROC curve to evaluate the accuracy of the Bayesian network.</p>	<p>Continuation of this study can, with the same methods, extend to include information regarding accommodation to provide more comprehensive travel information.</p>			
[5]	<p>Although there is numerous information provided on the web, the user gets puzzled in finding accurate information. In order to solve these web problems, the concept of semantic web comes into existence to have communication between human and computer. In this paper, we propose intelligent recommendation system</p>	<p>Since this paper takes the help of artificial intelligence, ratings given to places won't be recommended and so the same results will keep coming thus making the results cold.</p>			

III. PROBLEM DESCRIPTION

This system is helpful for people that want to visit a city without having much information about it. The main idea is to design a system that is able to find a route using user criteria. It should also be able to estimate time needed to travel from one object to the next and if it is possible, advise which bus line or other public means of transport may be used. It also tracks the user's location using GPS.

The system will work in 4 phases, as follows:

Phase 1(Analysis and Planning): -

We have taken reference of 6 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers are secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and are found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

Phase 2(Design and Implementation): -

The intelligent tree algorithm covers both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. To make sure that the front end (login forms and the questionnaire) are connected to the databases and function properly. Python will be used for coding the frontend

Phase 3(Testing): -

To check if the information filtering process works properly and try to use it with the GPS system to make sure suggestions favouring the user's interests are given.

Phase 4(Deployment): -

Implementation of the filtering process and making sure the user's forum functions properly.

Following functionalities are included in the project:

- Login and register: User enters his/her credentials and is able to use the services provided by our website.
- Interactive questionnaire: Series of questions asked to find out where the user's interest lies.
- Personalization: Caters to the specific needs of the user and makes it user friendly
- Location tracking- When the user logs in to our site, his current location is automatically tracked using GPS enabled system, based on which nearby places are suggested to him.
- Information filtering: With the help of machine learning, places liked and suggested by the previous users stored in the

online database will be generated for the next user based on similar interest

- Online forum: Not only do they help in gathering useful feedback but helps the user in shortlisting the places he would like visiting.

IV. PROPOSED METHODOLOGY

The proposed methodology is illustrated in Fig. 1. It contains following steps:

- Data Extraction
- Data Preprocessing
- Data Integration and Transformation
- Feature Selection
- Classification

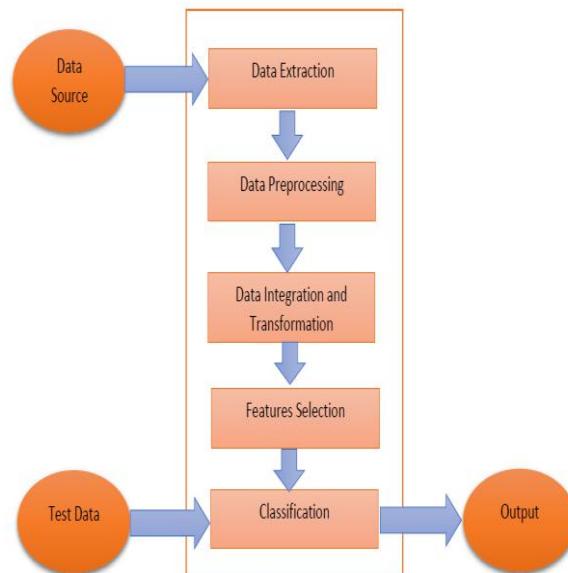


Fig. 1: Proposed Methodology

We need to extract important characteristics form user's data, pre-process them, integrate and transform all data into a unified format, select appropriate features and them perform classification for the user. Following classification techniques can be used:

Random Forest:

In bagging (Bootstrap Aggregation), numerous replicates of the original dataset are created to reduce the variance in prediction. Random Forest use multiple decision trees which are built on separate sets of examples drawn from the dataset. In each tree, we can use a subset of all the features we have. By using more decision trees and averaging the result, the variance of the model can be greatly lowered.

Decision Tree:

decision tree is commonly used in classification in machine learning, it is a tree-like graph or model to make decisions and get consequent results. Decision tree implicitly perform variable

screening or feature selection and the best feature of using it is easy to interpret and explain!

kNN:

K nearest neighbors is a non-parametric method used for classification and regression. The input normally consists of k closest training instances in the space. An object is classified to a class because it is most common among its k nearest neighbors. It is chosen because knn don't need make any assumption on the underlying data distribution, it is also a lazy algorithm and don't use training data to do any generalization.

Naïve Bayes:

Naive Bayes classifier is a probabilistic classifier based on Bayes' theorem. It depends on the assumption that some features are conditionally independent. It is super simple and can converge quicker than some other models, so it has a chance to achieve fast and easy process.

Following feasibility study has been conducted for our system:

- Technical Feasibility – The application will operate on Web Browser. The system will be implemented using python, django python framework, and web technology as required. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms which supports python.
- Economic Feasibility– This system has functionality to find a route using user criteria. We are unable to estimate the time needed to travel from one place to another and know which bus line or public transport to use in order to reach our destination. This reduces the cost of a tourist guide and the effort needed to look for places in a new city.
- Legal Feasibility-The project is developing under legal license of windows, python and other software's required. The data will not be used for some illegal purpose and will be used for making the news relevant to user as well as for publisher for making under certain constraint. Data will not be misused or misguided under any circumstance.
- Operational Feasibility -An overview of the process portrayed that the proposed Machine Learning for Online Tourist System be beneficial for news publisher to monitor the growth and improvement in the future article content. Also to user to view and read the relevant content.
- Social Feasibility- General users have to go to the place where the system is held. Because of that, the system affects the existing culture of the user. And it is a time waste. It can also be a reason to avoid this system. People do not have a method to compare old and new products. The system provides this facility to users.

Fig. 2 shows the process of training the model using training data and implementing the model for getting the testing data. By training the model it will be used for input testing the input data.

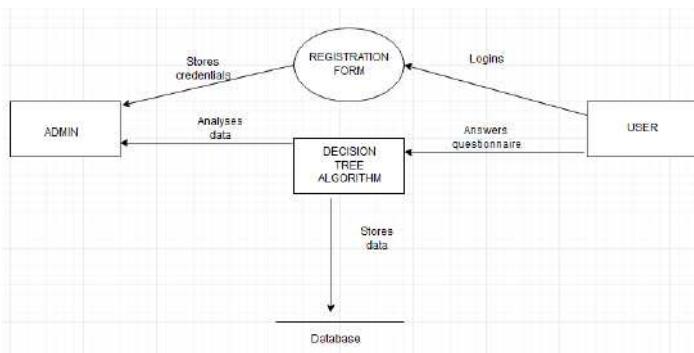


Fig. 2 Basic flow of information

Fig. 3 shows how the model in which the sample articles are being segregated into popular and unpopular articles using the model which is trained using the 49k sample.

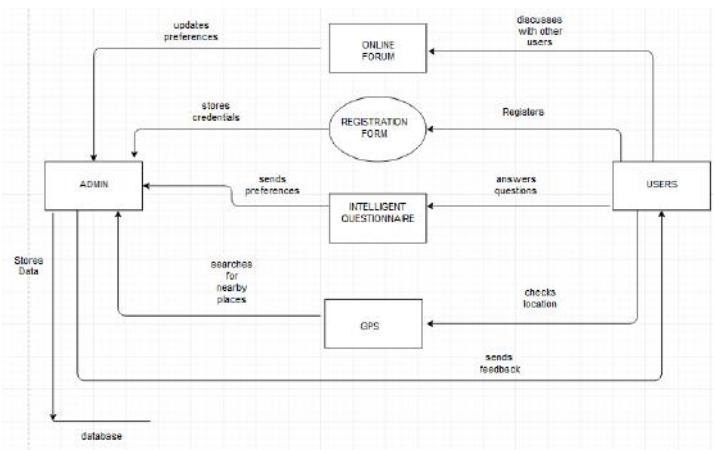


Fig. 3 Model of data flow

FLOW CHART

Fig. 4 shows the collection of raw data from online repository and training the data using different machine learning algorithm like random forest, decision tree, knn. etc.

In proposed system, the user will have to activate the application in their mobile or devices they are using. They can get every relevant information related to the place they wish to visit. The search mechanism and finding up of relevant path feature is the best, this intelligent tourist information application contains

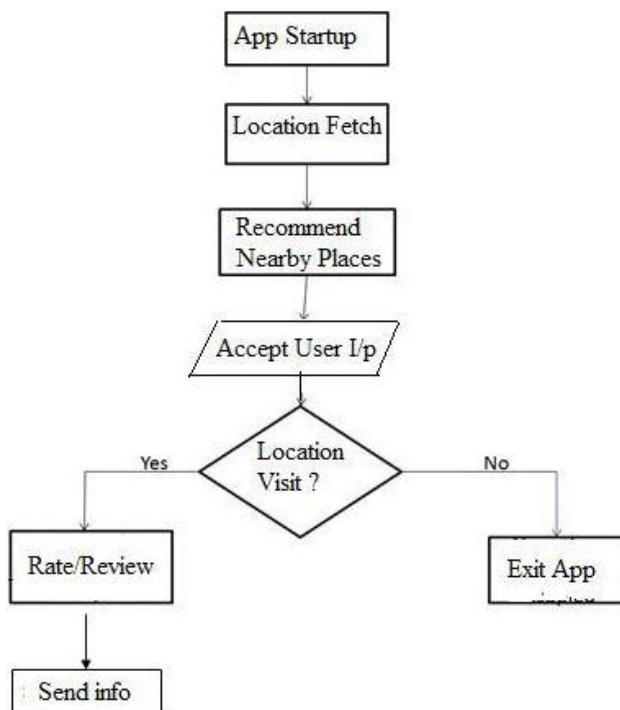


Fig. 4 Flowchart of application usage

Phase 1.1: Planning

- Proposing statement of work, scope definition and scope boundary for planning the prototype from problem definition to decide what needs to be done and what not to be done
- Finalizing each and every requirement with the customer

Phase 1.2: Analysis

- Finalizing the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase

Phase 1.3: Design

- Proposing network architecture for the website and its modules to implement the surveyed case studies and synopsis
- Creating an interactive questionnaire for analyzing user's query

Phase 1.4: Coding

- Various modules will be created like

- Creating the frontend using HTML,CSS and Python
- Creating the Backend using SQL

Phase 2.1: Integration

- Constructing the various modules of the proposed solution to integrate them into a prototype.

V. CONCLUSION

Machine Learning (ML) uses computers to simulate human learning and allows computers to identify and acquire knowledge from the real world, and improve performance on some tasks based on this new knowledge.

This system solves all problems of the client just by asking some questions. And it provides three best options that fit into users requirements along with the place details and facilities. It saves their money and time in finding and consulting a travel agency where they charge more.

This application provides recommendations to the tourists of the nearby locations worth visiting automatically without the need to search for the locations explicitly. The system provides information about the locations in the form of notifications and also provides the shortest path to reach the destination location.

So basically, this application provides accurate suggestions according to user preferences and gives them other's users feedback and can even interact with them using the online forum

REFERENCES

1. J. Kämppä, "Recommendation systems in the context of tourism", 2016.
2. P. Portugal, D. Alencar, D. Cowan, "The use of machine learning algorithms in recommender systems: a systematic review", 2015.
3. H. Fang-Ming, Y. Lin, T. Ho, "Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps", Expert Systems with Applications, 2012.
4. S. Walunj, P. Bhalere, Y. Bhaidkar: "Machine Learning Based Tourist Place Recommendation System", 2016.
5. C. Choi, M. Cho, J. Choi, "Travel Ontology for Intelligent Recommender System", 2009.
6. A. Umanesta, A. Ferreria, N. Leita, "A Recommender system with social interaction", 2014.

Appendix

Ad-Hoc Network and Database Object Analytics

Aniket Pathak

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*
pathakaniket2009@gmail.com

Sagar Parmar

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*
sagarparmar16@gmail.com

Pankti Parikh

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*
panktiparikh2003@gmail.com

Vikas Kaul

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*
kaul.vikas@thakureducation.org

Ranganathan Iyer

*Manager Information Services and BI
BhaktiVedanta Hospital and Research Institute
Mumbai, India*
ranganathan@bhaktivedantahospital.com

Manoj Nambier

*Manager Information Technology and Billing
BhaktiVedanta Hospital and Research Institute
Mumbai, India*
manojn@bhaktivedantahospitalt.com

Abstract — Ad-hoc Network and Database Object Analytics provides a way of solving the problems faced by the businesses by saving time and effort of analysing the data. This paper discusses about creating dynamic dashboards which allows the decision makers to analyse the data of networks, database objects and other important data of healthcare unit. Here a business intelligence tool named QlikView is used as a front end for displaying the dashboards. Data is obtained from SQL Server through stored procedure. The job for the stored procedure is scheduled on the daily basis and is fetched to the QlikView management console. This job would run every day on fixed time. The dashboard will be stored on management console site of the healthcare unit whose link will be sent to the IT-Department so that they can directly access the dashboards and do their analysis on daily basis. It will take around 2-3 minutes to fetch the data from database.

Keywords – analysis; business intelligence; networks; database object; database; data integration; dashboard; QlikView; healthcare unit database; schema; SQL server; management; IT.

Machine Translation

A Hybrid Statistical & Rule-Based Machine Translation System

Rajesh Bansode

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Harsh Khanna

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Abhishek Pandey

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

SahilTikoo

Department of Information Technology,
Thakur College of Engineering and Technology,
Mumbai, India

Abstract— People from different communities with different capabilities and knowledge become moot when they are not able to expand and spread their ideas due to the language barrier. Many government documents and records are also stored in the regional languages of that state, whose digitization becomes a problem. Machine Translation becomes a tool that supports these functionalities while aiding expansion of knowledge and ideas of people from different communities and removes the language barrier. The development of a hybrid model to use the benefits of representation models and overcoming the small hurdles such as increasing the translation rate by the parallel corpora is beneficial in accommodating the new domains of various languages to improve probabilistic nature of translation of texts of various languages.

Product Based Review System Using Data Mining

Raj Desai
 Department of Information Technology
 Thakur College of Engineering & Technology
 Mumbai, India
raj.desai459@gmail.com

Sangita Padshala
 Department of Information Technology
 Thakur College of Engineering & Technology
 Mumbai, India
sangita.padshala@gmail.com

Shraddha Paghdar
 Department of Information Technology
 Thakur College of Engineering & Technology
 Mumbai, India
Shraddha.paghdar@gmail.com

Anil Vasoya
 Department of Information Technology
 Thakur College of Engineering & Technology
 Mumbai, India
vasoyaanil32@gmail.com

Abstract - E-commerce is widely increasing business in the world with increasing revenues every year by manifold times. This is simple indication that more people each day are moving on-line for shopping. It suggests that the preference of the people are changing. People are preferring to buy products more and more from on-line e-commerce websites. They are successfully able to bring the customer in confidence and overcoming the risk involved in on-line transactions by carefully analyzing the behavior of the buyers/sellers and cultivating confidence among the resellers and buyers by developing strategies to attract more business and participation from people.

Here comes the existence of product review analysis into picture which enable the buyers to get product review from different E-commerce websites on one platform and processing those reviews to get the final review of the product, thus facilitating the buyers to purchase the product. There are many factors that are involved in developing a successful product. The success of the product is defined by the usefulness and the influence the product has on the human life.

The product review analysis using data mining enables the comparison of products reviews from different online websites on the basis of the prices and other parameters. To obtain prices and other parameters of the products from different E-commerce websites , a web Crawler is used that can crawl to different E-commerce websites and fetch the URLs of the products. Once the URL of the products are available, scrapper scrape the details that were abstracted within the URL. This scrapped data is then stored in the database. Then the classification model-Naïve Bayes will be used to classify the positive and negative reviews.

This way the Project aims to help out the end-users (customers) with the genuine reviews of the product from multiple ecommerce sources on a single platform they are wishing to purchase thus saving user's time , money and efforts.

Capstone Project-Big Data

Rahul Neve

Thakur College of Engineering & Technology

Mumbai, India

Nishant Singh

Thakur College of Engineering & Technology

Mumbai, India

Pranay Surana

Thakur College of Engineering & Technology

Mumbai, India

Anurag Singh

Thakur College of Engineering & Technology

Mumbai, India

Abstract- Drive better business decisions with an overview of how big data is organized, analyzed, and interpreted. It provides an introduction to one of the most common frameworks, Hadoop, that has made big data analysis easier and more accessible increasing the potential for data to transform our world!. By following along with provided code, you will experience how one can perform predictive modeling and leverage graph analytics to model problems. This specialization will exploration of large, complex datasets. Prepare you to ask the right questions about data, communicate effectively with data scientists, and do basic Build an ecosystem using tools and methods of Big Data. Analyze a data set simulating big data generated from a large number of users from a game In this paper, we are going to analyze the given datasets using modern tools like Hadoop, Splunk, KNIME etc.

Data Mining in Result Analysis

Tanvi S.Patil

Thakur College of Engineering & Technology
Mumbai

tanvipatil2704@gmail.com

Vikas Kaul

Thakur College of Engineering & Technology
Mumbai

kaul.vikas@thakureducation.org

Abstract- The main objective of the educational organization is to provide high quality and necessary education to its students. The two goals of data mining in Indian education system is to analyze and enhance the chronicle way of recent educational data mining advances development; the second is to preserve, organize and discuss the content of the result which is produced by a data mining approach. The use of various data mining techniques such as random forest, decision tree, neural network etc in Indian education processes will help to improve students' performance and provide a broad decision management skill in selection of courses as per their retention rate. The aim of this study is to use data mining techniques for predicting the students' performance at university using only pre-university marks and examination marks of early years at university, no socio-economic or demographic features are used.

Pathology Lab Management System Using Cloud

Varun Agarwal

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

.Tejas Mahajan

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

NehalChaurasia

*Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India*

Kamal Shah

*Department of Information Technology
Thakur College of Engineering and Technology Mumbai, India*

Abstract—Now a days there are many problem like we have to maintain all the medical documents, which we have to carry along us whenever we visit the doctors. There may arise a situation where the physical records(Documents) may get misplaced or destroy, which can be a very big issue. Most hospitals face several challenges with Pathology Lab Management System because some of them are still using manual processes, while the ones that use the computerized method are also faced with the challenge of adjusting to it. Such problems includes a) High cost of software development, deployment and improvement (b) Difficulty in migrating from manual processes, because both staff and patients are used to the manual processes and so are unable to speedily cope with the new system. (c) Lack of IT friendly medical personnel is also presenting several challenges. (d) Huge influx of patients visiting government hospitals makes the process of migrating to automated processes highly difficult. Considering the above, there is need for the improvement of computerized Pathology Lab management system to such hospitals as it would help provide and customize clinical data, enable faster diagnosis with ready-made templates, allow doctors to follow advanced medical prescription patterns, and so on. For a smooth management of pathology lab we have designed this web application in which there are wide range of features for the smooth flow and working of pathology lab. This includes uploading of the reports of the patients, and saving his medical history, the doctors can view this medical history from his login and treat the patient accordingly, and the patient also can view and search his history from the patient login after the payment of the report is done. We have used different algorithms like binary search, quick sort for enhancement of the web application. Pathology lab management system using cloud is web application developed for easy access of medical report to the patient as well as for the doctor's.

Design, Implementation, and Evaluation of SemHD: A New Semantic Hierarchical Sensor Data Storage

Mehdi Gheisari

*School of Computer Science and Educational Software,
Guangzhou University,
Guangzhou, Guangdong Province,
China
mehdi.gheisari61@gmail.com*

Masihollah Mahmoudpour

*Sharif University of Technology,
Tehran, Iran
Masih.Mahmoudpour@gmail.com*

Mostafa Khakbaz

*Department of Electrical Engineering,
Islamic Azad University,
South Tehran branch, Tehran, Iran*

Fatemeh Tarameshlu

*Ministry of Education,
Iran
Univers90fa@gmail.com*

Abstract: Sensor networks are condensed wired or wireless networks for gathering and distributing environmental data. Sensors allow machines to capture and detect characteristics of physical substances and features of natural incidents. Most of the current efforts on sensor networks are focused on networking and service development for numerous applications, but less on processing the emerging data. Sensor networks generate immense amount of data, which requires advanced analytical processing and interpretation by machines. Processing and interpretation of huge amounts of heterogeneous sensor data and utilizing a coherent structure for this data is an important aspect of a accessible and interoperable sensor network architecture. In this paper, we describe a new semantic hierarchical sensor data storage named SemHD, which organized sensors in hierarchical form and each sensor send their data to cluster ahead in semantic model.

A Novel Scheduling and Service Provisioning Strategy for On-Demand Data Packet Delivery to Heterogeneous Wireless High Speed Railway Network

Shah Manshi

*L.D. College of Engineering,
Ahemedabad, India*

Monali Prajapati

*Government Polytechnic,
Ahemedabad, India*

B. S. Sedani

*L.D. College of Engineering
Ahemedabad, India*

K. R. Parmar

*V. G. E. C.,
Ahemedabad, India*

Abstract-In order to meet customer expectations, enable flexible train configurations and reliable delivery of enhanced services, industrial wireless technology for High Speed Rail Communication needs to be exploited. In this paper the challenging issues of providing desired on-board seamless internet service is discussed and a novel resource allocation strategy is proposed here for different quality of services provisioning over time varying channel conditions. A series of track-side access points (TAPs) with high-speed data rates along the rail lines is deployed as one of effective solution and the theoretical analysis of delay performance considering TAP association scheme is discussed. From the simulation results, we verify the optimality of earlier proposed scheduling algorithms that significantly reduces End-to-End delay by introducing TAPs in the heterogeneous network and more effective scheduling strategy for on-demand packet delivery to investigate the tradeoff between individual fairness and high system performance is suggested for delay sensitive applications.

Study of an Approach Using Agile Software Development Methodology with Scrum

Purvi Sankhe

Department of Information Technology,
Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Ratna Nayak

Department of Information Technology,
Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Shruti Mathur

Department of Information Technology,
Thakur College of Engineering &
Technology
Kandivali East, Mumbai.

Neha Patwari

Department of Information Technology,
Thakur College of Engineering & Technology
Kandivali East, Mumbai.

Nishtha Mathur,

Department of Information Technology,
Thakur College of Engineering & Technology
Kandivali East, Mumbai.

Abstract: Customer is the main focus of software development, customer satisfaction plays vital role in development and deployment of software product so now a days companies are aiming to adopt customer centric methodologies which will help to produce valuable software in short span of time within available budget. Agile Software development methodology is one of the software development approach which helps to best meet changing requirements of customers. This paper provides introduction of agile methodology and different agile models thereby concentrating on Scrum.

Achievable uplink capacity with linear detectors and Tradeoff Between Spectral Efficiency and Energy Efficiency for Massive MIMO system

Maharshi K. Bhatt

*Electronics and Communication Department,
L.D. College of Engineering,
Ahmedabad, India*

Bhavin S. Sedani

*Electronics and Communication Department,
L.D. College of Engineering,
Ahmedabad, India*

K. R. Parmar

*Electronics and Communication Department,
V. G. E. C.,
Ahmedabad, India*

Abstract— This paper performs analysis of massive MIMO system with uplink (UL) and downlink (DL) of multi-cellular time-division duplexing (TDD) systems, assuming that base station (BS) have N number of antennas and user terminals (UTs) have K number of antennas per cell are large. This system model consider for channel estimation, path loss, pilot contamination and antenna correlation for each link. We have derived in this work, approximations of achievable rates with several linear detectors like MF (Matched filter) and MMSE (Minimum Mean Square Error). We simulated achievable rate versus number of base station antenna N for MF and MMSE. We observed that performance of MMSE detector is superior in terms of achievable rate compared to MF detector in massive MIMO system. Second, we analyze spectral efficiency and energy efficiency of massive MIMO system with imperfect CSI (Channel State Information measured from uplink pilots). In this analysis we have used MRC (Maximum Ratio Combining) detector and derived spectral efficiency and energy efficiency for different values of transmitted power. From simulation we can observe that as we increase transmitted power, spectral efficiency increased but at the cost of energy efficiency. So there is always trade off between spectral efficiency and energy efficiency. We also observed from simulation that as we increase number of base station antenna trade off curve shifts upwards.

A Review on Modulation Techniques Used for Parallel Operation of Inverters with Droop Control of Voltage and Frequency

Rakesh S. Kushwah
 Electrical Engineering
 TCOE Boisar
 Mumbai, India
 rakesh.kushwah027@gmail.com

AkramPathan
 Electrical Engineering
 TCOE Boisar
 Mumbai, India
 akramkhan452@gmail.com

Talha Saleh
 Electrical Engineering
 TCOE Boisar
 Mumbai, India
 salehtalha@yahoo.com

Abstract—In this paper a control strategy using a three levelspace vector modulation (S VM) for two parallelly connected distributed generation units is presented. While connecting DG units parallelly load sharing between them becomes important factor. This controlling technique is basically based on P-F/Q-V droop control. P-F/Q-V droop control using S VM is helpful in maintaining the voltage and frequency stability as well as sharing the load proportionally from individual inverter. This method of droop control is implemented and compared with the P-F/Q-V droop control using S PWM. The simulation regarding the same is implemented in Matlab / Simulink and the results indicate that S VM based droop control is more effective than S PWM based droop control.

Design of Guest Tagging System and Surveillance Using RFID

Satyendra Yadav

*Thakur College of Engineering and Technology
University of Mumbai, India*

Akash Singh

*Thakur College of Engineering and Technology
University of Mumbai, India*

Lokesh Ucheniya

*Thakur College of Engineering and Technology
University of Mumbai, India*

Deven Shah

*Thakur College of Engineering and Technology
University of Mumbai, India*

Abstract — Considering the present scenario at Govardhan Eco Village, it has a manual way of authorizing the visitors by providing them simple Identity cards for authentication purposes. Being a 70-acre sustainable farming community, Govardhan Eco Village (GEV), doesn't have an adequate way of controlling the access of the visitors due to the large area and require something which can automatically do it. Around a few hundred guests visit daily and the village is not so secured. To overcome this real-world issue proposed solution is "IoT Based Guest Tagging System". RFID Reader only reads the data through RFID tag but through IoT all data entries will be collected & it will be stored online on remote server. Analysis will be performed on it so as to manage the flow of guests in different zones effectively, reduce the waiting time and serve the visitors in a more convenient manner. The proposed system uses a RFID passive tag and RFID reader to record the current location and timestamp of the visitors (guests). The main idea behind implementing IoT Based Guest Tagging System is to automate the guest monitoring process. It will enhance the existing security system of the Govardhan Eco-Village and also be able track down visitors who are crossing the zones where they are not allowed to go.

Work-Stake Hybrid: A Trustless Consensus Mechanism for Blockchain

Yash P. Chetnani
Computer Engineering
V.E.S.I.T., Mumbai University
Mumbai, India
yash.chetnani@ves.ac.in

Nupur Giri
Computer Engineering
V.E.S.I.T., Mumbai University
Mumbai, India
nupur.giri@ves.co.in

Abstract-The authors aim to propose a new mechanism for trustless consensus in peer-to-peer networks using Blockchain. The previously used techniques of Proof of Work and Proof of Stake have been combined into a new parameter called the ‘Cost factor’ for a more efficient system. The Cost factor will now be the criterion the miners will have to fulfill in order to be eligible to mine new blocks. This new parameter takes into account both, a miner past performance, and its current capacity to determine its right to mine a block. The block size and buffer period have been thus introduced to achieve the objective of this paper. The proposed system leads to faster confirmations, reduced operating costs and immunity against the Sybil attack.

Operating System Support for Virtual Machines

Amit G Mistry

Thakur College of Engineering and Technology
Mumbai

Bijit Marakarkandy

Thakur College of Engineering and Technology
Mumbai

Abstract:

A virtual-machine monitor (VMM) is a useful technique for adding functionality below existing operating system and application software. One class of VMMs (called Type II VMMs) builds on the abstractions provided by a host operating system. Type II VMMs are elegant and convenient, but their performance is currently an order of magnitude slower than that achieved when running outside a virtual machine (a standalone system). In this paper, we examine the reasons for this large overhead for Type II VMMs. We find that a few simple extensions to a host operating system can make it a much faster platform for running a VMM. Taking advantage of these extensions reduces virtualization overhead.

Share Market Analysis for Share Selection using Data Mining Technique

Munaf Patel

*Thakur College of Engineering and Technology
Kadivali, Mumbai*

Zahir Alam

*Thakur College of Engineering and Technology
Kadivali, Mumbai*

Abstract -Share market has been an field of vast interest both for those who wish to make money by trading shares in the share market. Generally there is an opinion about share market like high risk and high returns. Even though we have a huge number of potential investors, only very few of them are invested in the share market. The main purpose is thenot able to take risk of taking skill of investors. Though get low returns they want to save their money. One important reason for this problem is that, they don't have a proper guidance for making their portfolio. In this paper we focus the real world problem; we had selected three indices such as SENSEX, NIFTY. The analysis is purely based on the data collected from past three years. The Data mining technique, Time series interpretation is applied for the Data analysis to show the ups and downs of a particular index. The correlation and Beta are the tools which gives the suggestion about the share and its risk. The correlation tool is used to identify the relationship between the index and company individually. This Beta is used to identify the risk associated with the share.

Blockchain Technology and Allied Research

Sudhir Dhekane

*Information Technology Department
Thakur College of Engineering & Technology
Mumbai, India
sudhirdhekane1987@gmail.com*

Sandip Bankar

*Information Technology Department
Thakur College of Engineering & Technology
Mumbai, India
sandipbankar1@gmail.com*

Abstract— Blockchain is a decentralized transaction and information retrieval technology started mainly for cryptocurrencies. Blockchain technology growth can change many aspects of life. Blockchain technology provides high protection for transactions, authenticity for our data and data integrity makes it a better approach for safer transactions, personal information stored in this will transformed into a numeric value(hash) which is difficult to identify . Bitcoin and such cryptocurrencies can serve a huge market for currency exchange at global market thus leading more interlinking of world economy. Cryptocurrency can be sub divided into currency, security and commodity. Creating and managing such vast and decentralized database is a big challenge in itself with its own flaws having some disadvantage too. Our main objective is to study about working of blockchain technology and to discuss research work of the same presented by different authors.

Keywords-*Blockchain, Bitcoin, Cryptocurrency*

Author Index for IC-CSOD 2018

A

Aaditya Desai	346-352
Aakash Tripathi	225-229
Aanchal Singh	27-30
Abdul Samea	1-4
Abhijit Chaudhary	212-215
Abhileen Pandey	247-249
Abhishek Naik	208-211
Abhishek Pathak	238-241
Abhishek Singh	27-30
Abuizer Malvi	1-4
Aditya Gupta	287-290
Aditya Rao Nuty	234-237
Adnan Husain Sayed	140-143
Aishwarya Nair	364-366
Ajinka Padave	188-191
Akanksha Chand	389-392
Akanksha Das	353-356
Akash Naik	361-363
Akash Singh	53-58
AkramPathan	152-156
Akshata Desai	238-241
Alireza Chaviwala	31-34
Aman Cheval	208-211
Amit Singh	13-16
Anand Ranka	364-366
Anil Vasoya	35-38
Anil Vasoya	122-126
Ankit Gupta	13-16
Ankita Gaud	361-363
Arpit Vakharia	230-233
Arvind Gupta	204-207
Ashish Soni	99-103
Ashwini Nakil	63-67
Ataul Islam	305-308
Atulkumar Mishra	212-215
Avani Vasant	262-268
Ayush Dutta	201-203
Ayush Kedia,	234-237

B

Bharat Paliwal	367-370
Bhavesh Mehra	166-168
Bhavesh Radhika Gokani	166-168
Bhushan Nemade	9-12
Bhushan Nemade	353-356
Bijith Marakarkandy	80-82

C

Chaitanya Roygaga	277-282
-------------------	---------

Chanda Jha	361-363
Charvika Banjan	122-126
Chirag Rai	234-237
D	
Darshana Patel	262-268
Deep Modi	49-50
Deepak Yadav	378-380
Deepti Chavan	9-12
Deven Shah	17-22
Deven Shah	53-58
Deven Shah	169-174
Deven Shah	242-246
Dharti Singh	393-397
Dhiraj Patil	341-345
Dinesh Panchi	96-98
Divya Sharma	230-233
G	
G. R. Sinha	257-261
Gaurav Gianani	327-329
Gayatri Mulchandani	327-329
Govinda Borse	295-297
Gunjan Deshmukh	295-297
H	
Harish Baheti	238-241
Harsh Bharucha	375-377
Harsh Doshi	117-121
Harsh Parekh	117-121
Harsh Singh	381-383
Hetal Amrutia	35-38
Hetal Amrutia	117-121
Hitesh Seedani	96-98
Huney Pandita	31-34
Huzaifa Saboowala	135-139
I	
Isheta Tilara	334-337
J	
Jagdish Choudhary	320-322
Jasmina Ambani	63-67
Jatin Sajnani	96-98
Jayshree H.Sali	152-156
Jewel Correia	242-246
Juhি Vadia	230-233
Jyoti Vikram	389-392
K	
Kamal Shah	35-38
Kamal Shah	59-62

Kameshwar Gupta	353-356
Karan Pawar	287-290
Kartik Iyer,	117-121
Kautilya Joshi	49-50
Khan Shaheer	1-4
Kirtan Patel	99-103
Kirti Vishwakarma	371-374
L	
Lokesh Ucheniya	53-58
Lutuf Islam	305-308
M	
Mahendrasingh Parmar	225-229
Manasi Kadam	161-165
Mandar S. Shinde	144-147
Mary Margarat	27-30
Mary Margarat Valentine Neela	83-85
Maulik Shah	99-103
Mayank A. Mehta	144-147
Meet Gadoya	357-360
Meghna Singh	247-249
Mehdi Gheisari	269-276
Mehrnoosh Bazrafkan	269-276
Mirza Zainab	131-134
Misal dave	175-179
Moeenuddin Mohammed Mohsin	298-304
Mohammed Ahmed Abdul Mannan	140-143
Mohammed Umair Shaikh	283-286
Mufaddal Kapadia	135-139
Mustafa Poonawala	135-139
N	
Namdeo Badhe	68-70
Nazneen A. Pendhari	127-130
Neha Kapadia	23-26
Neha Kapadia	39-42
Neha Patwari	75-79
Neha Patwari	157-160
Neha Patwari	313-315
Neha Patwari	367-370
Nehal Ansari	283-286
Nemish Gandhi	68-70
Nishita Desai	216-219
Nishtha Mathur	75-79
Nishtha Mathur	313-315
NishthaMathur	157-160
Nita Patil	148-151
Nupur Giri	96-98
Nupur Giri.	5-8

P

Padmaja Shukla	389-392
Pankaj Malviya	49-50
Parth H Parikh	80-82
Pooja Kammala	161-165
Pooja Mistr	68-70
Pooja Yadav	371-374
Prachi Jain	169-174
Pradeep Kotian	309-312
Prajakta Bhosale	148-151
Pramod Chauhan	338-340
Pranay Rao	13-16
Pranjali Aswani,	208-211
Pranjali Kasture	247-249
Pranjali Kasture	291-294
Prashant Chaudhary	309-312
Pratik Mishra	353-356
Pravin Jodha	320-322
Prince Jain,	320-322
Priyanka Verma	393-397
Purvi Sankhe	75-79
Purvi Sankhe	157-160
Purvi Sankhe	225-229
Purvi Sankhe	313-315

R

Rakesh S. Kushwah	152-156
Radhika Gokani	166-168
Radhika Kotecha	287-290
Radhika Kotecha	393-397
Raees Ahmad Noor Mohammad	298-304
Raghavendra R. Sedamkar	127-130
Rahatullah Khan Rahmatullah Khan	298-304
Rahul Vishwakarma	212-215
Raj Asha	357-360
Raj Singh,	247-249
Rajat Nareshkumar Khemka	323-326
Rajat Nareshkumar Khemka	384-388
Rajesh Bansode	43-48
Rajesh Bansode	63-67
Rajesh Bansode	86-90
Rajkumar Patra	257-261
Rakesh S. Shirasath	144-147
Ratna Nayak	75-79
Ratna Nayak	313-315
RatnaNayak	157-160
Ravi Sitapara	309-312
Rina Yadav	86-90
Rishab Dubey	13-16
Rishab Tiwari	99-103
Riya Parikh,	161-165
Rohan Shetty	393-397

Rohit Raja	257-261
Rohit Yadav	378-380
Rupesh Vishwakarma	27-30
Ruth Baptista	17-22
S	
S K Narayankhedkar	242-246
S.D. Sawarkar	148-151
Samad Ansari,	283-286
Sandeep I. Yadav	104-108
Sandip Bankar	381-383
Sangeeta Vhatkar	71-74
Sangeeta Vhatkar	104-108
Sangeeta Vhatkar	183-187
Sanjana Puthran	334-337
Satyendra Yadav	53-58
Saurabh Khandelwal	122-126
Saurabh Maydeo	357-360
Saurav Jangir	225-229
Seema Jamal	71-74
Seyed Mohamad Bagher Davoodi	269-276
Shamuwel Ahmed Ansari	140-143
Shashank Shukla	250-255
Shehbaz Ansari	140-143
Shital H. More,	109-116
Shivani Rege	180-182
Shivani Subnani	327-329
Shreya Mistry	216-219
Shreysijay Singh	371-374
Shridhar Kamble	91-95
Shridhar Kamble	109-116
Shridhar Kamble	309-312
Shruti Mathur	75-79
Shruti Mathur	313-315
ShrutiMathur	157-160
Shubham Maheshwari	91-95
Shweta Mishra	334-337
Shwetha Yadav	316-319
Siddesh Rao	192-196
Siddhant Acharya	220-224
Sneha Yadav	330-333
Su Su Hlaing	257-261
Sudhesh Singhania	220-224
Sudhir Dekhan	109-116
Sunil Chaurasia	338-340
Sunil Rai	346-352
Suraj Punjabi,	277-282
Suresh Rajpurohit	381-383
Swapnil Sampat	277-282
Swathi Yadav	59-62
Swati Abhang	109-116

T

- Tanay Sanghavi 216-219
Tanishq Mehra 91-95
Tanuja Sarode 277-282
Tejas Sutone 375-377

U

- Ujjwal Sharma 220-224

V

- V.S Dhaka 131-134
Vamil Gandhi 122-126
Vedprakash Jha 367-370
Veena Kamat 68-70
Veque Ansari 31-34
Vijay Karotra 91-95
Vijay Varma 378-380
Vikas Kaul 242-246
Vikas Kaul 330-333
Vikas Kukreja 96-98
Vikram Deshmukh 364-366
Vinayak Sharma 327-329
Vinesh Bagul 43-48
Vinesh Yadav 338-340
Vivek Kumar 367-370

W

- Waheeda Dhokley 283-286

Y

- Yamisha Savalia 63-67
Yash Kosambia 381-383
Yash Mujumdar 197-200
Yash P. Chetnani 5-8
Yogesh S. Kolhe 144-147

Z

- ZahIr Aalam 341-345
Zahir Aalam 49-50
Zainab Mirza 1-4

