||Jai Sri Gurudev||

Sri Adichunchanagiri Shikshana Trust ®

SJB Institute of Technology

BGS Health and Education City Kengeri, Bengaluru

Proceedings of 14th National Conference On "Integrated Intelligent Computing, Communication & Security" [NCIIC-2019]

04th May 2019

Prepared By

Department of Computer Science & Engineering

SJB Institute of Technology

BGS Health and Education City, Uttarahalli Road,
Kengeri, Bengaluru-60
Website: www.sjbit.edu.in

ABOUT THE TRUST & INSTITUTION

Established in 1974, Sri Adichunchanagiri Shikshana Trust (R) is guided by the commitment to excellence, under the Blessings of His Divine Soul Jagadguru Padmabhushan Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji. Importance is given to impart three components of education – intellectual, spiritual & physical needs of students. With this vision, the trust has started 500 educational units across the country, benefiting more than one lakh students every year. Presently, the trust is being administered by the Chief Pontiff, His Holiness Sri Sri Sri Dr. Nirmalanandanatha Maha Swamiji. Beneath the trust, BGS Group of Institutions has established many institutions right from Nursery, High School, Pre University College, Pharmacy, Ayurvedic, Management, Engineering, Medical etc., Revered Sri Sri Dr. Prakshnath Swamiji, Managing Director, SJB & BGS Group of Institutions has diligently associated with Sri Adhichunchagiri Shikshana Trust.

SJBIT is established in the year 2001 with the blessings of his divine soul, Jagadguru Padmabhushan Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji of Sri Adichunchanagiri Shikhshana Trust. SJBIT is one among the 500 educational institutions being managed by the trust. Spreading across sprawling 50 acres of land; the SJBIT is located amidst natural serene atmosphere, 14 km from Bengaluru city, near Kengeri. SJBIT is affiliated to Visvesvaraya Technological University, Belgavi and approved by AICTE, New Delhi. SJBIT is ranked No. 1 in Karnataka, (Among VTU affiliated Private Engg Colleges); Ranked No.5 in south India based on all India Data Quest-Cyber media research' survey – May 2015. SJBIT has been honoured as Excellent Training and Placement Institute in Karnataka" by VTU, AICTE, CMAI, AIU, Govt. of India. The institution has own the national employability award during 2013-14. Accredited by NAAC with 'A' grade 3.22 CGPA during the year 2017.

The institution offers 6 under graduate programmes, 7 post graduate programmes and 10 departments have been recognized as research centres.

ABOUT THE DEPARTMENT

Computer Science and Engineering Department was established in the institution during the year 2001, with an intake of 60 and increased to 120 in the 2008 later 180 in the year 2011. Being one of the core departments, it is headed by Dr. Krishna A N. The Department is strengthened with qualified Doctorates, Professors, Associate Professors, Asst. Professors and Technical staffs.

The department has well equipped laboratories with state-of-the-art infrastructure, department library and internet facilities. Supported by experienced & dedicated staff, the department offers B.E in CSE and M.Tech in CSE and Ph.D programs.

ABOUT THE CONFERENCE

The tremendous growth in the field of Computer Science and Technology over the past decades has led to challenging applications in various topics. The goal of the conference is to bring together technical persons from different organisations, universities and industries for exchanging ideas and experiences among them.

Conference Objectives:

The conference NCIIC aims -

- To present the overall perspective of the recent development, challenges and emerging trends in the field of computing**.**
- To bring together researchers, practitioners to put forth their challenging innovations and progress strategies.
- To provide an area for the budding technocrats expose their new visions.

Divine Blessings

His Divine Soul Jagadguru Padmabhushan Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji
Founder President, Sri Adichunchanagiri Shikshana Trust ®

His Holiness Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Maha Swamiji
President, Sri Adichunchanagiri Shikshana Trust ®

Revered Sri Sri Dr. Prakashnath Swamiji
Managing Director, SJB & BGS Group of Institutions, Bengaluru

Message

His Holiness Jagadguru Sri Sri Sri
Dr. Nirmalanandanatha Maha Swamiji

President, Sri Adichunchanagiri Shikshana Trust ®

It is a matter of pride that the Computer Science and Engineering department is organizing "**14th National Conference on Integrated Intelligent Computing, Communication and Security NCIIC-2019"** which aims at bringing together researchers, professionals and industry experts from all states across the country and renowned delegates from major industries and institutions in India, who deliberate and discuss about the latest technological developments.

I am confident that under the able guidance of the managing director Revered **Sri Sri Dr. Prakashnath Swamiji** and at the dynamic leadership of the Principal of the college **Dr. Puttaraju**, such endeavours will bring maximum exposure to the institution in national arena for it to grow from strengths and to meet the challenges of tomorrow.

I would like to express my appreciation to the organizing committee and hope the participants learn much, remain inspired to champion new efforts, and commit to sharing your future successes at every opportunity and wish the conference a grand success.

(Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Maha Swamiji)

Message

Revered Sri Sri Dr. Prakashnath Swamiji

Managing Director,
SJB & BGS Group of Institutions, Bengaluru

I am exceptionally glad to know that all the departments of SJB Institute of Technology are conducting National Conference at its campus on 4th MAY 2019.

I am contented to note that Computer Science and Engineering department is actively involved in organizing **"14th National Conference on Integrated Intelligent Computing, Communication and Security NCIIC-2019"** with the aim of bringing together research scholars, professionals and industry experts from frontline states across the country and the experts from major industries in India to discuss the interlinks of different sectors of computer science and engineering.

I hope that this platform will provide an excellent exposure to the participants and research scholars to impart their knowledge on their research areas. I am sure that the institution will be able to showcase much bigger events in the years to come and take the name of the institution to greater heights under the able guidance and dynamic leadership of the principal **Dr. Puttaraju**. My wholehearted wishes to the organizers, participants for the fruitful deliberations and the conference a grand success.

(Revered Sri Sri Dr. Prakashnath Swamiji)

Message

Dr. Puttaraju

Principal, SJBIT

SJB Institute of Technology comes under the aegis of Sri Adichunchanagiri Shikshana Trust and was started in the year 2001. For the past 18 years since the inception, we have been striving to raise the standards in terms of the quality of the students as well as the faculty strength. I would like to thank Management of Sri Adichunchanagiri Shikshana Trust for lending support at all levels.

I am immensely happy that Department of Computer Science & Engineering is organising “**14th National Conference on Integrated Intelligent Computing, Communication and Security NCIIC-2019”** on 4th May, 2019. The conference aims to bring different ideologies under one roof and provide opportunities to exchange ideas on recent technological advancements. The themes of this conference are indicative of relevant research areas to give the prospective authors innovative prepositions about the ambit of discussion.

I also congratulate the HOD, teaching and non-teaching staffs, students and participants from other colleges in making this conference a grand success. I thank all the contributors, reviewers and experts for their unconditional support towards the conference.

Dr. Puttaraju

Message

Dr. Krishna A N

Professor & Head
Dept of CSE, SJBIT

Each day the dependability on computer and its technologies on daily life have been increased enormously. Advancements in the intelligent computing activities are directed towards economic advancements and upliftment of the society. The ability of systems connectivity and communication between them intelligently and securely is a big challenge. These challenges have encouraged a lot for researchers to work on various innovative technologies.

This "**14th National Conference on Integrated Intelligent Computing, Communication and Security NCIIC-2019"** is a unique conference on Intelligent Computing has attracted 82 papers from around the country. After the peer review process in which paper was reviewed by two reviewers, finally 35 papers were accepted. This National conference aims at bringing together the researchers from academicians, research scholars and members of the industry across the country to allow understanding the challenges faced in these technological perspective. I wish all the participants to make use of this opportunity and make this conference a grant success

I would like to congratulate the Advisory Committee, Technical Program Committee, Conference Coordinators and Organizing Committee Members for Successfully Organizing this Conference. I am thankful to the entire teaching and non teaching faculty members of CSE Department for their support and efforts in bringing this proceedings and making this conference a success.

Dr. Krishna A N

Committees for NCIIC - 2019

Blessings:

His Divine Soul Jagadguru Padmabhushan
Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji
Founder President, Sri Adichunchanagiri Shikshana Trust(R)

Chief Patron:

His Holiness Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Maha Swamiji
President, Sri Adichunchanagiri Shikshana Trust(R)

Patron:

Revered Sri Sri Dr. Prakashanatha Swamiji
Managing Director, SJB & BGS Group of Institutions, Bengaluru

ORGANIZING COMMITTEE

Conference Chairman:

Dr. Puttaraju, Principal, SJBIT, Bengaluru

Conveners:

Dr. Krishna A N, Professor & HOD, Dept. of CSE, SJBIT

Dr. Srikantaiah K C, Professor, Dept. of CSE, SJBIT

Advisory Committee:

Dr. L M Patnaik, IISC, Bengaluru

Dr. Venugopal K R,Vice Chancellor,Bangalore University

Dr. B G Prasad, BMSCE, Bengaluru

Dr. C B Akki , IIIT, Dharwad

Dr. P Deepa Shenoy, UVCE, Bengaluru

Dr. Rajan M A, TCS, Bengaluru

Dr. T G Basavaraj, SKSJTI, Bengaluru

Dr. Ravishankar S, Vidya Vikas, Mysusru

Technical Program Committee:

Dr. K G Mohan, Presidency University

Dr. Chandrashekar S N, CBIT,Kolar

Dr. R Balakrishna , RRCE,Bangaluru

Dr. G T Raju , RNSIT, Bangaluru

Dr. Thimmaraju, VTU PG Studies, Mysuru

Dr. Ramesh babu, DSCE, Bangaluru

Dr. Siddaraju, Dr.AIT, Bangaluru

Dr. Mustafa,BEARYS, Bangaluru

Dr. D V Ashoka,JSSATE, Bangaluru

Dr. Srinivasa Naik C L,UBDTE, Davanagere

Dr. D V Ashoka,JSSATE, Bangaluru

Dr. Yogisha H. K., MSRIT, Bengaluru

Dr. Arun Biradhar, EWIT, Bengaluru

Dr. Sagar , RVCE, Bengaluru

Co-Ordinators:

Dr. Naveena C, Professor, Dept. of CSE, SJBIT

Mr. Chandrashekar D K, Assistant Professor, Dept. of CSE, SJBIT

Organizing Committee:

Mrs. Nirmala H

Mr. Shantha Kumar H C

Mrs. Chaitra H K

Mrs. Chandana C

Mrs. Chaitra M

Mrs. Shubha T V

Mrs. Prakruthi M K

Mrs. Manasa B S

Mrs. Manjula H S

Mrs. Basamma Patil

14th National Conference on Integrated Intelligent Computing, Communication and Security (NCIIC-2019)

Copy Right and Reprint Permission

Note: Papers are accepted and printed in this proceeding with the understanding that the same has been neither submitted nor published elsewhere. Concurrent submissions to other publications or journals are viewed as serious breach of ethics. The authors are responsible for obtaining copyright and other clearance prior to submitting materials for consideration and acceptance.

Views, ideas and designs expressed in the articles/papers are solely from respective authors. Organizers are not responsible for authentication of the articles/papers published in this proceeding. The authors are solely responsible for legal matter resulting out of violation of copyright act. The proceeding is being published in the interest of author and research community for private circulation only.

Table of Contents

A compendium of the technologies in monitoring vehicular emissions

Aparna S Bandekar[1], Divya Chatty[2], Mayuri M Hangarge[3], Manjuprasad B[4]
8th Semester B.E Student[1, 2, 3], Assistant Professor[4]
Department of Computer Science and Engineering
GSSS Institute of Engineering and Technology for Women, Mysuru, Karnataka
Affiliated to VTU Belagavi, Karnataka, INDIA
aparna.bandekar.ab@gmail.com[1], divyachatty@gmail.com[2], hangargemayuri1234@gmail.com[3] manjuprasad32@gsss.edu.in[4]

***Abstract*— One of the critical issues in today's scientific world is the need to tackle climate change. A major contributor to global warming is vehicular emissions. India ranks third in the world's list of vehicular emissions with 596 million tons of greenhouse gases being produced at a 43% rate of increase each year. According to the present system, all vehicles need to get a periodic emission and an annual fitness check for safety and road-worthiness. Even then, the objective of cutting down pollution is unachieved to a large extent due to lapses at various levels. The system to track the vehicles that are failing to meet norms is non-existent. The main aim of this paper is to identify and analyze the various existing solutions and methods which exist to monitor and control the vehicular emissions. And this paper also presents the basic sensor requirements for understanding the implementation aspects in vehicular emissions.**

Keywords—Vehicular Emission; ThingSpeak; Data Analysis; Report Generation; Internet of Things; Raspberry Pi; Predictive Model; Participatory Sensing.

I. INTRODUCTION

Vehicular emission is an important concern for every country of the world, especially for a developing country like India. A major contributor to global warming is vehicular emissions. India ranks third in the world's list of vehicular emissions with 596 million tons of greenhouse gases being produced at a 43% rate of increase each year. India's carbon dioxide emissions had been projected to increase at a rate of 6.3% by 2018. The 2018 Global Carbon Project report attributes this increase to strong economic growth of around 8% per year. [1]
Currently the Bharat Stage IV Emission Norms is in place, which is equivalent to the European emission standards. Bharat Stage Emission Norms VI is to be enforced by 2020 which focuses on "On-Board Diagnostics (OBD)". [2] It mainly aims to reduce the carbon footprint from individual vehicles and creating awareness on pollution levels in public.

Presently, all vehicles need to undergo a periodic emission and an annual fitness check for safety and road-worthiness in the pollution under control (PUC) centers, where the vehicle under inspection is examined for the emission levels by using the sensing tools approved by the pollution control board. The image of the vehicle's number plate is captured to start the session. After which, the examiner issues an emission certificate which will certify the emission from the vehicle is within the range. It will also contain other parameters such as the vehicle number, owner's name, validation period and many other parameters.

The objective of reducing pollution is not achieved to a large extent by the present system due to lapses at every level due to mismanagement and negligence of every stakeholder involved. A system to track the vehicles failing to meet emission norms is non-existent. Therefore, there is an undeniable need for technological advancements in the methods used to validate a vehicle on the grounds of emission.

This paper presents a comprehensive study on the recent trends in achieving On-Board Diagnostics to tackle vehicular emissions. The results and learning gleaned from various sources are summarized to give a clear understanding of the use of Internet of Things, Machine Learning and Cloud Computing to solve this problem. Section I gives a concise introduction to the problem. Section II presents a summary of the papers surveyed. Section III illustrates a comparative study of the papers that have been surveyed. Section IV presents the key technologies for On-Board Diagnostics. Section V draws logical conclusions regarding the trends in monitoring and controlling vehicular emissions.

II. LITERATURE SURVEY

This section presents a detailed discussion on the various papers surveyed on the domain of monitoring vehicular emissions.

Gupta et al., proposed a system assuming road lanes are in coordinate geometry. It uses Arduino for processing the sensor data, ADC to convert the analog signal to digital signal, Raspberry Pi to control the camera at the signal light and ZigBee networks to connect the sensors. The model collects sensor data from the sensors placed in the vicinity of the traffic signal. The defaulter vehicle is identified by capturing the image of the number plate, employing DIP for the further process on Microsoft Azure or Amazon Web Services. [3]

Saha et al., presents a project which employs Raspberry Pi 3B, MQ135 Gas Sensor to sense NH3, NOx, Smoke and CO2 gases, M393 as sound sensor. The Internet connectivity needed for the functionality of the IoT system is made available from the mobile data, the associated SIM card is embedded into the system. It also suggested to notify the user through mail. It also provides a web interface to the user for analyzing the emissions from the vehicle. [4]

Kulkarni et al., in their project use the ARM 7 processor to process the data collected near the exhaust of the vehicles by the smoke sensor MQ2. If the maximum threshold is violated then the vehicle is brought to halt by manipulating the fuel injector. The driver is notified about the halt action by a buzzer sound indicating motor shut down. The model also includes an LCD to display the pollution levels. Notification is sent to the nearby pollution control office through SMS. [5]

Mehta et al., implemented an air quality detection, analysis and prediction system using the cloud domain. The system would propose an alternate route to the user where he would encounter least pollution. The analysis of the pollution level in a particular area was achieved using sensors, video processing and pattern matching algorithms. Edge and contour detection algorithms are incorporated to classify the identified objects as vehicles. [6]

Chandrasekaran, et al., proposed a system which employed semiconductor sensor to detect the gases from the vehicle. The sensing module is specifically placed near to the exhaust of the vehicle to read accurate pollutant values. The system has predefined values of tolerance for the sensor values, on violation of the threshold the system is initialed to halt by manipulating the fuel supply to the engine thereby providing the user with ample time to park the vehicle. The user is also sent the location of the nearest service station location to get the vehicle serviced. [7]

Samsudin et al., in their project leveraged the ThingSpeak cloud to log the data from the humidity and temperature sensors mounted on the Raspberry Pi device. DHT11 Temperature and Pressure sensors were used. The main focus was on data visualization using the ThingSpeak cloud. It also demonstrated the utilization of the Twitter API to post tweets directed at the user in case the threshold value is crossed. [8]

Kumar et al., proposed an air pollution controlling system using Raspberry Pi and Arduino. Various sensors like DSM501A, DHT22, BMP180, MQ9 and MQ135 have been used to detect particulate matter, Temperature, Humidity and Pressure Carbon Monoxide and Carbon Dioxide. The visualization of the pollution level and the monitoring was achieved using the Device Centric Analysis feature of the IBM Watson IOT Platform [9]

Mhatre et al., implemented a humidity and temperature monitoring system using Arduino UNO and Raspberry Pi. It used ThingSpeak to visualize the data so collected. It demonstrates how to connect sensors to a cloud using a wireless setup. [10]

Sheorey et al., built a prototype to demonstrate collision avoidance system on highway. The prototype used sensors to monitor the state of the vehicle and regulate its speed. It executed controlling mechanism on the vehicle by manipulating the axle. After the vehicle crossed the tolerance level of the system, it would be brought to halt using the aforementioned mechanism. [11]

Devi et al., in in their paper spoke about integrating the cutting-edge technologies such as WSN, ZigBee and Cloud. They propose a methodology to monitor vehicular emission using cloud. The system used Amazon Web Services for the cloud aspect of their proposed system. It also generated a comprehensive report of the emission level and produced it to the pollution control board. The board would further notify the user of the same in appropriate method. [12]

Kiruthika, R et al., proposed an IoT-based system that attempts to solve the problem of environmental pollution. It uses various sensors like gas density, humidity, temperature and soil moisture interfaced with a Raspberry Pi device. The observations are taken at periodic intervals. The machine learning model predicts upon the newly collected data to suggest relevant solutions to any problems that are detected. There is a reporting mechanism in place that will keep the users informed of the results of the prediction. [13]

Reshi et al., proposed a vehicular pollution monitoring platform, which uses Wireless Sensor Networks. It measures the concentration of different types of pollutants produced by the vehicle. This module is proposed to be placed inside the vehicle for accuracy. [14]

Siregar et al., proposed a paper to detect various gases in urban area. The system proposed bound the WSN technology, 3G network and Wasp Motean open source sensing node together to create a firm system. The data is processed on-board and the result is communicated to the cloud and the storage unit using appropriate protocols. It also mentions about a web interface where the user is notified the violation of the rules. [15]

Caya et al., implemented a system using a Raspberry Pi, Gizduino mini microcontroller. It basically uses the dust sensor to record the suspended particulate matters and gas sensors to record CO emissions levels. After the computation is performed on the data that is read from the sensor, a notification is sent via email to the intended user informing him about the values read. [16]

Phala et al., in this project implements a real-time monitoring system using MQ2 and humidity sensor. These sensors are placed inside the vehicle to read an accurate value. The vehicle is evaluated against the predefined values of threshold. If the values are violated, then the defaulter vehicle is brought to halt using the fuel injection manipulation technique. The user of the system is provided with a web user interface to monitor and track his vehicles emission levels. It also speaks of reporting mechanism to the higher authority in regular interval. [17]

III. Comparative Analysis

This section presents the detailed analysis carried out for the identified survey papers in Section II. These analyses are classified according to the specific domain of the identified papers in terms of their features, inferences and identified future enhancements.

Table-1 presents the comparative analysis of all the papers which are related to the utilization of ThingSpeak.

Table 1: Analysis based on utilization of ThingSpeak

Paper	Salient Features	Inference	Future Enhancements
[8]	Humidity sensor, Temperature sensor ThingSpeak Twitter API	Connect Raspberry Pi and ThingSpeak	Control mechanism for the system
[10]	Arduino UNO Raspberry Pi ThingSpeak	Connect Sensor and Cloud via a wireless setup.	Implement a control mechanism.
[13]	IoT, WSN Raspberry Pi Cloud ThingSpeak Data Analysis Population Density.	Real time analysis of sensor data	Extend ThingSpeak capabilities to produce elaborate and verbose web interfaces.

Table-2 summarizes the salient features of all the papers related to various halting mechanism for the vehicle.

Table 2: Analysis based on different halting mechanisms.

Pape r	Salient Features	Inference	Future Enhancements
[5]	MQ2 ARM-7 Fuel Injection Manipulation Buzzer LED Display.	Halt vehicle using software.	Revocation of the vehicle, Data Collection, Analysis.
[7]	Semiconductor sensor Buzzer GPS ATMEL 89S52 Fuel Injector.	Halt vehicle using Fuel Manipulation.	Incorporate Analysis and Predictive Model.
[11]	Regulate speed, state of the vehicle Axle Manipulation.	Halt vehicle by Axle Manipulation	Mechanical Control over the vehicle.
[17]	WSN GSM module Sensor Array Sink Node Server.	Halt vehicle using Fuel Injection Manipulation.	Verbose way to halt and identify constraints in implementation

Table-3 summarizes the salient features of all the papers related to analysis on the collected sensor data.

Table 3: Summary of papers related to analysis on the collected data.

Paper	Salient Features	Inference	Future Enhancement
[4]	Raspberry Pi 3B MQ135 M393 GSM GPRS	Predictive models can be built.	Cloud based analysis.
[6]	Edge Algorithm Contour Detection Algorithm Blob analysis Video Processing Pattern Recognition	Real-time monitoring system.	Decentralized analysis and centralized monitoring.

Table-4 summarizes the salient features of all the papers incorporating the reporting mechanism.

Table 4: Summary of papers based on reporting mechanism.

Paper	Salient Features	Inference	Future Enhancement
[12]	WSN Zigbee Cloud based monitoring ARM microcontroller	Transfer real-time data to cloud using Zigbee network	Generate report on the analyzed data, implement conceptual model.
[13]	IoT WSN Raspberry Pi Cloud ThingSpeak Data Analysis Population Density.	Real time analysis of sensor data, Report generation.	Reduce dependency on population density.
[17]	WSN GSM module Sensor Array Sink Node Server.	Halt vehicle using the Fuel Injection Manipulation, Report higher Authority.	Provide a User Interface.

Table-5 summarizes the salient features of all the papers incorporating different mechanism to identify the defaulter vehicles.

Table 5: Summary of papers on the basis of the

mechanism used to identify the defaulting vehicle.

Paper	Salient Features	Inference	Future Enhancement
[3]	WSN Zigbee Cloud based monitoring ARM microcontroller	Transfer real-time data to cloud using Zigbee network	Generate report on the analyzed data, implement conceptual model.
[6]	Edge Algorithm Contour Detection Algorithm Blob analysis Video Processing Pattern Recognition	Real-time monitoring system. Identify the vehicle after Video Processing.	Intra vehicle sensing.

Learning from the Analysis:

The literature survey and analysis throw light on various technologies that can work together to achieve the mark of reading accurate data and acting upon them accordingly. Technologies such as image processing and video processing can be employed along with on-board sensors such as MQ7, MQ135, MQ9, MQ2 and dust sensors. Drastic actions can be implemented as actions on the emission data, this can include halting the vehicle. Various mechanisms can be used to manipulate the functionality of the vehicle, fuel manipulation and axel manipulation are some of them to mention.

From the cloud aspect, various vendors available in the market to provide the compute, storage and visualization of the data so collected from the vehicles. Some of them are IBM Watson, Amazon Web Services and ThingSpeak which is used specifically to visualize the data in terms of graph and plots.

IV. Key technologies for on-board diagnostics.

1) ThingSpeak:

ThingSpeak is a Cloud platform that is open-source and helps in connecting IoT devices and retrieving data from them. It facilitates the transfer of readings from sensors via the HTTP protocol over the Internet or via a Local Area Network. It logs the sensor values from the application and help in visualizing the data with appropriate graphs.

2) MQ7 sensor:

Figure 1: MQ7 Sensor

MQ7 is a sensor to detect the levels of carbon monoxide (CO) in the air. A typical MQ7 sensor can detect CO-gas concentrations in the range of 20 to 2000ppm.

3) MQ135 sensor:

Figure 2: MQ135 Sensor

It is a gas sensor with wide detection scope. It has fast response and is highly sensitive to the emissions. It operates at a voltage of +5V. It is capable of detecting NH3, NOx, alcohol, Benzene, smoke, CO2, etc.

4) Raspberry Pi:

Figure 3: Raspberry Pi

It is a 64-bit ARM based SBC (Single by Raspberry Pi Foundation. It runs Debian based Linux operating system named Raspbian and ports of many other OSes exist for this SBC. It also houses a Bluetooth and Wi-Fi modules on board.

5) MQ9:

Figure 4: MQ9 Sensor

It is a combustible gas sensor useful to detect gas leakage. It’s made up of highly sensitive SnO2 material, which has lower conductivity in clear air. It is highly sensitive to gases such as methane, LPG and propane.

V. Conclusion

On a broader aspect there are various methods and techniques available to monitor and control the vehicular emissions. All of these are individual entities which are powerful and capable for functioning to its fullest. The data so collected from the sensor after the analysis are not being utilized to its fullest, this data can be put to good use in terms of trend analysis, predictive analysis and to generate useful reports of the same. Currently the monitoring system and the controlling system are bound together to achieve a functional unit. The efficiency of this model can be improved many folds by incorporating the aforementioned technologies of predictive analysis and trend analysis.

Acknowledgments

This work was carried out at the “Innovative Smart Systems and IoT Laboratory” sponsored by Vision Group on Science and Technology (VGST), Govt. of Karnataka under the scheme of Centre of Innovative Science, Engineering and Education (CISEE). The authors would like to acknowledge the VGST, Dr. S. Meenakshi Sundaram, Professor and Head, Department of CSE and Dr. M. Shivakumar, Principal, GSSSIETW for their encouraging inputs and continuous support in carrying this work on.

VI. References

[1] Economic Times, [Online], available at [https://economictimes.indiatimes.com /articleshow/66963109 .cms?utm_source=contentofinterest&utm_medium=text&utm_c ampaign=cppst] accessed [6 December 2018]

[2] Embedded Blog [Online], available at [https://www.embitel.com/blog/embedded-blog /what-is-bharat-stage-6-and-bs-vi-norms-in-automotive-electronics] accessed [18 February 2019]

[3] Gupta, Karan, and Nitin Rakesh. "IoT Based Automobile Air Pollution Monitoring System." In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 14-15. IEEE, 2018.

[4] Saha, Arnab Kumar, Sachet Sircar, Priyasha Chatterjee, Souvik Dutta, Anwesha Mitra, Aiswarya Chatterjee, Soummyo Priyo Chattopadhyay, and Himadri Nath Saha. "A raspberry Pi controlled cloudbased air and sound pollution monitoring system with temperature and humidity sensing." In Computing and Communication Workshop and Conference (CCWC), 2018 IEEE 8th Annual, pp. 607-611. IEEE, 2018.

[5] Kulkarni, Anita, and Ravi Teja. "Automated System for Air Pollution Detection and Control in Vehicles." International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 3, no. 9 (2014): 12196-12200.

[6] Mehta, Yash, MM Manohara Pai, Sanoop Mallissery, and Shwetanshu Singh. "Cloud enabled air quality detection, analysis and prediction-a smart city application for smart health." In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1-7. IEEE, 2016.

[7] Chandrasekaran, Siva Shankar, Sudharshan Muthukumar, and Sabeshkumar Rajendran. "Automated control system for air pollution detection in vehicles." In *2013 4th International Conference on Intelligent Systems, Modelling and Simulation*, pp. 49-51. IEEE, 2013.

[8] Samsudin, Muhamad Fazril Afif, Roslina Mohamad, Saiful Izwan Suliman, Nuzli Mohamad Anas, and Hafizal Mohamad. "Implementation of wireless temperature and humidity monitoring on an embedded device." In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 90-95. IEEE, 2018.

[9] Kumar, Somansh, and Ashish Jasuja. "Air quality monitoring system based on IoT using Raspberry Pi."

[10] In *Computing, Communication and Automation (ICCCA), 2017 International Conference on*, pp. 1341-1346. IEEE, 2017

[11] Mhatre, Leena, and Neha Rai. "Integration between wireless sensor and cloud." In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 779-782. IEEE, 2017.

[12] Sheorey, Tanuja, and Nikhil Vivek Shrivas. "Development of sensor based front end collision avoidance system for highways." In *2015 IEEE International Conference on Information and Automation*, pp. 594-598. IEEE, 2015.

[13] Devi, P. Kausalya, and Naveen Ravichandran. "An automated cloud based vehicular emission control system using Zigbee." *In INTERNATIONAL CONFERENCE ON SMART STRUCTURES AND SYSTEMS-ICSSS'13, pp. 189-192. IEEE, 2013.*

[14] Kiruthika, R., and A. Umamakeswari. "Low cost pollution control and air quality monitoring system using Raspberry Pi for Internet of Things." In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 2319-2326. IEEE, 2017.

[15] Reshi, Aijaz Ahmad, Shabana Shafi, and A. Kumaravel. "VehNode: Wireless Sensor Network platform for automobile pollution control." In *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, pp. 963-966. IEEE, 2013.

[16] Siregar, Baihaqi, Ahmad Badril Azmi Nasution, and Fahmi. "Integrated pollution monitoring system for smart city." In *2016 International Conference on ICT For Smart Society (ICISS)*, pp. 49-52. IEEE, 2016.

[17] Caya, Meo Vincent C., Angeline P. Babila, Alyssa Moya M. Bais, Seoi Jin V. Im, and Rafael Maramba. "Air pollution and particulate matter detector using raspberry Pi with IoT based notification." In *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1-4. IEEE, 2017.

[18] Phala, Kgoputjo Simon Elvis, Anuj Kumar, and Gerhard P. Hancke. "Air quality monitoring system based on ISO/IEC/IEEE 21451 standards." IEEE Sensors Journal 16, no. 12 (2016): 5037-5045

Implementation of Artificial Neural Network Classification for Kidney Stone Detection

Rachitha.C[1], Rajeshwari.B.S[2], Sajeeda Khanum[3], Mrs.Ramya.B.N[4]
Asstistant Professor[4]
Information Science and Engineering
East West Institute of Technology, bengaluru, India
sajeedakhan686@gmail.com[1],aradhya.racchu@gmail.com[2],rajuvanivani@gmail.com[3],ramyanarayana2003@gmail.com[4]

***Abstract*— The mutation of the kidney can be perceived by sonographic pictures. The kidney may have deformities like kidney damage, change in its area and may additionally show up because of the shaping of stones. For doing medical procedure it is extremely basic to decide the particular and deliberate district of stone in the kidney. The sonographic pictures are of low inconsistency and have an undesirable commotion. This can prompt troubles in recognizing the issue. Hence pre-preparing of sonographic pictures are done to dispose of undesirable clamor. First reclamation is completed to limit the undesirable clamor and filtration is accomplished for smoothening. Also, ordinarily the histogram examination is accomplished for improvement. The resultant picture is partitioned into equivalent tomahawks utilizing level set division. Level set division includes impetus and strong term separately, which helps in contrasting and removing the highlights of kidney zones. We utilize FCM calculation and applying wavelet technique helps in getting the vitality levels from which the stones in kidney can be perceived. These vitality levels are extremely one of a kind from typical vitality level and are prepared by Multilayer Perceptron and Back Propagation Artificial Neural Networks to decide the assortment of stone .This encourages us to distinguish the stones in the kidney at beginning time and even exact moment stones can likewise be perceived. This technique encourages us with the precision of 97-98%.**

Catchphrases—Level Set Segmentation, Multilayer Perceptron, Kidney stone, and Back Propagation, Wavelet Method, and Sonographic pictures, Energy levels, Fuzzy c-implies Clustering.

I. INTRODUCTION

Kidney stone aggravation is one of the bet for the life in everywhere throughout the world, and most individual with stone framing in kidney early don't concern it as illness and it hurts the appendage (organ) relentlessly. Before watch disorders, most personover-expand by incessant kidney breakdown because of diabetes mellitus and hypertension, glomerulonephritis and so forth. Since kidney failing can be perilous, determination of irritation in the untimely stages is basic. The as of now accessible alternatives incorporate sonographic picture which is one of the non-intrusive minimal effort, broadly utilized imaging methods for diagnosing kidney sicknesses [l]. Stun wave lithotripsy (SWL), percutaneous nephrolithotomy (PCNL), relative super immersion (RSS) are the procedures to test pee. The Robertson Risk Factor Algorithms (RRF An) are open and are utilized for laparoscopic medical procedure,

these calculations are held for remarkable [15].Special cases. Hyaluronan has a focal job in various procedures that can eventually prompt renal stone sickness, including pee focus, Uric corrosive, Salt structure precious stone, crystallization hindrance, gem maintenance, Magnesium ammonium phosphate and amino acid [16].

Mohammad Shorif Uddin ,Tanzila Rahman, proposed decrease of dot commotion and division from sonographic pictures are talked about. This aides in distinguishing the locale of intrigue which likewise upgrade picture quality[1].The kidney US pictures were isolated into four divergent classifications: ordinary, bacterial contamination, cystic sickness, kidney stones, in light of dark dimension co-event lattice (GLCM),which was proposed by wan Mahani Hafizah. From these classes specialists distinguish that the kidney is typical or unusual [2]. Gladis Pushpa had proposed Hierarchical Self Organizing Map (HSOM) for cerebrum tumors utilizing division, wavelets bundles, and the outcomes were right up to most extreme 97% [3]. Norihiro Koizumi proposed high power centered ultrasound (HIFU) procedure, utilized for pulverizing tumors and stones [4, 13]. The numerous highlights for anomalies ID and fake neural system (ANN) for grouping and the outcomes says that the most extreme effectiveness is about 90.47%, and precision 86.66% which was proposed by bommanna raja [5].The Non-obtrusive blend of renal utilizing beat cavitation US treatment proposed stun wave lithotripsy (ESWL) has turned into a standard for the treatment of calculi situated in the kidney and ureter [10]. Mohammad E. Abou EI-Ghar anticipated area of urinary stones with unenhanced processed tomography (CT) utilizing half-radiation (low) portion contrasted and the standard portion and of the 50 patients, 35 patients had a solitary stone while the remainder of them had numerous stones[II]. So as to fathom the nearby minima and division issue the thord Andersson, Gunnar Lathen proposed changed slope inquiry and level set division [12].The

Emmanouil Skounakis proposed layouts based method with precision of 97.2% and mutation in kidneys at an exactness of 96.1% [13].

The paper proceeds as follows: In section II Issue declaration defined, section III methodology, in section IV image segmentation to locate the kidney stone, in section V calculation of energy boosting for segmentation, in section VI wavelets based energy extraction, in section VII artificial neural networks classifiers used is described, in section VIII experiments results are discussion and in the last section we conclude the paper with future work.

II. PROBLEM STATEMENT

The flawed working of kidney movement could be unsafe forever, so finding the kidney stone in the earlier stage is fundamental. So as to convey out careful movement to clean up kidney stone it is important to recognize the kidney stone. The ultrasonic pictures of kidney include mottle crash and are of discouraged inconsistency which empowers the spotting of kidney anomaly a requesting obligation. As a response the specialist may have issue to break down the unassuming kidney stones and their assortment effectively. To subject this issue a changed dimension set segment to distinguish spot of the stone, Wavelets remain by band to excert the stamina zone of the stone and MLP-BP ANN calculations for requesting is normal and considered [9].

III METHODOLOGY

Fig.1. indicates all out strategies utilized for kidney stone location. It comprises of the accompanying squares Kidney picture Directory, Pre-handling, Segment, Wavelet planning and ANN Classification.

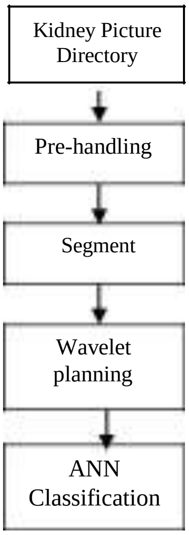


Fig:1: Block Diagram of proposed method

A. Kidney Picture Directory

The 500 kidney picture of MRI of all together for example standard and sporadic/surprising kidney are made from different wellbeing administration focuses of different sufferers and are spared in registry. Interesting picture is gotten to from the catalog and unleash to stone recognizable proof.

B. Image Pre-handling

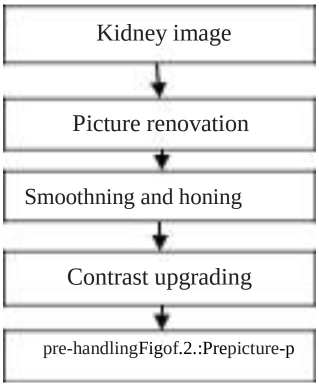
The 500 kidney picture of MRI of all together for example standard and unpredictable/strange kidney are created from different wellbeing administration focuses of different sufferers and are spared in catalog. Extraordinary picture is gotten to from the index and unleash to stone ID

Fig.2. indicated pre-preapring of sonographic picture which comprises of the accompanying advances:

1. Picture Renovation

2.Smoothing and honing

3.Contrast upgrade Kidney picture.



1.Picture Renovation

The real capacity of picture remodel is to reduce the weakening such an incite over the span of acquisition of kidney picture examination. At present here structure being proper area, level set action do pre-possessed. Away influencing activity as concerns game plan circular segment signal, shape smoothens, reduce withstand eventually disappears [1].

2. Smoothing and Honing

Straightforwardly towards recover most noteworthy settlement now by and large dimensional alongside consistency concern, Gabor channel stay reused that progression during the time spent restricting pass channel since the constrained adjoining redundancy circulation[4]. The run of the mill disparity ascribed to the Gaussian conduct would do diverse to familiarize force of fix

3. Contrast Upgrading

To update difference and to increase steady concentraion circle diagram comparision is thought about. The present increase can suffer reused on full picture or lump of an image. Here specific association, augment the contrariety of pictures is done by modifying the models in a genuineness picture, with the goal that the histogram of the resultant picture for the most part contend a specific disperse chart. The yeild earth shattering is of comparative data class as the given flag.

IV. PICTURE SEGMENTATION

Figure.3. displays the level set breaking approach needed to bisect the place of kidney stone. recommended task contains of two altered declivity declivity form. Initially is using a momentum term and second is based on resilient propagation (Rprop) term. intention of the segmentation is to overcome difficulties involved in energy function. The energy function depends on properties of the pictures such as gradients, curvatures, energy levels and regularization terms, e.g. flatening constraints. These are simple, but effective modifications of the basic method are directly compatible with any type of level set implementation. The first proposed method is based on a modification which essentially adds a momentum to themotion in so lution space [15, 19].

This arouse the physical properties of propulsion and often allows the search to neglect local vertex and huge footsteps in helpful way. In order to avoid the complication of inclination descent search, R$_{prop}$ provides a modification which uses ones adaptive step sizes and the signs of the inclination components.

1.Propulsion term

Rotating to inclination descent with propulsion will adopt a search vector according to:

$$a_l = - \eta\ (1-m)\ \nabla f_l + m a_{l-1}$$ -- (1)

Where η is the learning rate and m ∈ [0, 1] is the propulsion. Note that m = 0 gives standard inclination descent $a_l = - \eta\nabla f_l$, while m = 1 gives "infinite momentum" $a_l = a_{l-1}$

2. R$_{prop}$ term

The hindrances of standard tendency descent(SID) is overwhelmed by joining versatile advance sizes ∇_l considered new qualities in which each measurement will have one new esteem for example diminish (∇_l)=dim(x_l). The tendency size is never utilized in Rprop. The new principle considers just the indications of the halfway subordinates. Another favorable position of Rprop, which is critical in functional use, is the bulkness of its parameters; Rprop will work out of the crate in numerous applications utilizing just the standard estimations of its parameters [18, 20] .

We will presently portray the Rprop calculation quickly, however for execution subtleties of Rprop we allude to [23. 21]. For Rprop, we pick a pursuit vector sl as indicated by:

$$s_l = -sign\ (\nabla f)_l * \nabla_l$$ -- *(2)*

Where ∇_l is a vector containing the current new-qualities and sign (.) the component astute sign capacity.

V. ENERGY BOOSTING FOR SEGMENTATION

The division difficulties can be drawn nearer by utilizing the analytics of varieties where vitality capacities is characterized speaking to the target of the trouble. The extraordinary to the utilitarian are discovered utilizing the Euler-Lagrange condition [10, 22] which is utilized to infer conditions of movement, and the relating vitality tendency, for the shape [17]. Utilizing these tendency, a tendency plunge look in shape space is performed to discover an answer for the division issues. Consider, for instance, the inference of the weighted area depicted by the accompanying utilitarian:

$$f(p) = \int_\Omega\int_p g(x,y)\ dxdy$$ -- (3)

Where p is a 1D bend installed in a 2D space, **Ω p is the district within p, and g(x, y) is a scalar capacity. This utilitarian is utilized to expand some amount given by g(x, y) inside p. In the event that g(x, y) = 1 for instance, the zone will be augmented Computing the primary variety of Eq. 1 yields the**

$$\partial p/\partial t = -g(x,y)\eta$$ --- (4)

Where η is the bend typical. Utilizing g(x, y) = 1 which is consistent stream the negative typical way. The bend is frequently verifiably spoken to by the zero dimension of a period subordinate marked separation work, known as the dimension set capacity. The dimension set strategy was presented by Osher and Sethian [6]. Formally, a bend p is portrayed by:

$$p = \{x: \varphi(x, t) = 0\}.$$

The bend p is advanced in time utilizing a lot of fractional differential conditions (PDEs). A movement condition for a parameterized bend

:$\partial p/\partial t = \gamma\eta$ is in general translated into at the level set equation $\partial\varphi/\partial t = \gamma|\nabla\varphi|$

Eq. 2 gives the commonplace at level set condition:

$$\partial p/\partial t = -g(x,y)\ |\nabla\varphi|$$--- --(5)

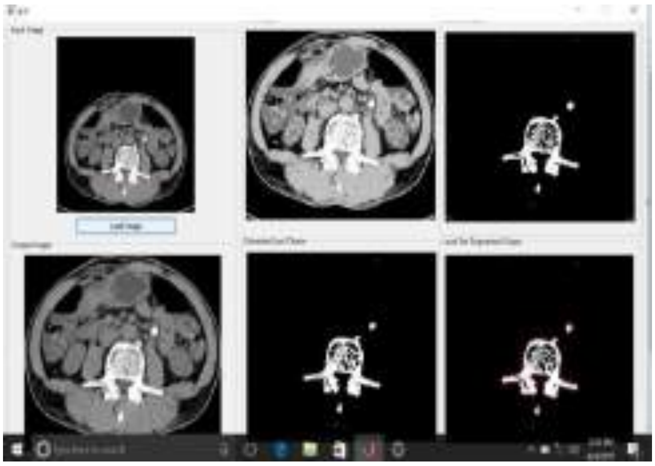


Fig:3.Level set segmentation of kidney stone detection

VI. WAVELET PROCESSING

The sectioned pictures which has from past square is connected to wavelet preparing square. It incorporates of Daubechies channel (DbI2), Symlets channel (symI2) and Biorthogonal channel (bi03.7, bi03.9 and bi04.3). Daubechiesjilter (Db12) in this the number 12 alludes to the quantity of evaporating minutes. Essentially, the higher the quantity of evaporating minutes, the smoother the wavelet (and longer the wavelet channel) and the length of the wavelet (and scaling) channel is multiple times that number [3]. Symletsfilter (symI2) extricate highlights of kidney picture and dissect discontinuities and sudden changes contained in signs, one of the twelfth - request Symlets wavelets is utilized. Biorthogonalfilter (bi03. 7, bi03.9 and bi04.4) channel's wavelet vitality marks were considered and midpoints of even and vertical coefficients subtleties were determined. Each channel will give distinctive vitality levels or vitality highlights. These vitality highlights will demonstrate noteworthy distinction, if there is any stone is available in the specific district or area. The ID of sort of stone is depicted in next area.

VII. ANN CLASSIFICATION

In ANN Classification two designs are utilized to be specific, Multilayer Perceptron and back spread which are portrayed in detail in the accompanying areas.

1. *Multilayer Perceptron (MLP)*

A multilayer perceptron is a feed forward counterfeit neural system calculation that maps sets of vitality esteems acquired from wavelets subbands vitality extraction appeared in the tablel. These vitality esteems are nourished to include layer and increased with introductory loads as in condition (6). The back spread is altered form of straight perceptron in which it utilizes at least three concealed layers with nonlinear initiation work. The back engendering is the most broadly connected learning calculation for multilayer perceptron in neural systems and it utilizes angle drop to limit the squared mistake between the system yield esteem and wanted yield an incentive as in condition (7). These mistake signals are utilized to compute the weight refreshes which speak to intensity of information learnt in the system [7]. Multilayer Perceptron with Back Propagation (MLP-BP) are the fundamental calculations. In light of the writing review, MLP-BP calculation was observed to be superior to anything the others as far as precision, speed and execution [14]. The stages associated with ANN are forward stage and in reverse stage as appeared in fig4. In back engendering, loads are refreshed after each example and by taking one example m at once as pursues:

References

[I] T.anzila Rahm an, Moham m ad Shorif Uddin, "Speckle N oise Redu ction and Segm entation of Kidn ey Regions from Ultrasoun d Im age", 978-1-4799-0400-6113, 2013 IEEE.

[2] Wan Mahani Hafizah, "Feature Extraction of Kidney Ultrasoun d Im ages based on Intensity Histogram and G ray Level Co-occurrence Matrix" 2012 sixth Asia Modeling Sym posium , 978-0-7695-4730-5112, 2012 IEEE.

[3] V. P. Gladis Pushpa Rathi, "Detection and Characterization of Brain Tumor Using Segm entation based on HSOM, Wavelet packet feature spaces and ANN", 978-1-4244-8679- 3111, 2011 IEEE.

[4] Norihiro Koizumi, "Robust Kidney Stone Tracking for a Non -invasive Ultrasound Theragnostic System -Servoing Performance and Safety Enhancem ent", 2011 IEEE International Conference on Robotics and Autom ation Shanghai International Conference Center May 9-13, 2011, Shanghai, China.

[5] K. Bommanna Raja, "Analysis of ultrasound kidney Images using content description multiple fearures for disorder identification and ANN based classification", Proceedings of the international conference on com puting: Theory and applications (ICCTA'07) 0-7695-2770-1/07, 2007.

[6] S. Osher and J. A. Sethian, "Fronts propagating with curvature depen dent speed: Algorithm s based on Hamilton - Jacobi form ulations," J. Comput. Phys., vol. 79, no. I, pp. 12-49, Nov. 1988.

[7] M. Stevenson, R. Weinter, and B. Widow, "Sensitivity of Feedforward Neural Networks to Weight Errors," IEEE Transactions on Neural Networks, Vol. l, No. 2, pp 71-80, 1990.

[8] N.Dheepa "Autom atic seizure detection using higher order m om ents & AN" IEEE- international conference on advan ce in Engineering scien ce and m anagem ent (ICAESM-2012) m arch 30,31,2012 with ISBN: 978-81-909042-2-3, 2012 IEEE.

[9] Joge Martinez- carballido, "Metam yelocyte nucleus classification uses a set of m orphologic tem plates", 2010 electronics, Robotics an d Autom atic Mechanics conference 978 -0-7695-4204-1110, 2010 IEE.

[10] P. M. Morse and H. Feshbach, "The variational integral and the Euler equations," in Proc. Meth. Theor. Phys., I, May 1953, pp. 276-280.

[II] D em etrius H. Bagly, Kelly A. Healy, "Ureteroscopic treatm ent of larger renal calculi (>2cm)", Arab Journal of Urology (2012) 10, 296-300 produ ction and hostin g by Elsevier.

[12] William G. Rob ertson, "Methods for diagnosin g the risk factors of stone form ation", Arab Journal of Urology (2012) 10, 296-300 produ ction and hosting by Elsevier.

[13] Moham ed E. Abou EI-G har, "Low-dose unenhanced com puted tom ography for diagnosin g stone disease in ob ese patients". 2090-598X, 2012 Arab Association of Urolog, Produ ction and hosting by Elsevier B. V, 10,279-283.

[14] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in Froc. IEEE Int. Con! Neural Netw., vol. 1. Jun. 1993, pp. 586-591.

[15] William G Robertson, "Methods for diagnosing the risk factors of stone fonnation", 2090-598X, 2012 Arab Association of Urolog, Produ ction and hosting by Elsevier B.V, 10,250

A Review on Big data in Healthcare

Jayalakshmi D S[1], Arpitha G N[2]
Dept. of CSE, Ramaiah Institute *of Technology,* Bangalore, India

Abstract

The globe is generating a high volume of data in all domains, such as industries, stock markets, social media and healthcare systems. Most of data volume has been generated in the past years. This massive amount of data can bring benefits and draw knowledge to individuals, governments and industries and assist in decision making. In healthcare, an enormous volume of data is generated from healthcare providers and stored in digital systems. Hence, data are more accessible for reference and future use. The ultimate vision for working with health big data is to support the process of improving the quality of service in healthcare providers, reducing medical mistakes and providing a promoting consultation in addition to providing answers when needed. This paper gives systematic review of importance of big data analytics in healthcare system and also we have discussed various applications and analysis of different techniques of big data in healthcare.

Keywords: Big Data, Healthcare, EHR, Machine learning.

1. INTRODUCTION

The big data analytics application in healthcare has a lot of positive and also life-saving results. Big data refers to the tremendous amounts of data made by the digitization of everything that gets united and broke down by explicit innovations. Applied to healthcare, it will utilizeparticular health information of a population (or of a specific individual) and possibly help to avoid epidemics, cure disease, cut down costs, and so on.

Now that we live longer, treatment models have changed and many of these changes are namely driven by data. Doctors want to understand as much as they can about a patient and as early in their life as possible, to pick up warning signs of serious illness as they arise – treating any disease at an early stage is far simpler and less expensive. With healthcare data analytics, prevention is better than cure and managing to draw a comprehensive picture of a patient will let insurances provide a tailored package. This is the industry's attempt to tackle the siloes problems a patient's data has: everywhere are collected bits and bites of it and archived in hospitals, clinics, surgeries, etc., with the impossibility to communicate properly.

To be sure, for quite a long time gathering colossal measures of information for therapeutic use has been expensive and tedious. With the present continually improving innovations, it winds up simpler not exclusively to gather such information yet in addition to change over it into significant basic bits of knowledge, which would then be able to be utilized to give better mind. This is the reason for human services information investigation: utilizing information driven discoveries to anticipate and take care of an issue before it is past the point of no return, yet additionally survey techniques and medicines quicker, monitor stock, include patients more in their very own wellbeing and enable them with the devices to do as such.

1.2 Challenges Big Data Healthcare

One of the greatest difficulties obstructing to utilize enormous information in drug is the means by which curative information is spread crosswise over numerous sources administered by several states, emergency clinics, and managerial offices. Coordination of these information sources would need building up another foundation where all information suppliers work together with one another.

Likewise imperative is executing new web based revealing programming and business insight methodology. Human services needs to make up for lost time with diverse ventures that have officially moved from standard relapse based methods to progressively future-situated like prescient examination, AI, and diagram investigation.

Be that as it may, there are some radiant examples where it doesn't fall behind, for instance, EHRs (especially in the US.) So, paying little respect to whether these organizations are not some tea, you are a potential patient, consequently you should consider new medicinal services examination applications. Furthermore, it's great to investigate some of the time and observe how diverse ventures adapt to it. They can rouse you to adjust and embrace some smart thoughts.

1.3 Big Data Applications in Healthcare

➢ **Patients Predictions for an Improved Staffing** For our first case of huge information in human services, we will take a gander at one great issue that any move administrator faces: what number of individuals do I put on staff at some arbitrary time period? If you put on an over the top number of masters, you hazard having trivial work costs incorporate. Too couple of authorities, you can have poor customer organization results which can be savage for patients in that industry.

Enormous information is dealing with this issue, at any rate at several facilities in Paris. A Forbes article nuances how four centers which are a bit of the Assistance Publique-Hôpitaux de Paris have been using data from a combination of sources to consider step by step and hourly estimates of what number of patients are required to be at each emergency clinic.

One of the key informational indexes is 10 years of medical clinic affirmations records, which information researchers crunched utilizing "time arrangement examination" strategies. These examinations empowered the investigators to see huge precedents in affirmation rates. By then, they could use AI to find the most exact figuring that foreseen future affirmations design.

Summing up the outcome of this work, Forbes states: "the outcome is an internet browser-based interface intended to be utilized by specialists, medical caretakers and clinic organization staff – untrained in information science – to conjecture visit and confirmation rates for the following 15 days. Additional staff can be drafted in when high quantities of guests are normal, prompting decreased hanging tight occasions for patients and better nature of consideration."

- ➢ **Electronic Health Records (EHRs)**

It's the most no matter how you look at it usage of immense data in the solution. Every patient has his own one of a kind mechanized record which fuses economics, remedial history, sensitivities; explore office test outcomes, etc. Records are shared by methods for secure information structures and are accessible for providers from both open and private section. Each record is incorporated one changeable report, which suggests that pros can realize changes after some time with no work area work and no hazard of information replication.

EHRs can likewise trigger admonitions and updates when a patient ought to get another lab test or track solutions to check whether a patient has been following specialist's requests. In spite of the fact that EHR are an incredible thought, numerous nations still battle to completely execute them. U.S. has made a noteworthy jump with 94% of medical clinics embracing EHRs as per this HITECH investigate, yet the EU still falls behind. Be that as it may, an eager mandate drafted by European Commission should transform it: by 2020 incorporated European wellbeing record framework ought to turn into a reality.

- ➢ **Real-Time Alerting**

Various instances of huge information examination in medicinal facilities share one critical usefulness – continuous alarming. In emergency clinics, Clinical Decision Support (CDS) programming investigations therapeutic data on the spot, furnishing wellbeing specialists with counsel as they settle on prescriptive choices.

Another model is that of Asthma polis, which has begun to utilize inhalers with GPS-empowered trackers so as to recognize asthma patterns both on an individual dimension and taking a gander at bigger populaces.

- ➢ **Enhancing Patient Engagement**

Numerous buyers and henceforth, potential patients as of now have an enthusiasm for brilliant gadgets that record each progression they take, their pulses, dozing propensities, and so forth, consistently. This crucial data can be combined with other tractable information to recognize potential wellbeing dangers hiding. Incessant sleep deprivation and a raised pulse can flag a hazard for future coronary illness for example. Patients are legitimately associated with the observing of their own wellbeing, and motivating forces from medical coverage can push them to lead a solid way of life (e.g.: giving cash back to individuals utilizing shrewd watches).

- ➢ **Big Data Might Just Cure Cancer**

Another captivating instance of the usage of gigantic data in social protection is the cancer Moonshot program. Earlier quite far of his second term, President Obama thought of this program had the target of achieving 10 years of headway towards reestablishing illness down the center that time.

Remedial researchers can use a great deal of data on treatment plans and recovery rates of dangerous development patients in order to find examples and meds that have the most surprising rates of accomplishment as a general rule. This data can in like manner brief abrupt points of interest, for instance, finding that Desipramine, which is a stimulant, can help fix specific sorts of lung threatening development.

Regardless, to make these sorts of encounters progressively available, calm databases from different associations, for instance, medicinal facilities, universities, and philanthropies ought to be associated up.

- ➢ **Predictive Analytics in Healthcare**

We have authoritatively seen perceptive examination as one of the best business knowledge designs two years in a row, yet the potential applications achieve far past business and much further later on. Optum Labs, the US investigate network, has accumulated EHRs of in excess of 30 million patients to make a database for judicious examination gadgets that will improve the movement of thought.

The goal of therapeutic administrations business learning is to enable experts to settle on data driven decisions inside seconds and improve patients' treatment. This is particularly important if there ought to emerge an event of patients with complex restorative narratives, encountering different conditions. New gadgets would in like manner have the ability to envision, for example, which is in peril of diabetes, and thusly be instructed to make use in regards to additional screenings or weight the administrators.

- **Reduce Fraud and Enhance Security**

A few investigations have demonstrated that this specific industry is 200% bound to encounter information breaks than some other industry. The reason is basic: individual information is amazingly significant and productive on the illicit businesses. What's more, any break would have sensational results. Obviously, huge information has innate security issues and many feel that utilizing it will make the associations more powerless than they as of now are. Be that as it may, progresses in security, for example, encryption innovation, firewalls, hostile to infection programming, and so forth, answer that requirement for greater security, and the advantages brought generally surpass the dangers.

Moreover, it can help forestall misrepresentation and off base cases in a fundamental, repeatable manner. Investigation help streamlines the handling of protection claims, empowering patients to show signs of improvement returns on their cases and parental figures are paid quicker. For example, the Centers for Medicare and Medicaid Services said they spared over $210.7 million in fakes in only a year.

- **Telemedicine**

Telemedicine has been accessible for over 40 years, anyway only today, with the passage of online video gatherings, PDAs, remote devices, and wearable, has it had the ability to come into a full grow. The term suggests the transport of remote clinical organizations using development.

It is used for basic gatherings and beginning the investigation, remote patient watching, and restorative preparing for prosperity specialists. Some continuously express usage join telesurgery – experts can perform undertakings with the use of robots and quick steady data transport without physically being in a comparable region with a patient. Clinicians use telemedicine to give redid treatment plans and deflect hospitalization or re-attestation.

- **A Way to Prevent Unnecessary ER Visits**

Sparing time, cash and vitality utilizing enormous information investigation for human services are fundamental. Imagine a scenario where we disclosed to you that through the span of 3 years, one lady visited the ER in excess of multiple times. That circumstance is a reality in Oakland, California, where a lady who experiences psychological instability and substance misuse went to an assortment of neighborhood emergency clinics on a practically regular schedule.

> "Everyone had good intentions. Yet, she was being alluded to three distinctive substance misuse facilities and two diverse psychological wellness centers, and she had two cases the executive's specialists both dealing with lodging. It was not just awful for the patient; it was additionally a misuse of valuable assets for the two clinics."

2. Analysis of different machine learning approaches

In industry, how the market is developing can be anticipated utilizing information examination. Expectation is finished utilizing AI calculations. Distinctive elements are researched in making expectation. There ought to be earlier information about the class about which the forecast model is getting down to business. Two kinds of learning approaches utilized in human services are directed and unsupervised learning. To manage substantial measure of highlights, we apply dimensionality decrease ways to deal with acquire important highlights. Dimensionality decrease dispenses with the pointless highlights to accelerate calculation and forecast for precise choice [1]. In the accompanying segment of paper, distinctive information mining approaches are outlined.

A. Feature Selection and Evaluation

At first, information pre-handling is performed to diminish the commotion and repetitive information to speedup calculation. As the dataset is isolated into preparing and testing subsets, the preparation subset is utilized for the component extraction and determination. As such, subset from given highlights are chosen. These highlights are gotten from pixel force, hues and geometric highlights, for example, forms, edges and shapes [2]. These highlights are additionally utilized for end of excess and loud highlights. This progression will help for model elucidation. Distinctive methodologies can be utilized to acquire ideal element subset.

- Complete search

  This pursuit ensures the best arrangement. It very well may be connected for finding ideal arrangement of enormous information issues. Heuristic methodologies like branch and bound limits the looking entire component space

- Sequential search

In this kind of hunt, heuristic methodology is connected [3]. This methodology seeks either from entire list of capabilities or invalid component subset to get ideal arrangement by including or evacuating highlights, separately. Included highlights can't be expelled and evacuated highlights can't be included. This methodology does not ensure for ideal arrangement. In any case, arrangement of this methodology winds up worthy because of less handling time.

- Random Search

Random features are chosen to begin this pursuit. Nearby ideal arrangement can be accomplished utilizing irregularity. Execution of arbitrary inquiry can be expanded by incorporating with consecutive look for age of irregular subset like re-enacted strengthening and irregular begin slope climbing calculation.

Feature evaluation step includes values that are appointed to highlight subset relying upon explicit criteria. Similitude is resolved utilizing class marks for arrangement. Finding immaterial highlights is a difficult errand in grouping. The idea driving component assessment claims: "Good features subsets containfeatures highly correlated with class, yet uncorrelated witheach other" [4]. Commonly used for feature evaluation arewrapper, filter, and hybrid approaches.

Wrapper approach has capacity to give ideal arrangement which can be tuned for classifier learning. Calculation performs seeking to choose specific element subset dependent on foundation work (Fig. 2) [12]. Computational cost increments since calculation is kept running at every emphasis. This strategy includes high computational expense and not appropriate for taking care of huge information issues.

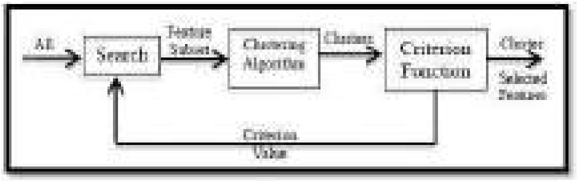


Fig. 2: Wrapper approach

Filter method can be utilized to decide significance between chose highlights (Fig. 3). Subset creation isn't utilized for classifier learning. Thusly, conventional outcome is created rather than explicit calculation tuning [5]. Its reasonableness turns out to be high for taking care of huge information issues. This methodology performs hitter than different methodologies like Relief traits estimator [6] [6].

Fig. 3: Filter approach

Hybrid methods achieve assessment on highlights and make a component subset by picking the best among them in further emphases. Examination between various sorts of highlights subsets is performed for ideal learning of calculation [7]. Mixture approaches have acquired significance as contrasted and different methodologies. It includes exchange off among time and effectiveness. From human services angle, it is more appropriate methodology than different methodologies. There is extent of problematic arrangement [8].

B. Classification

Grouping model orders input information and target class is utilized by classifier for preparing and testing reason [16]. Info is given to AI calculation; target class information is additionally given for performing right choice by classifier that include highlights of info information dependent on classifier demonstrate for target class. In the wake of preparing of classifier, next stage is trying in which input information is given to perform expectation about target class. Because of increment in volume of enormous information, existing grouping approaches have a few impediments and include high preparing expense [1]. Ordinarily utilized arrangement systems in medicinal services space are choice tree, bolster vector machine, neural system, k-closest neighbor, and Bayesian methodology [9].

A choice tree is a tree like structure utilized for characterization [10]. Choice tree is connected for increasing precise and quick outcomes because of its basic structure. It's hard to develop choice tree with gigantic measure of information in light of the fact that a great deal of time is required to build it. Choice tree is most regularly order approach in medicinal services area for critical thinking by doling out class name to tolerant. Fig. 4 speaks to an example choice tree.

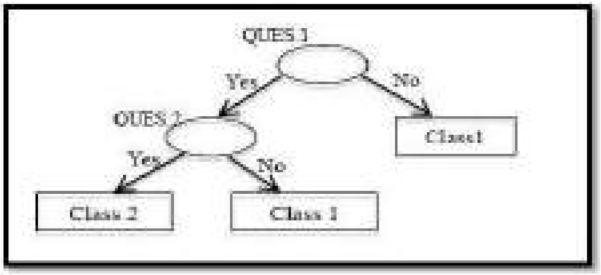


Fig. 4: Classification by decision tree

Support Vector Machine (SVM) is a factual model utilized for order. SVM is proficient for settling on choices on expansive informational index. SVM is exceptionally useful uniquely multi space applications in enormous information condition. SVM is scientifically intricate and computational costly [11]. Staggered or binomial characterization can be performed utilizing SVM [12].

SVM is most prevalent methodology among existing AI systems. The execution of SVM debases on bigger informational indexes that comprise of noisier information. This strategy performs expectation exceptionally quick subsequent to preparing [13]. Expectation is done dependent on hyper plane and bolster vector that performs division in higher dimensional space. To beat issue of loud information, SVM is joined with other AI procedures for getting better outcomes [14]. Fig. 5 indicates SVM characterization is spoken to by hyper plane for basic leadership. In customary AI methods, Neural Network (NN) portrayed part of variety. In NN, loads of associating joins are balanced between neurons until achieving ideal esteem. NN is most generally utilized for critical thinking is Multilayer Perceptron (MLP) [15]. It works like human mind. To begin with, NN is prepared to perform characterization. After finish of preparing, testing is performed on info information.

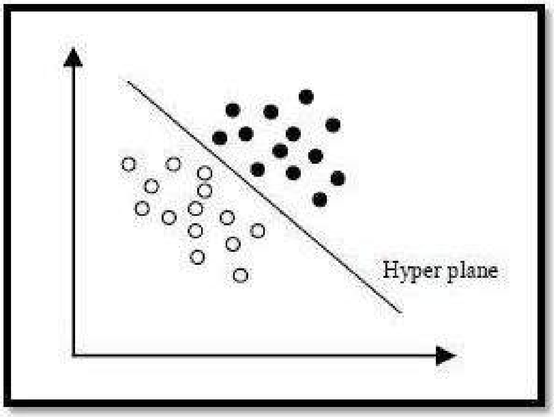


Fig 5: Classification by support vector machine

Significant downside of utilizing NN is calculation time for extensive informational index. Memory necessity likewise increments as size of informational index increments. NN is appropriate for increasing ideal outcomes however it requires more opportunity for bigger informational index. There is have to receive half and half way to deal with limit computational time by joining NN with different ways to deal with beat issue of computational time for bigger informational index [16].

Convolutional Neural Network (CNN) is the development type of NN. CNN is a multilayer neural system that takes contribution to vector structure. Anyway if there should arise an occurrence of medicinal pictures pixels or voxels are data source. In standard multi-layer neural system, Convolutional layers interlard with sub-inspecting layer pursued by completely associated layer is intended to hitter utilization of spatial data by accepting 2D or 3D pictures as information [17].

K-Nearest Neighbour (KNN) is a basic grouping model that works as per closest neighbour of existing class name [18]. In KNN, estimation of k is registered to discover closest neighbour where k speaks to number of closest neighbours.

To acquire exact outcomes, advancement calculations are connected to limit computational expense. KNN furnishes better outcomes when contrasted and Bayesian strategy in different applications. Bayesian strategy works as indicated by Bayes hypothesis. Bayes hypothesis gives scientifically grounded instruments to discover the vulnerability of a model. For bigger informational collection, Bayesian classifier can perform better in characterization. Guileless Bayesian model give high exactness just when characteristics esteems are autonomous. It is a factual model and gives high precision. This methodology expect all ascribes are autonomous as indicated by one another. This classifier can perform well in social insurance either by pre-preparing or without pre-handling.

C. Clustering

Clustering groups are comparable information together to make groups. Target class name isn't given at first in bunching. There exists higher similitude inside same bunch. Distinctive bunches have lower closeness between information focuses. Customary methodologies that are being utilized for comparability measures are Jaccard measure, Pearson connection, Euclidean separation and Cosine [19]. There is no requirement for past data about information is required to work after grouping. It is reasonable for applying on microarray information in which almost no data is required about qualities. Tapia et al. executed hereditary calculation on articulation information to break down it [20]. It has capacity to speak to data in minimal structure without losing much data by creating ideal groups with reference to enormous information [21]. There exists an alternate grouping calculation. Divided grouping includes predefined number of bunches.

In this method, informational index is arranged into predefined number of parcels. There is none of void segment and information has a place with precisely one group. This methodology can be additionally arranged dependent on group centroid and comparability measures, i.e., K-Medoids and K-Means. In K-Medoids, medoids are utilized rather than centroid. Focus purpose of a bunch is medoid that exist in informational index. Belciug et al. connected bunching for identification of bosom malignant growth to get better exactness [22].

In various levelled bunching, it isn't important to pre-characterize number of groups [21]. There are two principle classes of various levelled bunching with deference of working. Agglomerative system is a base up strategy in which each datum point is expected as bunch while two distinct groups are converged based on couple of closeness measures [23]. Wanted bunch can be acquired after some cycle. Significant disadvantage of this methodology is combination can't be rollback. Troublesome method is top-down procedure in which all information focuses are treated as a bunch. After some cycle, group is additionally arranged into two classes dependent on certain measures. Required number of bunches can be accomplished by over and again running this calculation. Customary methodologies that are being utilized for comparability measures are Jaccard measure, Pearson connection, Euclidean separation and Cosine [19]. There is no requirement for past data about information is required to work after grouping. It is reasonable for applying on microarray information in which almost no data is required about qualities. Tapia et al. executed hereditary calculation on articulation information to break down it [20]. It has capacity to speak to data in minimal structure without losing much data by creating ideal groups with reference to enormous information [21]. There exists an alternate grouping calculation. Divided grouping includes predefined number of bunches.

This methodology likewise has real downside that once bunch is separate into sub-groups, these can't be gathered to make unique group. In various levelled bunching calculations, emphases can be stop by increasing wanted number of groups.

3. Results

All existing traditional data mining techniques have notcapability to perform better classification for big data in healthcaredomain. Therefore, there is need to merge the distincttechniques together to perform better classification. Onepopular approach for requiring relationship between data isAssociation. It is necessary to obtain relationship betweendiseases for finding similar treatment. Apriori algorithm isapplied in association to find relationship between items andalso to perform separation between similar as well as differentitems. PSO optimization approach is mostly combined withSVM to perform optimization first on feature set and thenclassifier is applied to separate data. KNN is fused withfuzzy logic to decrease computation time. It is betterto combine different techniques in healthcare domain forobtaining better classification.

4. Conclusion

Due to rapid enhancement in big data prediction and analysis,healthcare domain has got a valuable attention from recentfew years. All machine learning techniques have less capability to perform better classification of big data in healthcare domain. Hence fusion of different techniques gives the better classification in healthcare domain.

References

[1] S. Suthaharan, “Big data classification: Problems and challenges in network intrusion prediction with machine learning,” SIGMETRICS Perform. Eval.Rev., vol. 41, no. 4, pp. 70–73, 2014.

[2] A. S. Panayides, C. S. Pattichis, and M. S. Pattichis, “The promise ofbig data technologies and challenges for image and video analytics inhealthcare,” in Signals, Systems and Computers, 2016 50th AsilomarConference on. IEEE, 2016, pp. 1278–1282.

[3] M. A. Hall, “Correlation-based feature selection for machine learning,”Ph.D. dissertation, The University of Waikato, 1999.

[4] H. Liu and H. Motoda, Computational methods of feature selection.CRC Press, 2007.

[5] Y. Saeys, I. Inza, and P. Larra˜naga, “A review of feature selectiontechniques in bioinformatics,” bioinformatics, vol. 23, no. 19, pp. 2507–2517, 2007.

[6] M. A. Hall, “Correlation-based feature selection of discrete and numericclass machine learning,” 2000.

[7] H. Liu and L. Yu, “Toward integrating feature selection algorithms forclassification and clustering,” IEEE Transactions on knowledge and dataengineering, vol. 17, no. 4, pp. 491–502, 2005.

[8] R. Dharavath and A. K. Singh, “Entity resolution-based jaccard similaritycoefficient for heterogeneous distributed databases,” in Proceedingsof the Second International Conference on Computer and CommunicationTechnologies. Springer, 2016, pp. 497–507.

[9] G. Kesavaraj and S. Sukumaran, “A study on classification techniquesin data mining,” in Fourth International Conference on Computing,Communications and Networking Technologies (ICCCNT). IEEE,2013, pp. 1–7.

[10] D. Tomar and S. Agarwal, “A survey on data mining approaches forhealthcare,” International Journal of Bio-Science and Bio-Technology,vol. 5, no. 5, pp. 241–266, 2013.

[11] D. Wang, X. Liu, and M. Wang, “A dt-svm strategy for stock futuresprediction with big data,” in IEEE 16th International Conference onComputational Science and Engineering (CSE). IEEE, 2013, pp. 1005–1012.

[12] S. Suthaharan, “Support vector machine,” in Machine Learning Modelsand Algorithms for Big Data Classification. Springer, 2016, pp. 207–235.

[13] B. Sch¨olkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “Newsupport vector algorithms,” Neural computation, vol. 12, no. 5, pp.1207–1245, 2000.

[14] T.anzila Rahman, Mohammad Shorif Uddin, "Speckle Noise Reduction and Segmentation of Kidney Regions from Ultrasound Image", 978-1- 4799-0400-6113, 2013 IEEE.

[15]Wan Mahani Hafizah, "Feature Extraction of Kidney Ultrasound Images based on Intensity Histogram and Gray Level Co-occurrence Matrix" 2012 sixth Asia Modeling Symposium , 978-0-7695-4730-5112, 2012 IEEE.

[16] V. P. Gladis Pushpa Rathi, "Detection and Characterization of Brain Tumor Using Segmentation based on HSOM, Wavelet packet feature spaces and ANN", 978-1-4244-8679- 3111, 2011 IEEE.

[17] D. Tomar and S. Agarwal, “A survey on data mining approaches forhealthcare,” International Journal of Bio-Science and Bio-Technology,vol. 5, no. 5, pp. 241–26

Implementation Of Fruit Quality Management System Based On Image Processing

Tuhin pal bhowmik[1],Sagar Gurung[2],Manish sah[3],Natesh Kumar S[4]

Department of Electronics and communication Engineering

East west institute of Technology

Bengaluru,india

***Abstract*— This project aims at presenting the concept of fruit quality management. In recent years automatic vision based technology has become more potential and are more important to many areas including agricultural fields and food industry. The desired system determines the quality of fruits by its color, size. Sorting tons of fruits manually is a time consuming, costly, and an inaccurate process and to developed in order to increase the quality of food products made from fruits. The sorting process depends on capturing the image of the fruit and analyzing this image using image processing techniques to discard defected fruits. Color is most striking feature for identifying disease and maturity of the fruit. The main emphasis is to do the quality check with a short span of time so that maximum number of fruits can be scrutinized for quality in minimum amount of time. The absolute reference point is the way to perceives and interpret the quality of fruit. The present assessment of fruit quality requires new tools for size and color measurement and capturing the fruit side view image, some fruit characters is extracted by using detecting algorithms. This system performs the sorting using MATLAB software and gives some advantages over traditional practices.**

Keywords—Fruit mangement, color, size, insert (key words**)**

I. INTRODUCTION

In India the ever-increasing population, losses in handling and processing and the increased expectation of food products of high quality and safety standards, there is a need for the growth of accurate, fast and objective quality determination of food and agricultural products. Agriculture is one of the largest economic sectors and it plays the major role in economic development of our country. In our country the ever-increasing population, losses involved in processing and the increasing demand of fruits of high quality with good appearance, there is a need for the development of accurate, fast and focused quality determination of food and agricultural products like fruits and vegetables. Whereas grading is done based on the overall quality features of fruits by considering a number of attributes like shape, size, color etc.

Classification is necessary for the quality evaluation of Agricultural produce like fruits and vegetables. Fresh market fruits like Apples, Oranges, and Banana are graded into categories based on several factors such as color, shape, size and presence defects or bruises, blemishes on it. Fruit market is getting highly selective, requiring their suppliers to distribute the fruits of high standards of quality and presentation as well. So there is an increasing need to supply quality fruits within a short period of time has given rise to the development of automated Grading of fruits to improve the quality. As the major source of national income is from agriculture, it becomes the backbone of every countries Economy. India ranks first among the other countries in the world, in the production of milk, pulses, jute and jute-like fibers; second in cereal crops, cotton, vegetables and fruits production; and is one of the leading producers of spices and plantation crops as well as fisheries and poultry.

If the overall production is good then it will directly increase the annual income of the cultivators and ultimately the national income of the country. Therefore currently researchers are trying to develop innovative and automated methods using science and technology to increase the production of agricultural industry.

Generally, the quality of fruit shape, color and size, default and so on cannot be evaluated on line by the traditional methods. The development of image processing technology and computer software and hardware, it becomes more attractive to detect fruits' quality by using vision detecting technology. At present, most existing fruit quality detecting and grading system have the disadvantage of low efficiency, low speed of grading, high cost and complexity. So it is significant to develop high speed and low cost fruit size detecting and grading system.

Mostly, they provide two choices for grading either by color and size. In first case, we are going to sort circular shaped fruits according color and grading is done according to size. The proposed automated classification and grading system is designed to combine three processes such as feature extraction, sorting according to color and grading according to size. Software development is highly important in this color classification system and for finding size of a fruit. The entire system is designed over MATLAB software to inspect the color and size of the fruit.

II. LITERATURE SURVEY

Hongshe Dang, Jinguo Song, Qin Guo [1] have proposed fruit size detecting and grading system based on image processing. The system takes ARM9 as main processor and develops the fruits size detecting program using image processing algorithms on the QT/Embedded platform. Authors in [2] have proposed system which finds size of different fruits and accordingly different fruits can be sorted using fuzzy logic, here author proposed matlab for the features extraction and for making GUI. John B. Njoroge. Kazunori Ninomiya. Naoshi Kondo and Hideki Toita [3] have developed an automated grading system using image processing where the focus is on the fruit's internal and external defects. The system consists of six CCD cameras. Two cameras are mounted on the top,two on the right and another two cameras mounted on the left of the fruit. X-ray imaging is used for inspecting the biological defects. Image processing is used to analyze the fruit's features; size, color, shape and the grade is determined based on the features. The developed system is built from a combination of advanced designs, expert fabrications and automatic mechanical control. J. V. Frances, J. Calpe, E. Soria, M. Martinez, A. Rosado, A.J. Serrano, J. Calleja, M. Diaz [4] presented a procedure to improve the performance, whether increasing speed or accuracy, of the load-cell-based weighting subsystem in a fruit sorting and grading machine to achieve an accuracy of + l gram. Wong Bing Yit, Nur Badariah Ahmad Mustafa, Zaipatimah Ali, Syed Khaleel Ahmed, Zainul Abidin Md Sharrif [5] proposed new MMS-based system design and developed with signal processing for fruit grading for consumers. The prototype network architecture, integration of wireless messaging system with signal processing between mobile consumers for development purposes was studied, proposed and designed. Here in all above work the grading is done by considering each attributes separately. This paper suggest an integrating system which provides different options for grading fruits i.e. either according to color and size or weight and finally by using GSM module system can send message automatically to consumer or to head office to know about the grading process progress.

III. SYSTEM OVERVIEW

This automated system is designed to overcome the problems of manual techniques. Here the hardware model is designed which contains conveyor system, grading assembly which contains two container in either direction of the model which servo motor is connected for moving in clockwise or anticlockwise direction, digital camera, UDM sensor, Arduino uno, LCD display on field, and grading assembly. The block diagram of a system is shown in Figure 1.

The image could be captured using a regular digital camera. The system arrangement is done as shown below the basic aim is to obtaining the fruit's features. The system consists of several steps like feature extraction, sorting and grading. As proposed in [1], to avoid shadow, two annular lights are used to supply well- distributed light. The white background color in image is easier to extract the fruit edge characters later. So the background is set white in whole process of image capture. The light and camera location is as shown in Figure 2.

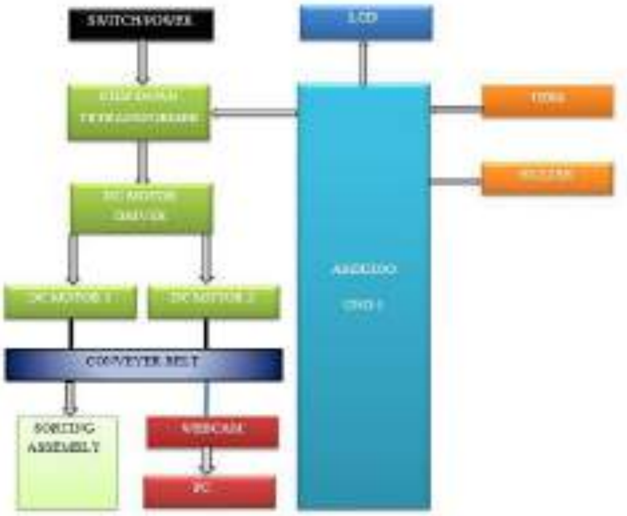


Figure 1: Block diagram

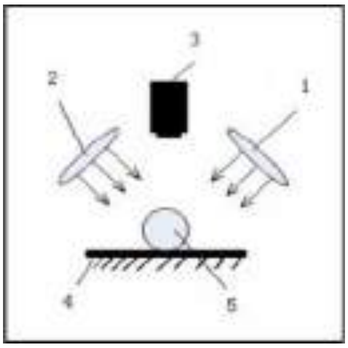


Figure 2: Fruit system light and camera location.1-light; 2-light; 3-Web camera; 4-conveyor belt; 5-fruit

IV. FRUIT SIZE AND COLOR DETECTION

A. Processing flow

Take apple as the processing example, according to [1], the apple size is its diameter, which is the longest distance in the apple's cross section. So the detecting program is focused on how to calculate the diameter in an apple side view image. The fruit image size detecting and grading processing flow is shown in Fig. 3.

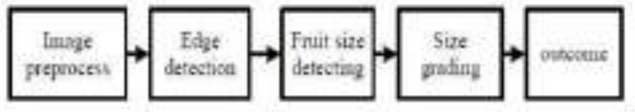


Figure 3: processing flow

B. Colour Detection

In the process of fruit color is detected according to RGB values, here fruits are sorted according to color and size. So for e.g. two fruits are considered say tomato having red color and guava having green color, so in this step work is going to find out color of a fruit by using RGB values of an image taken from the camera, this image can be processed by using Matlab software and accordingly color can be detected i.e. green or red.

Color detection algorithm:

```
 1) Start
 2) Read the input color image using imread fuction.
 3) Read the input pixel of color image in three
    different planes (RGB) and store it into
    three variable r, g, b.
 4) Read the small region of fruit to detect color
    of fruit.
 5) Store in different variable r1, g1, b1.
 6) Calculate the mean of r1, g1, b1 and store
    into variable r2, g2, b2.
 7) Compare the value with threshold.
 8) If g2>threshold, Color detected is green.
 9) If r2>threshold, Color detected is Red.
10) End
```

C. Edge Detection

Once color is detected, there is a need to find out size of a fruit. The size of circular shaped fruit is its diameter [1]. The edge extraction is key factor for size detecting. After gray image, the most powerful edge-detection method that finds edge is the canny method. The Canny method differs from the other edge-detection methods in that it uses two different thresholds (to detect strong and weak edges), and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges.

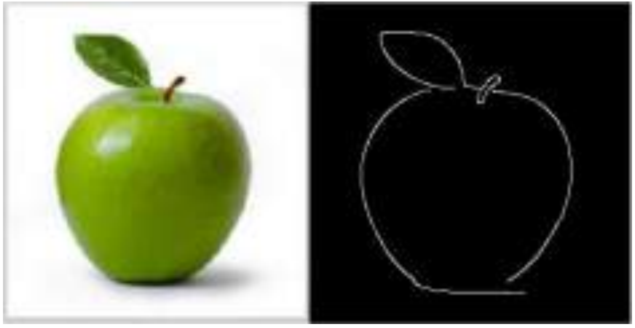


Figure 4: Edge detection

Fruit Size detecting algorithm:

In order to calculate this diameter, the fruit's natural symmetry is considered, so the fruit size detecting algorithm based on its symmetry mainly contains two parts: finding the center coordinate of fruit's shape in image and finding fruit's axis in image. The algorithm is described as follows:

1) Finding the center coordinate of fruit's shape in image: The center coordinate can be easily calculated once finding the edge sequence points. Suppose the finding edge sequence points is q (xi, yi), i=1,…,n. the center coordinate of fruit's shape is (cx , cy), it can be calculated by (1) and (2) as in [1]:

$$c_x = \frac{\sum_{k=1}^{n}[y_k(x_k^2 - x_{k-1}^2) - x_k^2(y_k - y_{k-1})]}{2\sum_{k=1}^{n}[y_k(x_k - x_{k-1}) - (y_k - y_{k-1})]} \qquad (1)$$

$$c_y = \frac{\sum_{k=1}^{n}[y_k^2(x_k - x_{k-1}) - x_k(y_k^2 - y_{k-1}^2)]}{2\sum_{k=1}^{n}[y_k(x_k - x_{k-1}) - x_k(y_k - y_{k-1})]} \qquad (2)$$

2) Finding the fruit's axis in image: After get the center coordinates of fruit's shape in image, the diameter sequence from the edge point to the center can be also acquired, that is p(j) . j=1,…, n. and then it's even points selected from p(j) . called r(j) ,j=1,…,m. suppose h □ 1,…,m / 2 . So the r(j) can be divided in two parts by h, and then calculating the g, which is described by (3).

$$g = \sum_{i=1}^{\frac{m}{2}} |r(h+1) - r(h-1)| \; (h = 1,2, \ldots\ldots, m/2)$$

$$\text{If} |h-l| \leq 1, \; r(h-l) = r(m+l-h-1) \qquad (3)$$

The direction of r (h) is the fruit's axis in image while the g getting its minimum. Following the below method, the fruit's axis point and center point is found in image as shown in Figure. 6. Once known the axis point and the center point, a line through the center point which is vertical to the line from axis point to center point will be crossed with the edge sequences, two edge points that on the line will be searched. Suppose the two points is (x1, y1) and (x2, y2) in order to improving the system's speed, the diameter is calculated by (4) directly, this diameter value can approximately indicates the fruit's real maximal diameter in image. From the detecting result in Figure 8, this method can find the axis point accurately in a fruit image. And also, it still can find the two points while the fruit's location changed. So this method can satisfy the fruit size detecting on line which its location changed often.

$$d = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2} \qquad (4)$$

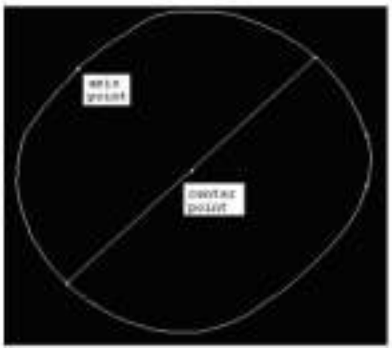


Figure 5: Fruit's axis point and the center point location

E. Fruit Size Grading

According to apple state criterion, size grading is judged by the detected diameter of an apple, [6] the criterion is shown as the table 1.

Criteria	Diameter
Big Fruit	>=60mm
Small Fruit	<=50mm

V. Conclusion

The proposed system is a demo version, so for a large scale production the number of cameras and length of conveyor system can be modified. This work presents the integrated techniques for sorting and grading of different fruits. Generally image capture is a big challenge as there is a chance of high uncertainty due to the external lighting conditions, so we are taking the advantage of gray scale image which are less effected to the external environment changes as well as beneficial for finding size of a fruit. Same way while collecting fruit from conveyor system weight can be the other important quality parameters for the sorting and grading of fruits. So the weight measurement of a fruit can be added to the design for further categorization of fruits. Speed and efficiency of a system can be further improved by using other advance microcontroller for the same purpose.

FUTURE SCOPE

Further design can be modified by increasing size of conveyor belt so that it is possible to perform quality inspection of large fruit than apple, and increase accuracy of the system so that it can differentiate between artificial , hybrid color from original fruit color.

ACKNOWLEDGMENT

We would like to thank my all the staff members of EWIT, Bengaluru for being moral support through the period of my project study, whose help and shared knowledge was the main support to our project.

REFERENCES

[1] Hongshe Dang, Jinguo Song, Qin Guo "A Fruit Size Detecting and Grading System Based on Image Processing", 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics

[2] Harshavardhan G. Naganur, Sanjeev S. Sannakki, Vijay S Rajpurohit3, Arunkumar R, "Fruits Sorting and Grading using Fuzzy Logic", ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 6, August 2012 117

[3] Manoj B. Avhad, Satish M. Turkane, "ARM Based Fruit Grading and Management System Using Image Processing", ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 1, January 2013 243

[4] Ms.Rupali S.Jadhav, PROF. S.S.Patil, "A Fruit Quality Management System Based On Image Processing", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 8, Issue 6 (Nov. - Dec. 2013), PP 01-05 www.iosrjournals.org www.iosrjournals.org

[5] Mr. Sumit S. Telang, Prof. S.M.Shirsath, "Fruit Quality Management using Image Processing", International Conference on Ideas, Impact and Innovation in Mechanical Engineering (ICIIIME 2017) ISSN: 2321-8169 Volume: 5 Issue: 6 727 –733

[6] Manali R. Satpute, Sumati M. Jagdale, "Automatic Fruit Quality Inspection System",

[7] Ebrahim.Aghajari, D.C. Gharpure, "Fuzzy C-Means Clustering Algorithm for Quality Inspection of Fruits based on Image Sensors Data"

[8] Hadha Afrisal, Muhammad Faris, Guntur Utomo P., Lafiona Grezelda, Indah Soesanti, Mochammad Andri F. "Portable Smart Sorting and Grading Machine for Fruits Using Computer Vision", 2013 International Conference on Computer, Control, Informatics and Its Applications

[9] Yogitha.S, Sakthivel.P, "A Distributed Computer Machine Vision System for Automated Inspection and Grading of Fruits"

[10] Chandra Sekhar Nandi, Bipan Tudu, Chiranjib Koley, "Machine Vision Based Automatic Fruit Grading System using Fuzzy Algorithm", 2014 International Conference on Control, Instrumentation, Energy & Communication(CIEC).

A Tool for Improving the Technical Knowledge of Students with Automatic Formative Assessment

Shruthi B Gowda[1], Mohan B C[2], Murali M[3], Pavan M[4], Narayan Muduli[5]

Department of Computer Science and Engineering Vivekananda Institute of Technology, Bengaluru, India

{Shruthi.b.gowda[1], gowdamohan009[2], muralimuniyan25[3], pavan21manju[4], narayankumarmuduli[5]}@gmail.com

Abstract---**Concept of doing this project is to train the Engineering students based on their academic background mainly to support and uplift the knowledge that is required to face the competitive world by providing successful training programs. The success of the program relies on a centralized approach process. To provide software platform for the students to learn the technical concepts. The application examines the overall impact of the program such as: 1) Preparations 2) Interest in student learning. 3) Ability to respond to questions. 4) Overall effectiveness. Technology on the other hand is rising and leading the planet and also the whole world is looking on technology, while not technology is extremely not possible to survive. So, the students has to match these latest trending technologies by gaining the knowledge day-by-day along with solving assessments. Project title mainly concentrates on the students, they are been provided with a software platform which would train the students. Effective use of technologies like Machine Learning is used to build our project. On daily or weekly basis set of suggested technical topics and videos are linked to their login credentials, the students has to work and periodically tasks will be assigned and graded automatically based on the performance of student. The goal of formative assessment is to monitor student learning to provide ongoing feedback that can be used by instructors to improve their teaching and students to improve their learning. At the end of all semesters the overall evaluation will be counted and it will be aggregated, this can be seen by the company officials which may give extra beneficial for the students who attend placement drive. Thus the application trains the students for upcoming generations. Random forest algorithm is a combination of both Naïve Base and Decision Tree algorithms. Regression and classification both the techniques are used in the Random forest algorithm, which will help us to make the evaluation automotive, effective and faster to generate the further process for the candidate.**

Keywords—**Centralized Approach, Machine Learning, Formative Assessment, Random Forest.**

I. INTRODUCTION

A COMMON issue facing in many schools and within universities is the ever increasing requirement for teaching assistants. Main theme in the literature is that the best training takes the form of on-the-job practice with a focus on self-reflection. Student satisfaction has to be grown in importance due to the competitive education environment. Various studies have pointed to the emergence of a new professional profile in organizations. Essentially, an individual with this profile has various types of knowledge, e.g., operational; technological, particularly in computing; Management and administration; social; and, finally, emotional. The emergence of this new type of professional is not coincidental: It is an inevitable outcome of the shift to a "knowledge Society [1]." The application of various kinds of knowledge to solve a range of problems has always played a central role in economic Development and social welfare. The relatively modern Concept of the knowledge society, however, refers to a society in which knowledge, rather than manual work, raw materials, and capital, is the most important source of economic and social development. In the context of higher education, a competence may be understood as the combination of skills, knowledge, attitudes, values, and abilities that underpin effective and/or superior performance in a professional area. However precedence is given to the learning of the technical course, in both academic and professional life, and crucial for engineering [6].

A. Problem Statement

The ideal goal of the project is to deal with current scenarios which are facing by the students who cannot be able to solve or crack the real world tasks with respect to technical aspects. Supporting analysis which leads to beneficial to help the students by providing a software platform which can monitor the performance of the individuals can be accomplished by:

- Giving certain technical tasks.
- Measuring the tasks performed.
- Analyzing the task.
- Periodical Mentoring of students.

By doing all these things we can azure that the students can be able to involve in dealing, the IT perspective tasks and can be enlightened to face the world.

B. Objective and Scope of the Project

The **goal** of **formative assessment** is to monitor students and help them to get more required materials for their improvement in lacking topics and check their progress in the course. More specifically, **formative assessments:** help students identify their strengths and weaknesses and target areas that need more hard work. help faculties to recognize where students are struggling and address problems immediately.

Formative assessment and feedback allows the student to formulate understanding of specified learning objectives while receiving feedback on their progress to assist in their improvement at a targeted subject; with the application of formative strategies including: inquiry (the student asking questions), immediate feedback (provided by the instructor to the student), self-reflection, and adjustments to the instructional process based on the students response and understanding of the current method.

Assessment is vital to the education process. This formative assessment technique can be used by the teachers to guide the students in a different way to reach their respected task and find the required skill needed for them.

- As the students receive feedback from the instructors, they self-reflect and assess their progress while determining how they may improve; students must actively be engaged in their learning to achieve positive outcomes.
- Research that was conducted by Havnes et al, reported that the practice of giving students feedback was more dependent on the subject being taught verses the student's reception or the school in which the subject is being taught.
- The students involved in the research further gave the following feedback as it relates to feedback during instruction in the classroom: feedback practice relates to the educational attitude and beliefs of the individual teacher, immediate review of subjects related to corrections after a test, and the perception of the feedback by the student.
- Formative assessment and feedback can be very useful in the classroom as we develop our perspective courses and analyze as we teach the course determining how well the class is learning the material presented and the need for change in the way we are teaching the course or the material being taught, or both.

C. Importance of the Project

The benefits of formative assessment include:

- Defined learning goals.
- Increased rigor.
- Improved academic achievement.
- Enhanced student motivation.
- Increased student engagement.
- Focused and targeted feedback.
- Personalized learning experiences.
- Self-regulated learners.

II. LITERATURE REVIEW

A. Development of Procedures to Assess Problem

In the context of higher education, a competence may be understood as the combination of skills, knowledge, attitudes, values, and abilities that underpin effective and/or superior performance in a professional area. The aim of the work reported here was to design a set of procedures to assess a transferable competence, i.e., problem solving, that is basic for learning, in both academic and professional life, and crucial for engineering. The study involved a total of 71 students enrolled at three universities at two different stages of their studies. The development phases of the assessment device included an analysis of the competence and its facets, the design of the assessment task, the development of criteria to rate student performance, and the analysis of the basic psychometric properties for assessment methods in the area of education. The conclusion was drawn that the training process and the elaboration of scoring criteria are costly but necessary if objectivity in the interpretation of results is to be guaranteed. The main achievement of this project was the development of a procedure that measures learning outcomes, and more specifically; problem solving [2].

B.*Using Learning Analytics to Assess Capstone Project Teams*

Succeeding in today's software engineering (SE) workforce requires mastery of practical teamwork skills. The unacceptably high failure rate of SE projects stems from failures in communication, organization, and team dynamics,_ as well as the difficulty of engaging in projects across diverse and geographically distributed teams. ACM and the IEEE Computer Society addressed these problems in Computer Science Curricula 2012, which placed much emphasis on the need for better coverage of teamwork-based learning (TBL) and project-based learning (PBL) in higher education.

Despite increased awareness of its importance, teamwork in SE projects has proved to be difficult to teach. Typically included as part of the capstone course in college level SE programs, PBL and TBL present unique challenges as student teams build projects during the course. Little is known about the best way to objectively and quantitatively assess student progress in acquiring these skills. In education, assessment can be defined as "the use of empirical data on student learning to re_ ne programs and improve student learning."- Here, we consider a novel approach to student learning assessment in the SE teamwork context that uses learning analytics based on modern machine learning (ML) tools. Leveraging the rich and objective data generated by such tools, it's more objective and consistent, and scales better, than traditional student learning assessment methods. More importantly, the approach allows early prediction of which SE teams are more likely to fail, facilitating intervention and thus education effectiveness. It's also applicable to SE training and project management in industry. The approach emerged from the SETAP project using data from joint SE classes concurrently taught at San Francisco State University, Florida Atlantic University, and Fulda University in Germany. The sidebar "SETAP: Software Engineering Teamwork Assessment and Prediction" outlines the project's goals and methodology [3].

C. *Measuring the Impact of Agile Coaching on Students' Performance*

Nowadays, considerable attention is paid to agile methods as a means to improve management of software development processes. The widespread use of such methods in professional contexts has encouraged their integration into software engineering training and undergraduate courses. Although several research efforts have focused on teaching Scrum through simulating a software development project, they covered only the learning of programming practices within a Scrum team. Furthermore, few studies tackle nontechnical skills other than the development practices themselves. The work presented here introduces an original Scrum-based training model enhanced with agile coaching to maximize student performance. This was validated by a case study on a capstone project in a Scrum course. This paper summarizes the positive results of introducing agile coaching, which resulted in approximately 22% more coverage of software engineering practices. In addition, a survey data showed that, compared to students who did not receive coaching, coached students gained valuable insight into the internalization of Scrum, problem solving, and guidance by means of checkpoint meetings [4].

D. Improving the Laboratory Learning Experience

The overall goal of our Software Engineering Teamwork Assessment and Prediction (SETAP) project is to develop effective machine-learning-based methods for assessment and early prediction of student learning effectiveness in software engineering teamwork. Specifically, we use the Random Forest (RF) machine learning (ML) method to predict the effectiveness of software engineering teamwork learning based on data collected during student team project development. These data include over 100 objective and quantitative Team Activity Measures (TAM) obtained from monitoring and measuring activities of student teams during the creation of their final class project in our joint software engineering classes which ran concurrently at San Francisco State University (SFSU), Fulda University (Fulda) and Florida Atlantic University (FAU). In this paper we provide the first RF analysis results done at SFSU on our full data set covering four years of our joint SE classes. These data include 74 student teams with over 380 students, totaling over 30000 discrete data points. These data are grouped into 11 time intervals, each measuring important phases of project development during the class (e.g. early requirement gathering and design, development, testing and delivery). We briefly elaborate on the methods of data collection and describe the data itself. We then show prediction results of the RF analysis applied to this full data set. Results show that we are able to detect student teams who are bound to fail or need attention in early class time with good (about 70%) accuracy. Moreover, the variable importance analysis shows that the features (TAM measures) with high predictive power make intuitive sense, such as late delivery/late responses, time used to help each other, and surprisingly statistics on commit messages to the code repository, etc. In summary, we believe we demonstrate the viability of using ML on objective and quantitative team activity measures to predict student learning of software engineering teamwork, and point to easy-to-measure factors that can be used to guide educators and software engineering managers to implement early intervention for teams bound to fail [5].

E. Existing System

The application of various kinds of knowledge to solve a range of problems has always played a central role in economic Development and social welfare. The relatively modern Concept of the knowledge society, however, refers to a society in which knowledge, rather than manual work, raw materials, and capital, is the most important source of economic and social development. In the context of higher education, a competence may be understood as the combination of skills, knowledge, attitudes, values, and abilities that underpin effective and/or superior performance in a professional area. The aim of the work reported here was to design a set of procedures to assess a transferable competence, i.e., problem solving, that is basic for learning, in both academic and professional life, and crucial for engineering.

Advantage:

- Determine Prediction Accuracy
- Discover factors that contribute to prediction
- Tool Logs.

Disadvantage:

- No suggestions will be given for the performed tasks.
- Not concerned about the technical growth of students.
- Periodic mentoring is not at all achieved.

Everything is uploaded from the previously uploaded databases, to all students.

F. Proposed System

Formative assessment and feedback allows the student to formulate understanding of specified learning objectives while receiving feedback on their progress to assist in their improvement at a targeted subject; with the application of formative strategies including: inquiry (the student asking questions), immediate feedback (provided by the instructor to the student), self-reflection, and adjustments to the instructional process based on the students response and understanding of the current method. Assessment is vital to the education process. Teachers using formative assessment approaches and techniques are better prepared to meet diverse students' needs – through differentiation and adaptation of teaching to raise levels of student achievement and to achieve a greater equity of student outcomes. As the students receive feedback from the instructors, they self-reflect and assess their progress while determining how they may improve; students must actively be engaged in their learning to achieve positive outcomes [1].

Advantage:

- Training the students to gain technical knowledge.
- Preparing them to solve real world programming tasks.
- Daily/Weekly basis suggested links will be uploaded.
- Fully Automated software evaluation.
- Overall evaluation of the tasks and assigned a final grade point.

II. METHODOLOGY

This section describes the development of a procedure to assess the basic transferable competence of problem solving. Students were set a task that consisted of a problem statement and a set of questions that measured Students' problem-solving ability. In addition, described is how basic objectivity and valid data for the assessment procedure were assured. Finally, some results are given on how the two academic levels participating in the study compare.

A. Participants

In designing the task, an aim was that it should be useful for measuring the progress in the performance of the competence (discriminant validity). The initial plan was to sample first -year students, which would involve all computer science students.

B. Procedure

The design of these assessment procedures included:
1) a detailed analysis of the facets of each competence; 2) the design of assessment task covered these competences and their specific facets and have various levels of difficulty to accommodate the development of the student over the two academic years covered by the study; 3) the development of assessment criteria with an acceptable inter-rater reliability when grading students' work; and 4) determination of the basic psychometric properties that any measurement device should show, such as inter-rater agreement, internal consistency, and validity (content and discriminant validity), following the standards for educational and psychological testing.

1) Curriculum Analysis and Student Recruitment: The curriculum of the study program as prescribed by the university was analyzed to find common content that could be used from first year. The tests were taken by student volunteers enrolled in those academic years.

2) Task Design: To develop a task appropriate to comprehensively measure how flow of procedures are associated with transferable competences, the competences were first analyzed in terms of their component parts, i.e., the aspects present in a problem-solving task analysis. The test questions were then mapped to this scheme: 1) identifying the problem; 2) creating a strategy to solve the problem; 3) finding additional information if necessary; 4) applying knowledge needed for problem solving; and 5) evaluating the adequacy of the solution and, if necessary, restarting the cycle as shown in. Fig. 1 below:

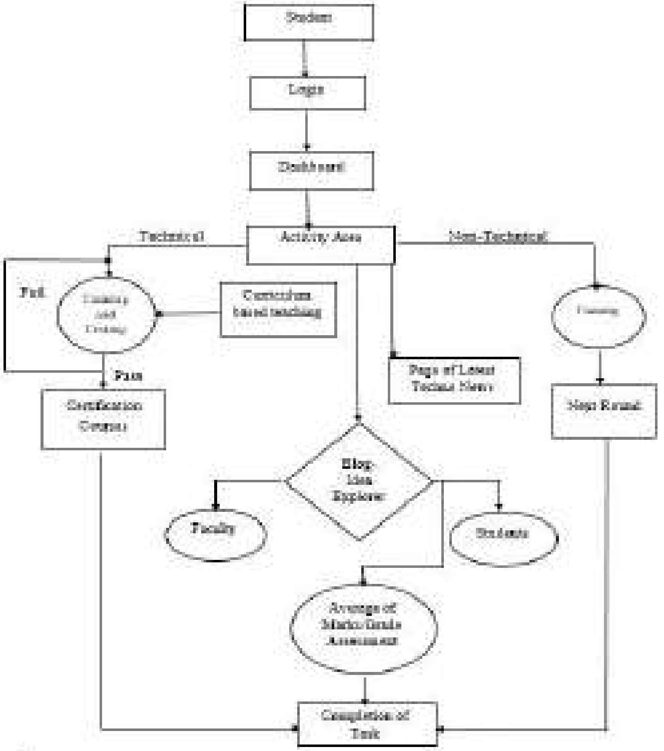


Fig. 1. Students Training Procedure Flow Diagram

3) Internal Consistencies of the Test: Naïve Bayes is a commonly used index for this feature, which reveals how the task focuses on what it intends to measure.

In this paper, we apply machine learning algorithms to detect the failure steps on taking tests and to search the required concepts that need to be overcome in order to learn the failed step.

C. Random Forest

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time [6]. It is also one of the most used algorithms, because its simplicity and the fact that it can be used for both classification and regression tasks. Random Forest is a supervised learning algorithm. Like you can already see from its name, it creates a forest and makes it somehow random. The forest "it builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. One big advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems. I will talk about random forest in classification, since classification is sometimes considered the building block of machine learning. Below you can see how a random forest would look like with two trees. Fig. 2.

D. Naïve Bayes

Bayes' theorem finds many uses in the probability theory and statistics. There's a micro chance that you have never heard about this theorem in your life. Turns out that this theorem has found its way into the world of machine learning, to form one of the highly decorated algorithms [7]. In this article, we will learn all about the Naive Bayes Algorithm, along with its variations for different purposes in machine learning. Bayes' theorem does exactly that.

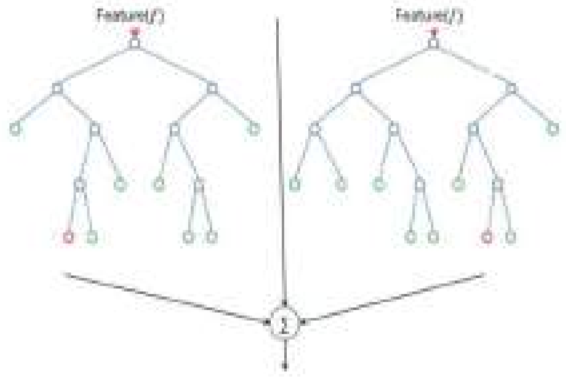


Fig. 2. A Splitted Random Forest tree which Combined on Final Traversal.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Applications of Naive Bayes Algorithms

- **Real time Prediction:** Naive Bayes is an eager learning classifier and it is sure fast thus, it could be used for making predictions in real time.
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System:** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information.

FUTURE ENHANCEMENT

1. The application can be extended in which it includes all the streams of engineering course.
2. The application can have online courses where the students can purchase any of the course for certification purpose.

III. CONCLUSION

In this article we documented the creation of online assessment, a programming platform whose intent is increasing student task completion and engagement, especially in girls, while teaching basic CS concepts, as a way of promoting interest towards CS-related careers and as a way of contributing to the increasingly important discussion of how to introduce engineering students to CS concepts in an engaging way. We evaluated this in two observational studies: an online competition. Combining the results of these studies allowed us to identified general findings with largely quantitative data, and confirm, construct explanations and understandings of these findings, from classroom field work conducted in real school settings. In both studies, most indicators of engagement (participation, task completion, interest, willingness to learn more and self-reported interest) were found.

ACKNOWLEDGEMENT

We would like to thank the project guide **Mrs. Shruthi B Gowda**, **Assistant Professor**, Department of Computer Science and Engineering for her guidance, continuous support and the students who participated in the experiments for their enthusiasm, and the reviewers for their valuable feedback.

REFERENCES

[1] Luciana Benotti , Maria Cecilia Martinez, and Fernando Schapachnik A Tool for Introducing Computer Science with Automatic Formative Assessment, IEEE Transactions On Learning Technologies, Vol. 11, No. 2, April-June 2018.

[2] Jorge Pérez, Carmen Vizcarro, Javier García, Aurelio Bermúdez, and Ruth Cobos Development of Procedures to Assess Problem-Solving Competence in Computing Engineering, IEEE Transactions On Education, Vol. 60, No. 1, February 2017.

[3] Dragutin Petkovic, San Francisco State University Using Learning Analytics to Assess Capstone Project Teams, Published byThe IEEE Compute R Society 0018-91 62 / 16 /$33.00 © 2016 IEEE.

[4] Guillermo Rodríguez, Álvaro Soria, and Marcelo Campo Measuring the Impact of Agile Coaching on Students' Performance, IEEE Transactions on Education 0018-9359 © 2016 IEEE.

[5] Sasha Nikolic, Member, IEEE, Peter James Vial, Senior Member, IEEE, Montserrat Ros, Member, IEEE, David Stirling, Senior Member, IEEE, and Christian Ritz, Senior Member, IEEEImproving the Laboratory Learning Experience: A Process to Train and Manage Teaching Assistants IEEE Transactions On Education, Vol. 58, No. 2, May 2016.

[6] Marc Sosnick-Perez, Kazunori Okada Rainer Todtenhoefer, Shihong Huang Using the Random Forest Classifier to Assess and Predict Student Learning of Software Engineering Teamwork 978-1-5090-1790-4/16/$31.00 ©2016 IEEE.

[7] Zhenchong Zhao, Xiaodan Wang Multi-segments
Naïve Bayes classifier in likelihood space IET
Computer Vision Volume: 12, Issue: 6 , 9 2015

A Survey on Phishing Detection and Prevention Techniques

Chandana R Gowda[1], Deepanjali B I[2], Nehal Bopaiah[3], Nisha K M[4], S Meenakshi Sundaram[5]
8th Semester B.E Student[1, 2, 3, 4], Assistant Professor & Head[5]
Department of Computer Science and Engineering GSSS Institute of Engineering and Technology for Women, Mysuru, Karnataka Affiliated to VTU Belagavi.
chanadanaravikumar97@gmail.com[1], deepanjali624@gmail.com[2], nehalbopaiah186@gmail.com[3], niskakm1234@gmail.com[4], 1965sms@gmail.com[5]

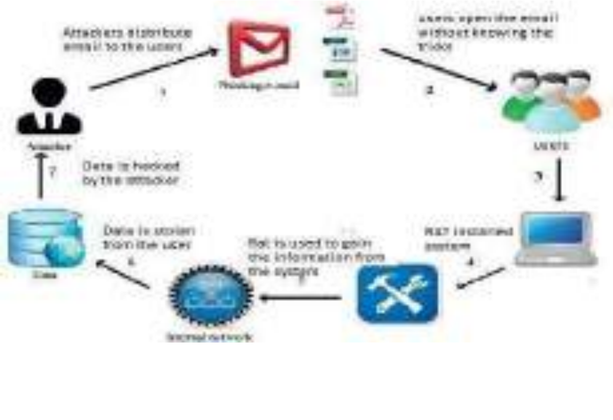
***Abstract*—Cybersecurity is the process of protecting the networks and confidential data from being accessed illegally by the attackers. The various cyber attacks include malware, man in the middle, phishing etc. Phishing is one of the major active attacks in Cybersecurity. According to the various statistics available phishing attacks has increased from 76% to 83% in 2018. This paper focuses on identifying and analyzing the various works related to phishing. This is achieved by summarizing some of the existing algorithms, datasets from trusted repositories. This paper also gives a brief introduction to various phishing tools.**

Keywords—Phishing; Link Guard; Cybersecurity; Metasploit; Phinn; Nmap; Email; Security;

I. Introduction

Security in the cyber world is necessary to protect the data from various threats. Phishing is the process of acquiring sensitive information such as username, Password, credit card details that takes place via email spoofing, instant messaging and many other modes of internet communication. Often emails that are phished contain links that direct the user to a fake webpage that have the look and feel of the genuine webpage. The user may then be asked to enter the personal information which the phisher uses for various other malicious activities.

Figure 1: Phishing attack steps [1]


Attackers distribute email to the users
Phishing e-mail
1
users open the email without knowing the trick
2
Attacker
Users
3
Data is hacked by the attacker
7
RAT installed system
4
Data is stolen from the user
6
Rat is used to gain the information from the system
5
Data
Internal network

A Phished email is sent to large number of people after which the phisher will calculate the percentage of people who has read and entered the information that is tracked by the phisher. It is very hard for us to differentiate between a genuine and a phishing website.

In December 2015, about 630,494 phishing sites were detected according to the anti phishing working group. The highest phishing attacks are found in USA and Belize having a percentage of 76.8 and 81.3 respectively.

As 2018 drew to an end, many cyber security reports published their findings on the most common types of attacks that targeted small and large organizations. The focus in 2018 shifted from private people to more and more attacks targeting businesses.

"Info security professionals reported a higher frequency of all types of social engineering attacks year over year. Phishing increased to 83% versus 76%. Spear phishing increased to 64% from 53%."[9]

A recent report by Kaspersky Lab states that 52% of business are worried about data breach stemming from their employees and acknowledge that they (especially non-IT employees) are the weakest link to their cyber security strategy. According to the reports about, 52% of businesses believe that employees are their major weakness in IT security, this is because of the careless actions of the employees. [10]

II. Literature Survey

A) Shekokar et al., describes phishing detection and prevention approach for both URL and Webpage similarity based detection. This is done with a help an algorithm called the Link Guard Algorithm. This algorithm has five conditions based on which the webpage or the URL is detected as phishing or not. The conditions includes the length of the URL, the special symbol @, IP address along with double slash and prefix of the URL. Once an URL is entered in the search tab the developed algorithm runs in the background and notifies if the entered URL is prone to phishing or not. A novel technique to visually compare if the page is suspicious. [2]

B) Abdulghani Ali Ahmed et al., proposes a solution that distinguishes the suspected web pages as legitimate web page and phishing web page by only checking the URLs. Features of the URLs are checked using the following criteria:

- The existence of the IP address is checked.
- The length of the URLs are checked (should not exceed 54 characters).
- The URLs should not have more prefixes or suffix separated by "-".
- The URLs should not contain "@" symbol.
- The position of the "//" symbol in the URL is checked. [3]

C) Barraclough et al., demonstrates the online toolbar which runs in the background of Internet Explorer continuously checking all websites and classifies it as a phishing, suspicious or legitimate website respectively against a set- data in real- time. The new toolbar system has been evaluated using a wide-ranging websites of 600 websites which consists of equal number of phishing websites, legitimate websites and suspicious websites (i.e. 200 each). The accuracy resulted was about 96%. [4]

D) Daeef et al., describes about the detection system that is developed based on URL features. This system is built on fact that most of the users make use of URLs to surf the internet. Since client side solutions were given utmost importance, this turned out to add more processing overhead at the client side which led to dissatisfaction, loss of trust among the users. Thus this detection method seems to be a good approach to detect fake URLs. The system proposed in this paper increased the rate of detection performance in real time which could also be integrated into such processes. The proposed system has the phishing URL detection accuracy of 93% and detects single URL in average time of 0.12 second. [5]

E) Naresh et al., demonstrates anti-phishing algorithm called Link Guard that extracts and uses the generic characteristics of hyperlinks .The analysis of the characteristics derived from an hyperlink were done on the data archive that was provided by the Anti-Phishing Working Group (APWG).Since the Link Guard works on the generic characteristics of an hyperlink, it can detect not only known but also unknown attacks of phishing. This Link Guard was implemented on Windows XP. Based on this implementation, experimental results verified that Link Guard works well in effective detection and prevention with minimal false negatives for both known and unknown phishing attacks. Their further analysis showed that Link Guard being light weighted can work well in real time and was successful in detecting 195 attacks out of 203 phishing attacks. [6]

F) Hewamadduma et al., provides information to the customers about the various phishing attacks, purpose of these attacks and the impact it causes to the customers and the bank. It mainly concentrates to find the various techniques to detect and prevent the phishing attacks such as unauthorized login attempts. The solution to detect and prevent these attacks is provided by the technologies such as analysis based on behavior, device identification and IP identification.

The Table 1 provides the security actions for login according to the deviations of the factors. Here the most important factors are cookie and IP address.

Table 1: Security actions to be taken for login in user anomalies [7]

IP	Device	Cookie	Time	Browser	OS	Action to be taken
Y	Y	Y	Y	Y	Y	Allow
Y	Y	Y	Y	N	N	Security Question
N	Y	Y	Y	Y	Y	Security Question
N	N	N	Y	Y	Y	Code+ Security Question
N	N	N	N	N	N	Email Notification +Block

Table 2: Overview of the literature survey

Paper	Methodology	Merits	Demerits
A	LinkGuard Algorithm	Accuracy is 96%.	Focuses mainly on the URL based phishing attacks. Implemented only in Windows XP.
B	Inspecting the URL's of fake web pages	Detects Phishing Webpage with accuracy-0.96	Only checks the validity of URL‟s
C	Feature-based online toolbar	Accuracy is 96% .	To be extended so that other web browsers such as Firefox, Chrome can be used.
D	URL Language Model: N-Gram	Accuracy is 93%.	Error rate are still high, requires additional efforts to reduce error
E	LinkGuard Algorithm	96% detection in real time	Does not handle CSS attack.
F	Detection based on anomalies, IP and device identification	Detect the unauthorized login attempts.	Accuracy of system is unknown or not calculated.

III. TOOLS

A. *Phinn*

Phinn’s look is similar to a login page in Google.There is allowance in Phinn for the company administrators to train and generate the customized extension of chrome,which is later distributed within the organization.With the help of machine learning algorithm i.e., convolutional neural network this particular extension analyses the content of the rendered page for visual similarity.

Phinn captures a screenshot when the login form on the browser is displayed which is then passed to the trained neural network and thus the result is displayed within fraction of seconds.

Convolutional neural networks have several benefits. As it works on raw image data obfuscation and minimization typically seen in phishing pages plays no role in the detection accuracy. Mainly it forces the attacker to deviate from the styling and branding that the user is used to, making the page appear more suspicious while lowering the burden on the users to be proactive.[11]

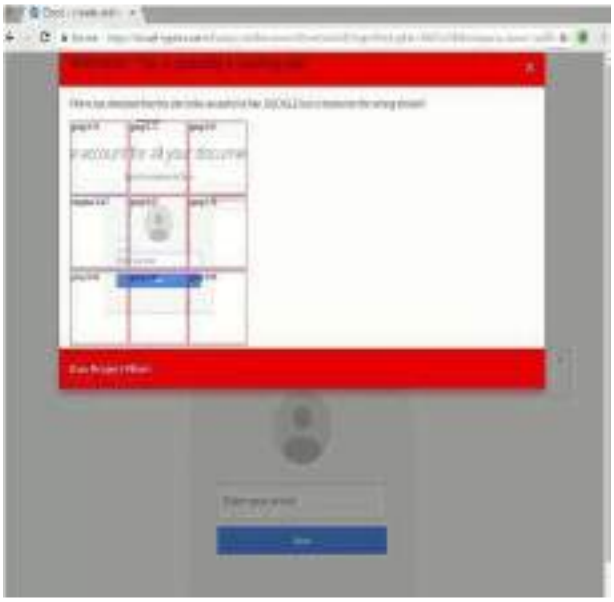


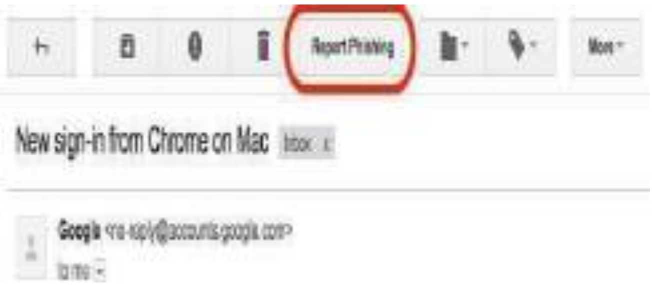
Figure 2: Working of Phinn [11]

B. *IsThisLegit*

IsThisLegit can be divided into three parts:

1. measuring
2. training
3. reporting

It measures organization's exposure to phishing by sending simulated phishing emails. This also teaches users how to spot real phishing attempts when they are inevitably stored in their inbox. It also enables users to easily report suspicious emails to their security team, by giving them the tools required to investigate and manage these reports. The dashboard lets analysts view, analyze, and respond to phishing reports. [11]



New sign-in from Chrome on Mac

Google <no-reply@accounts.google.com>
to me

Figure 3: Working of IsThisLegit [11]

C. *Metasploit*

Metasploit is one of the penetration testing tools available in Kali OS which is used find, exploit and validate the vulnerabilities. This tool holds the Metasploit framework along with many other commercial counterparts such as Metasploit Pro. The infrastructure, content and tools to perform variety as tests are provided by Metasploit framework. On the other hand Metasploit Pro holds multiple components which is combined together to form a complete testing tool.

D. *Nmap*

Nmap is another inbuilt security tool within Kali OS that provides the complete information about the systems present on the network. This information includes MAC address, IP address, the current OS the user is working and related information about the systems on the network. Nmap not only secures the network but also checks if the servers are configured correctly, ensures there is no open or unsecured ports thus ensuring the proper working of the firewall.

E. Social Engineering Toolkit

Social Engineer Toolkit (SET) is an open-source penetration testing tool available for social engineering in Kali OS. This tool is used to track the user information with the help of fake web pages that is developed using the SET toolkit. During the attack any credential entered by the user on the fake webpage is monitored by the attacker with the help of the IP address. The original webpage is faked intact where the user fails to distinguish between the real and fake one.

IV. DATASET

A. *UCI Repository*

Dataset Information

The dataset consists of different features related to legitimate and phishing websites. The features have been collected from 1353 different websites. The URLs of the phishing websites were collected from Phishtank data archive (www.phishtank.com). The URLs of the legitimate websites were collected from Yahoo, Chrome etc. There are 548 legitimate websites, 702 phishing websites and 103 suspicious websites out of 1353 websites. The websites which are suspicious can either be phishy or legitimate.

Table 3: Abstract of the dataset [8]

Characteristics of the Dataset	Multivariate
Count of Instances	1353
Attribute Representation	Integer
Count of Attributes	10
Tasks Associated	Classification of websites
Representation of Missing Values	N/A
Donation Date	2016-11-02
Number of Web Hits	47076

Attribute Information

The dataset consists of 10 features such as SFH, Request URL, prefix/suffix, URL Anchor, Number of sub domains, Web traffic, Domain age, URL length, IP, Class. The features holds categorical values such as legitimate, phishing and suspicious and these values have been represented with numerical values 1, -1 and 0 respectively.

B. Kaggle

DatasetInformation

The dataset is a multivariate and consists of 32 features. The features have been collected from 11055 websites . The features holds categorical values such as legitimate, phishing and suspicious and these values have been represented with numerical values 1, -1 and 0 respectively.

Attribute Information

The dataset consists of 32 features such as Index, IP address, Length of the URL, Shortening service, Presence of @ symbol, double slash redirecting, Presence of prefix/suffix, Number of subdomains, SSL final state, domain registration length, Favicon, DNS record, links pointing to page, web traffic, page rank, Google index, port, HTTPS token, request URL, Anchor URL, links in tags, SFH, submitting to email, abnormal URL, redirect, on mouseover, right click, pop up window, Iframe, Domain age, statistical report and result. [12]

CONCLUSION

Phishing is the most dangerous threat in today’s world. Victims of these attacks have increased tremendously during the last decade. Cybercriminals are changing their strategies to gain people’s personal information from time to time.

The paper consists of the survey done on some of the phishing detection and prevention techniques. Each technique that is discussed has its own drawbacks in terms of accuracy and performance. There is no single system developed till now to detect all types of phishing attacks, therefore in future there is a need to develop a single system to detect all these attacks with high accuracy and performance.

ACKNOWLEDGEMENTS

This survey was carried out at the “Innovative Smart Systems and IoT Laboratory” under the sponsorship of “Vision Group on Science and Technology”, Government of Karnataka under the scheme of “Centre of Innovative Science, Engineering and Education”. The team would like to acknowledge the VGST and Dr. M. Shivakumar, Principal, GSSSIETW for their continuous encouragement and support for the completion of the work.

REFERENCES

[1] Muhammet Baykara and Zahit Ziya Gurel. "Detection of phishing attacks." IEEE, 2018.

[2] Shekokar, Narendra M., Chaitali Shah, Mrunal Mahajan, and Shruti Rachh. "An ideal approach for detection and prevention of phishing attacks." *Procedia Computer Science* 49 (2015): 82-91.

[3] Abdulghani Ali Ahmed and Nurul Amirah Abdullah. "Real-time detection of phishing websites." IEEE, 2016.

[4] Barraclough, P. A., Graham Sexton, and Nauman Aslam. "Online phishing detection toolbar for transactions." In *Science and Information Conference (SAI), 2015*, pp. 1321-1328. IEEE, 2015.

[5] Daeef, Ammar Yahya, R. Badlishah Ahmad, Yasmin Yacob, and Ng Yen Phing. "Wide scope and fast websites phishing detection using URLs lexical features." In *Electronic Design (ICED), 2016 3rd International Conference on*, pp. 410-415. IEEE, 2016.

[6] Naresh, U., U. VidyaSagar, and C. V. MadhusudanReddy. "Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm." *Proc. IOSR* 14, no. 3 (2013): 28-36.

[7] Hewamadduma, Shammi Ishara. "Detection and prevention of possible unauthorized login attempts through stolen credentials from a phishing attack in an online banking system." In *Research and Innovation in Information Systems (ICRIIS), 2017 International Conference on*, pp. 1-6. IEEE, 2017.

[8] UCI data repository [online], available: url:https://archive.ics.uci.edu/ml/datasets/Website+Phishing#, [Accessed Feb.4,2019]

[9] Proofpoint: Annual state of Phish report, [Accessed Feb.4,2019]

[10] Lucy Security [online], available: url: https: //lucysecurity.com/category/ blog/ , [Accessed Feb. 4,2019]

[11] Duo Security, Inc. [US] available: url: https: //duo.com/blog/new-open-source-phishing-tools-isthislegit-and-phinn, [Accessed Feb.4,2019]

[12] Kaggle [online], available: url: *https://www.kaggle.com/akashkr/phishing-website-dataset*, [Accessed Feb. 4,2019]

Dynamic Geo Based Survey Document Retrieval

Mohit D P [1], *Nivedita* [2], *Pallavi V R* [3], *Vinutha H M* [4], *Prof. Vinayashree* [5]
8 th Semester B.E Student[1, 2, 3, 4] & Professor[5]
Information Science & Engg. Department, Vidya Vikas Institute Of Engineering & Technology, Mysore
mohitgowdacoorg994@gmail.com[1], *niveditasangolgikar7@gmail.com*[2], *pallaviradder97@gmail.com*[3],
vinugowda842@gmail.com[4], *vinaya448@gmail.com*[5]

Abstract— ***The availability of the locations using GPS has made many changes nowadays. The smartphones are having this inbuilt location-based feature of GPS where we use to display in maps. For Android Operating Systems, Google Maps API V2 is one of the location-based APIs available. This paper discusses the utilization of Google Maps API V2 to design an app which aims to simplify the Land Office and the people who want to buy the land to obtain more accurate information about the land which they want to buy . By using GPS in mobile, we can get location to perform a survey to determine the coordinate points of the land which the buyer wants to buy. The main objective of this paper is to identify the exact survey details of a piece of land and to reduce the litigation issues created by the land brokers or real estate people on the people who are purchasing it. The main advantages of this application is that ,it reduces the ambiguity in identifying the plot, it is accurate and reliable, litigation issues can be reduced, the buyer can be saved from incurring loss, fraudulent activities can be terminated.***

Keywords— ***Global Positioning System(GPS),Google Maps API V2, Android, National Land Agency (NLA),), Google Maps, Survey Department of Government.***

I. INTRODUCTION

Surveys are used to identify boundaries and features of land to determine ownership. Survey Department of Government surveys the land and draws boundaries across and are given the plot of land with specific survey numbers. They are employed in construction projects ranging from building fences to entire cities. Survey is essential for data collection, with the end result of a data and information in the form of maps. Generally, this results in the form of a hardcopy form. This contains the administrative boundaries of land ownership, which change at any time.

When a person is willing to purchase a land belonging to someone else he/ she has to first go to government offices collect the survey details and maps of that particular land and then manually go to that plot and check the measurement and other parameters. Sometimes the broker or the real estate people who act as mediators between the seller and the buyer play double game in order to take the benefits of commission and make profits. Sometimes the buyer of the land after purchasing the land he might have built a house or started work on constructing a building, suddenly encounters worst scenarios by others claiming that this particular land belongs to us or the survey number of this land is different etc. The buyer will end up in litigation issues and cases. The detailed produced by the seller or the broker will not be 100% genuine, this may put the buyer in despair.

In order to overcome all these fraudulent activities we have developed this application which will give all the information regarding the plot which the buyer wants to purchase, so that the buyer doesn't get carried away with false perceptions and incur loss.

II. EXISTING SYSTEM

In the existing system , if somebody wants to buy a piece of land, the buyer either collects the details of the plot which he/ she is interested to buy from the broker or real estate people. These people in order to make profits will not provide the full details of the plot such as the family tree details of the owner of the plot, the boundaries and other measurement parameters of *the plot. The buyer has to collect all these details from a government office or manually waste lot of time in gathering all the information regarding the plot. Time consuming. Tedious to frequently visit the physician for suggestions.*

In existing system, it's difficult to get proper details

about land or plot. The buyer has to collect all these details from a government office or manually waste lot of time in gathering all the information regarding the plot. Hence there is a need of application which overcomes from problems of existing system.

III PROPOSED SYSTEM

In the proposed system, an android application is developed for the survey department of the Government. This application can be used both by the Government as well as the buyers of the land. In this application complete survey details and the geographical location details of each portion or plot of land, owner details, owner's family tree, shape of the plot etc. will be accurately fed into a server. When the buyer goes to the plot and opens this application a clicks the button in, it will show the complete details of that particular plot with accurate survey number where he / she is standing. The buyer can even download the document, whole document can be downloaded with one click and can use this document in future.

In the proposed system, there are mainly two modules, one is Admin module and the other is user

module. In Admin module there is two other sub-modules, they are Manage User module and Manager Service Provider (Survey Department). In Manage User module admin can add, delete and view users registered into the applications. In Manager Survey Provider module, admin can manage service providers by adding details of service providers and also admin can view service providers. In User module, Register and Login module, the user has to get registered to our application and can login to view the app modules, Receive Notification module is for the users to get notifications about the confirmation message to access the land details, Request Land Details module is to request for land details provided by survey departments and View Plot/ Land Details module is used to view details about the lands or plots. The Manager Service Provider module contains Login Module where the service provider has to get logged-in to the application to view the app modules, View Request module is used by the service provider to view request sent by user to access the land details, Send Notifications module is used by the service provider send confirmation message to user for accepting service to provide land details and Provide Land Details module where the service provider allow the user to access the land details.

The advantages of the proposed system is that, it can reduces the ambiguity in identifying the plot, it is accurate and reliable, litigation issues can be reduced, the buyer can be saved from incurring loss and fraudulent activities can be terminated.

IV SYSTEM DESIGN

The purpose of the design phase is to plan a solution of the problem specified by the requirements document. This phase is the first step in moving from the problem domain to the solution domain. In other words, starting with what is needed; design takes us toward how to satisfy the needs.

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system. DFDs can also be used for the visualization of data processing (structured design). On a DFD, data items flow from an external data source or an internal data store to an internal data store or an external data sink, via an internal process. The data flow diagram for admin and user of our system is shown in Fig. I and Fig.2 respectively.

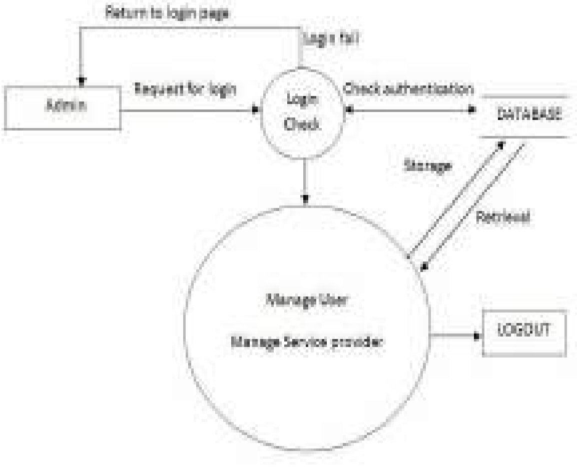
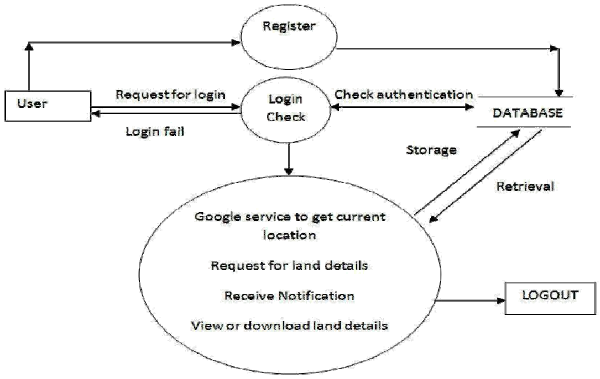


Fig. I DFD for Adm



V RESULTS AND DISCUSSIONS

Our android application is implemented using object oriented programming language. Following snapshots below shows our application pages.

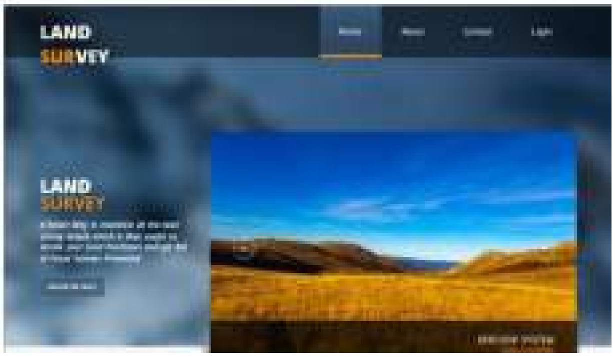
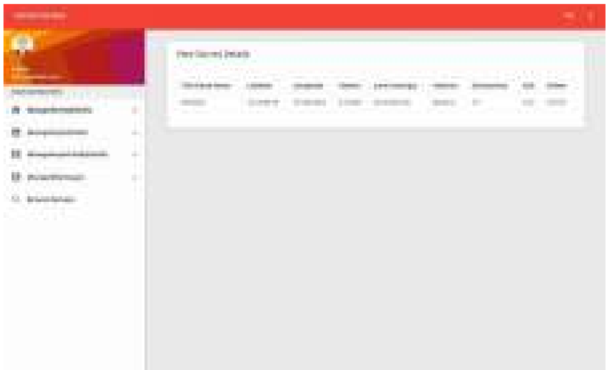


Fig. 3 Home Page

Fig. 3 shows the home page of our applications, where the user can login to our application, they can contact us and they can know about us more.



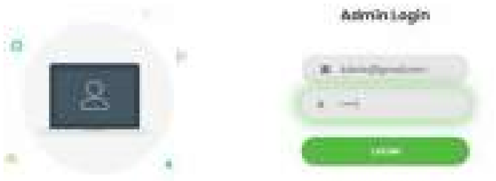


Fig. 4 Login Page

Fig. 4 shows the Login page where both the admin and the user can login to the application.

Fig. 5 Add New Survey Details

Fig. 5 shows Add New Survey Details page where the admin adds the new survey details of the land by giving details such as survey number of the land, its latitude and longitude, its radius, owner name, type of land e.g. Agriculture land, dimension of the land in acres, registration date and the address.

Fig. 6 Add Buyer Details

Fig. 6 shows the page to add the details of the buyer. This page takes the information such as parent survey number and new survey number of the land, *latitude and longitude, radius, dimension, address*

Fig. 7 View Survey Details

Fig. 7 shows the page to view the survey details of the land which are surveyed.

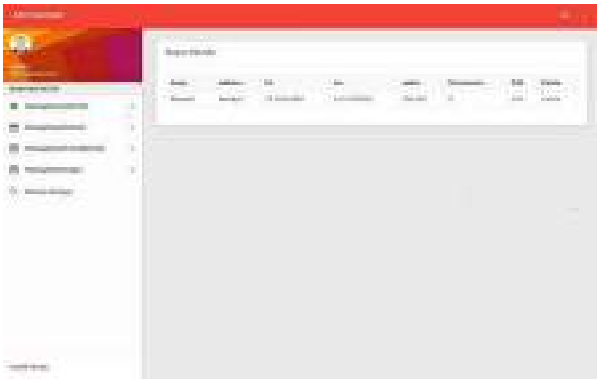


Fig. 8 View Buyer Details

Fig. 8 shows the page to view the details of the buyer who wish to buy the land.

Fig. 9 Add Buyer Family Details

Fig. 9 shows the page where the buyer needs to add their family details by uploading some certificates as

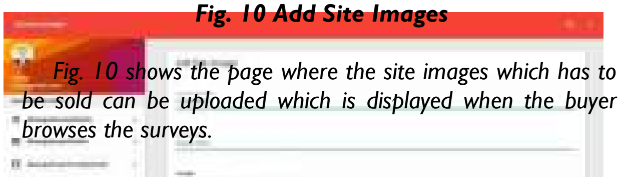


Fig. 10 Add Site Images

Fig. 10 shows the page where the site images which has to be sold can be uploaded which is displayed when the buyer browses the surveys.

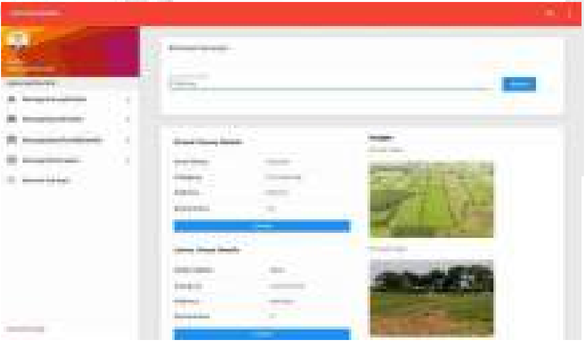


Fig. 11 Browse Surveys

Fig. 11 shows the page where the buyer can browse the surveys.

VI CONCLUSION

Our land survey is an android application using which users can view details of lands. It thus saves time, effort and reduces the litigation issues created by the land brokers or real estate people on the people who are purchasing it. This application can be used both by the Government as well as the buyers of the land

REFERENCES

REFERRED WEBSITES

[1] Android Developer Guide:http://developer. android. com /guide/index.html.

[2] Android API :http://developer.android.com/ reference

/packages.html

[3] http://www.java.com/en/download/faq/whatis-java.xml

REFERRED BOOKS

[1] Software Engineering, Ian Summerville, Sixth Edition,

Pearson Education Ltd, 2001.

[2] Wikipedia, the free encyclopedia

[3] Android Application Development By O'reily.

[4] Android Programmer's Guide by Jerome Dimarzio.

[5] Android Application Development by Reto.

[6] Abidin HZ et al. (2009) Crustal deformation studies

in

Java (Indonesia) using GPS. J. Earthquake Tsunam 3(2): 76-88.

[7] Thomas A.Powell, HTML, "CSS: The Complete Reference", 5th Edition, Tata McGraw- Hill education, 2010.

[8] Badan pertanahan nasional. http:// kotsalatiga. bpn.

go.id/Tentang-Kami/Sekilas.aspx, Accessed on. Agustus 08, 2017.

[9] Sahoo, B. P. S and Rath, Satyajit, " Integrating GPS, GSM and Cellular Phone for Location Tracking and

Monitoring," Proceedings of the International

Conference on Geospatial Technologies and Applications, IIT Bombay, Mumbai, India, 2012, February

Easy Delivery through Drone Stations

Narayan Muduli[1,] Shruthi B Gowda[2],
Student[1], Assistant professor[2]
Computer Science and Engineering,
Vivekananda Institute of Technology Bangalore, India
narayankumarmuduli@gmail.com[1],shruthi.b.gowda@gmail.com[2]

Abstract--**E-commerce and retail companies are seeking ways to cut delivery times and costs by exploring opportunities to use drones for making last mile delivery deliveries. This paper addresses the delivery concept of a truck-drone combination along with the idea of allowing autonomous drones to fly from delivery trucks, make deliveries, and fly to any available delivery truck nearby. We present a mixed integer programming (MIP) formulation that captures this scenario with the objective of minimizing the arrival time of both trucks and drones at the depot after completing the deliveries. A new algorithm based on insertion heuristics is also developed to solve large sized problems containing up to a hundred locations. Experiments are conducted to compare the MIP solutions with those obtained from different models with single truck, multiple truck and a single truck and drone system, as well as test the performance of the proposed algorithm.**

INTRODUCTION: Importance of drone delivery services is increasing. However, the operational aspects of drone delivery services have not been studied extensively. Specifically, with respect to truck-drone systems, researchers have not given sufficient attention to drone facilities because of the limited drone flight range around a distribution center. In this paper, we propose a truck-drone system to overcome the flight-range limitation. We define a drone station as the facility where drones and charging devices are stored, usually far away from the package distribution center. The traveling salesman problem with a drone station (TSP-DS) is developed based on mixed integer programming. Fundamental features of the TSP-DS are analyzed and route distortion is defined. We show that the model can be divided into independent traveling salesman and parallel identical machine scheduling problems for which we derive two solution approaches. Computational experiments with randomly generated instances show the characteristics of the TSP-DS and suggest that ourdecomposition approaches effectively deal with TSP-DS complexity problems.E-commerce and m-commerce increases theimportance of efficient logistics. In 2013, Amazonannounced drone technology as a future logistic innovation, and many companies have invested into drone research. For example, Amazon unveiled *Amazon Prime Air*, and Google announced *Project Wing* [1], [2]. Drones have many advantages over the typical truck delivery system [3], [4]. As drones

operate independently, they are free from operating labor costs and have relatively unlimited working time. Further, they movethrough the air and thus avoid the traffic congestion problemsof ground transportation. These advantages lead to the

highly energy-efficient use of drones. Moreover, the transportationcost per kilometer is much lower than that of othermeans. However, because of technological limitations, a dronecan carry only one parcel of limited weight and volume,and it can deliver to a single customer within a short flight range. To overcome these limitations, drone and truck deliveryservices can be used such that the characteristics of one complementthe other. Table I summarizes comparison of trucksand drones [3], [4]. To demonstrate the combined means ofdelivery, the HorseFly team at the University of Cincinnati

developed a system in which a drone can attach to and launchfrom a truck [4].Drones seem a good logistic alternative for industries, butthe technology needs further development to overcome some

realistic problems. Aside from controlling a certain type ofdrones [5] or motion sensing issues [6], battery capacity is amain concern for drone utilization. As many distribution centerswith drone facilities are located far from central cities,relatively few customers are serviceable by drones. For thisreason, large retail companies such as Amazon strive to buildmore distribution centers near major cities, but the expensesof constructing distribution centers are still a huge obstacle tocompletion. To deal with this logistical problem, a differentconcept of drone facilities is proposed. Roblin [7] introduced*Pylons* Dronairports, which contain drone recharge and shelterdevices. Designed by Bruni and Sardo, these compact devicescan be easily installed any place. In addition, Amazonplans to use street lights and church steeples as drone docking stations[8]. Another problem is that the weight and volume.

Capacities of drones are not enough to accommodate commercialdelivery services [1]. Drone security is also affectedby issues with GPS and sensor accuracies [4].Because many researchers and companies have tried to overcomethese problems, some companies have been able toutilize drones for commercial purposes. For instance, DHLexpress launched the first commercial delivery drone, called*Parcel copter*, in 2014 [7], [8], and the plan for building thefirst airport for drones is ongoing in Rwanda [3]. In contrast,research on the operational aspects of drone delivery hasbeen neglected, and only a handful of papers in drone-trucksystems have been presented. One of the initial papers .The traveling salesman problem (TSP) in tandem with droneswas conducted by Murray and Chu [4], who described twodifferent models. The flying sidekick TSP (FSTSP) describesthe way a single drone is used with a truck.Hybrid detection combines misuse and anomaly detection[4]. It is used to increase the detection rate of known intrusions and to reduce the false positive rate of unknown attacks. Most ML / DL methods are hybrids.

I.LITERATURE SURVEY

[1]D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook, The Traveling Salesman Problem: A Computational Study. Princeton, NJ, USA: Princeton Univ. Press, 2017. The TSP-DS is one variation of the TSP and the vehicle routing problem. A recent review of the TSP and a review of multiple TSP (MTSP) problems was written by Bektasto use drones for making last mile delivery deliveries. This paper addresses the delivery concept of a truck-drone combination along with the idea of allowing autonomous drones to fly from delivery trucks, make deliveries, and fly to any available delivery truck nearby.

[2] B. L. Golden, S. Raghavan, and E. A. Wasil, The Vehicle Routing Problem: Latest Advances and New Challenges, vol. 43. New York, NY, USA: Springer, 2016.Available: Other excellent overviews of the vehicle routing problem were provided In contrast, research on the operational aspects of drone delivery has been neglected, and only a handful of papers in drone-truck systems have been presented. The plan for building the first airport for drones is ongoing in Rwanda [13]. In contrast, research on the operational aspects of drone delivery has been neglected, and only a handful of papers in drone-truck systems have been presented.

[3] P. Toth and D. Vigo, Vehicle Routing: Problems, Methods, and Applications. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2014.E-commerce and retail companies are seeking ways to cut delivery times and costs by exploring opportunities to use drones for making last mile delivery deliveries. This paper addresses the delivery concept of a truck-drone combination along with the idea of allowing autonomous drones to fly from delivery trucks, make deliveries, and fly to any available delivery truck nearby. We present a mixed integer programming (MIP) formulation that captures this scenario with the objective of minimizing the arrival time of both trucks and drones at the depot after completing the deliveries. Armed drones are generally used for three types of tasks: close air support (giving support to troops on the ground by firing from the air), elimination of specific targets, and continuous surveillance of a specific area to allow suspected objects to be attacked immediately.

[4] A. Allah Verdi, C. T. Ng, T. C. E. Cheng, and M. Y. Kovalyov, “A survey of scheduling problems with setup times or costs,” Eur. J. Oper. Res., vol. 187, no. 3, pp. 985–1032, Jun. 2013.It move through the air and thus avoid the traffic congestion problems of ground transportation. These advantages lead to the highly energy-efficient use of drones.

[5] R. Ruiz and J. A. Vázquez-Rodríguez, “The hybrid flow shop scheduling problem,” Eur. J. Oper. Res., vol. 205, no. 1, pp. 1–18, Aug. 2013. A drone station can operate drones after a truck arrives and supplies parcels. This characteristic is closely related to the PMS with precedence constraints. Tanaka and Sato [2] studied a single machine scheduling problem with precedence constraints. The objective was to minimize total job completion time, and job idle time was not permitted. A successive sublimation dynamic programmingmethod was applied to find the exact solution

III. EXISTING SYSTEM

With respect to truck-drone systems, researchers have not given sufficient attention to drone facilities because of the limited drone flight range around a distribution center.

.

IV.PROPOSED SYSTEM

Proposing a truck-drone system to overcome the flight-range limitation. We define a drone station as the facility where drones and charging devices are stored, usually far away from the package distribution center. The traveling salesman problem with a drone station (TSP-DS) is developed based on mixed integer programming. Fundamental features of the TSP-DS are analyzed and route distortion is defined. We show that the model can be divided into independent traveling salesman and parallel identical machine scheduling problems for which we derive two solution approaches.

V.TECHNICAL BACKGROUND

A.NETWORK SECURITY DATA SET

Data constitute the basis of computer network security research. The correct choice and reasonable use of data are the prerequisites for conducting relevant security research. The size of the dataset also affects the training effects of the ML and DL models. Computer network security data can usually be obtained in two ways: 1) directly and 2) using an existing public dataset. Direct access is the use of various means of direct collection of the required cyber data, such as through Win Dump or Wire shark software tools to capture network packets.

B.DARPA INTRUSION DETECTION DATA SETS:

DARPA Intrusion Detection Data Sets [2], which are under the direction of DARPA and AFRL / SNHS, are collected and published by The Cyber Systems and Technology Group (formerly the DARPA Intrusion Detection Evaluation Group) of MIT Lincoln Laboratory for evaluating computer network intrusion detection systems. The first standard dataset provides a large amount of background traffic data and attack data. It can be downloaded directly from the website. Currently, the dataset primarily includes the following three data subsets: 1998 DARPA Intrusion Detection Assessment Dataset: Includes 7 weeks of training data and 2 weeks of test data. ·1999 DARPA Intrusion Detection Assessment Dataset: Includes 3 weeks of training data and 2 weeks of test data.

B. KDD Cup 99 dataset

The KDD Cup 99 dataset [6] is one of the most widely used training sets; it is based on the DARPA 1998 dataset. This dataset contains 4 900 000 replicated attacks on record. There is one type of the normal type with the identity of normal and 22 attack types, which are divided into five major categories:

R2L (Root to Local attacks), U2R (User to Root attack), Probe (Probing attacks) and Normal. For each record, the KDD Cup 99 training dataset contains 41 fixed feature attributes and a class identifier. Of the 41 fixed feature attributes, seven characteristic properties are the symbolic type; the others are continuous.

C. NSL-KDD dataset

The NSL-KDD dataset [5] is a new version of the KDD Cup 99 dataset. The NSL-KDD dataset improves some of the limitations of the KDD Cup 99 dataset. The KDD 1999 Cup Dataset Intrusion Detection Dataset was applied to the 3rd International Knowledge Discovery and Data Mining Tools Contest. This model identifies features between intrusive and normal connections for building network intrusion detectors. In the NSL-KDD dataset, each instance has the characteristics of a type of network data. It contains 22 different attack types grouped into 4 major attack types.

D. ADFA dataset

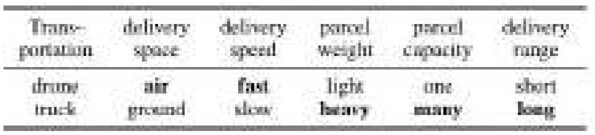
The ADFA data set is a set of data sets of host level intrusion detection system issued by the Australian defence academy (ADFA) [2], which is widely used in the testing of intrusion detection products. In the dataset, various system calls have been characterized and marked for the type of attack. The data set includes two OS platforms, Linux (ADFA-LD) and Windows (ADFA-WD), which record the order of system calls. In the case of ADFA-LD, it records the invocation of operating system for a period of time.

Kernel provides the user space program and the kernel space interact with a set of standard interface, the interface to the user program can be restricted access hardware devices, such as the application of system resources, operating equipment, speaking, reading and writing, to create a new process, etc. User space requests, kernel space is responsible for execution, and these interfaces are the bridge between user space and kernel space. ADFA-LD is marked for the attack type, as shown in the figure.

Linux system, user space by making system calls to kernel space to produce soft interrupts, so that the program into the kernel state, perform corresponding operations. There is a corresponding system call number for each system call.

It contains 5 different attack types and 2 normal types, as shown in Table

COMPARISON OF TRUCKS AND DRONES

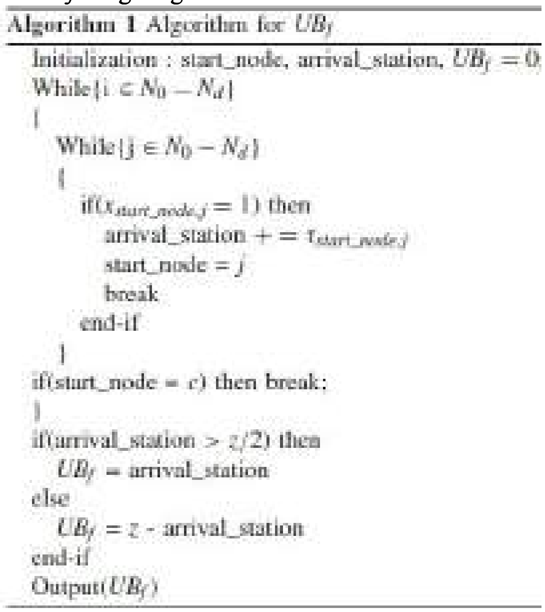


Trans-portation	delivery space	delivery speed	parcel weight	parcel capacity	delivery range
drone truck	air ground	fast slow	light heavy	one many	short long

E.TSP Technique:

The TSP-DS is an NP-hard problem, and the typical mathematical formulation can be solved very limited size of instances; it was hard to solve problems with more than 11 nodes of instances. One of our motivation is to reduce the complexity. By analyzing the mathematical structure of the TSP-DS, we found that there are special characteristics of the mathematical formulation and exploited them to derive decomposition methods which guarantee optimal solutions. For our problem, we address the situation in which the majority of customers are located far from the distribution center and the maximum flight distance of a drone from

the distribution center is less than the distance between the drone station and the distribution center.symmetric coupled random feedback binary unit neural network composed of a visible layer and a plurality of hidden layers. The network node is divided into a visible unit and a hidden unit, and the visible unit and the hidden unit are used to express a random network and a random environment. The learning model expresses the correlation between units by weighting.



Algorithm 1 Algorithm for UB_f

```
Initialization : start_node, arrival_station, $UB_f = 0$
While{i $\in N_0 - N_d$}
{
  While{j $\in N_0 - N_d$}
  {
    if($x_{start\_node,j} = 1$) then
      arrival_station $+= \tau_{start\_node,j}$
      start_node = $j$
      break
    end-if
  }
if(start_node = $c$) then break;
}
if(arrival_station > $z/2$) then
  $UB_f$ = arrival_station
else
  $UB_f$ = $z$ - arrival_station
end-if
Output($UB_f$)
```

In the study, Ding and Yuxin apply Deep Belief Nets (DBNs) to detect malware. They use PE files from the internet as samples. DBNs use unsupervised learning to discover multiple layers of features that are then used in a feed-forward neural network and fine-tuned to optimize discrimination. The unsupervised pre-training algorithm makes DBNs less prone to overfitting than feedforward neural networks initialized with random weights. It also makes it easier to train neural networks with many hidden layers.

E.TSP Technique:

The TSP-DS is an NP-hard problem, and the typical mathematical formulation can be solved very limited size of instances; it was hard to solve problems with more than 11 nodes of instances. One of our motivation is to reduce the complexity. By analyzing the mathematical structure of the TSP-DS, we found that there are special characteristics of the mathematical formulation and exploited them to derive decomposition methods which guarantee optimal solutions. For our problem, we address the situation in which the majority of customers are located far from the distribution center and the maximum flight distance of a drone from

VI.CONCLUSION

We define a new drone and truck-drone TSP by exploring use of a drone station with three features: 1) it can utilize many drones; 2) it is located far away from the distribution center; and 3) it is activated for delivery after a truck arrives with parcels. The TSP-DS was formulated based on mixed integer programming and we analyzed characteristics of the TSP-DS. We proved that the mathematical model can be divided into two different mathematical models, and derived the TSPMSand the TSMPMS to give the exact solution of the TSP-DS. Computational experiments showed that the fundamental characteristics of the TSP-DS and the TSMPMS could effectively reduce the complexity problem. Another experiments revealed that the TSP-DS is more effective than the PDSTSP when a majority of customers are located far from the distribution
center. We also showed that route distortion can be eliminated with relatively small number of drones. We expect our model can be used as a means to overcome the limits of drone facility problems, and it can be used to establish drone-truck delivery
systems in the near future.

ACKNOWLEDGMENT

I am very thankful to the Staff members and Colleagues who have directly or indirectly contributed for completion of this Paper. I am very much grateful to the entire CSE department of Vivekananda Institute of Technology, Bangalore for giving me all facilities and work environment which enable me to complete my task. I am also thankful to all researchers which I have mentioned in reference list. I am also thankful to all my family members and friends for supporting to complete my paper.

REFERENCES

[1] M. Grothaus. (Jan. 2016). This is How Google's Project Wing Drone Delivery Service Could Work. Available:https://www.fastcompany.com/3055961/fast-feed/this-is-how-googlesproject.

[2] D. Muoio. (Aug. 2016). Google's Secretive Drone Delivery Project Just Got Cleared for Testing—Here'S Everything We Know About the Program.[Online].Available:http://www.techinsider.io/google-projectwing-drone-service-2016-8

[3] N. Agatz, P. Bouman, and M. Schmidt, "Optimization approachesfor the traveling salesman problem with drone," Transport. Sci., vol. 52, no. 4, pp. 965–981, Apr. 2018. [Online]. Available: https://doi.org/10.1287.

[4] M. Wohlsen. (Jun. 2014). The Next Big Thing You Missed: Amazon'sDelivery Drones Could Work—They Just Need Trucks. [Online].
Available: https://www.wired.com/2014/06/the-next-big-thing-you-missed-delivery-drones-launched-from-trucks-are-the-future-of-shipping

[5] M. Tanaka, K. Tanaka, and H. O. Wang, "Practical model construction and stable control of an unmanned aerial vehicle with a parafoiltype
wing," IEEE Trans. Syst., Man, Cybern., Syst., to be published,doi: 10.1109/TSMC.2017.2707393.

[6] S. Cai, Y. Huang, B. Ye, and C. Xu, "Dynamic illumination optical flow computing for sensing multiple mobile robots from a drone,"
IEEE Trans. Syst., Man, Cybern., Syst., vol. 48, no. 8, pp. 1370–1382,
Aug. 2017.

[7] A. Robin. (Jan. 2015). Pylons Dronairports Would Serve As Charging Stations and Safe Storage.

Health Tweet: Reward Based Transactions using Element Crypto Currency

Anushree J[1], Pavithra V Bhat[2,], Priyanka K[3], Priyanka Krishna[4], Mrs. Usha Rani J[5]
Student [1,2,3,4] & Asst. Professor[5], Dept. of CSE/GSSSIETW, Mysuru, Karnataka, India
(E-mail: pavi.bhat130197@gmail.com)

Abstract- Blockchain is an evolving technology and it is a public ledger to which everyone has access but without a central authority having control. One of the best known applications of blockchains are the cryptographic currencies such as 'Bitcoin' and others, but many other applications are possible. It offers a secure way to exchange any kind of good, service, or transaction. Many implementations of blockchain technology are widely available today, each having its particular strength for a specific application domain. Our project uses this technology in order to provide a platform for all the like-minded people to share their thoughts and gain rewards.

Keywords*: Element crypto-currency, Bitcoin, X11 Algorithm, Ethereum, GSSS coin.*

I. INTRODUCTION

This paper is an effort to break the ground for presenting and demonstrating the use of Blockchain technology in multiple industrial applications. Blockchain technology is considered to be the driving force of the next fundamental revolution in information technology.

One such in the healthcare industry application called **"Health Tweet",** is formalized and developed on the foundation of Blockchain initiative. The application will enable, the people to interact with others by posting their perspectives and showing attentiveness related to health and fitness, and will be rewarded when the eligibility criteria is reached.

The application uses Block chain algorithm to build a very own crypto currency called **"GSSS COIN",** which is used for distributing rewards. This application helps in providing the health (both physical and mental) and fitness related realizations among the users who sign in, and also acts as a virtual community for the affinity group to communicate with each other.

II. EXISTING SYSTEM

There are many web applications similar to the proposed system like Twitter [1], but they lack in concentrating only on Health and fitness related views and concerns, apart from this there is no application which provides health related products and services for its users as rewards. From paper [1], we gained the knowledge of the current social media technology, APIs, documentation and developer tools. Some features like wall, tweet, post, like and follow has been adopted in our system for better implementation of interaction and community like appearance. Furthermore, we gained knowledge on data storage and data structure of the social networking site which helped us design our storage of data. Collecting user information and network information has also been in consideration for better data storage and display. Major visualization methods have been improved, mainly text and network display. It has been a source for decision making and intelligence gathering.

No application present in market which uses its own crypto-currency for claiming the rewards. The systems in existence leaves out the integrating part of health, motivation and availability of product related to health, all in one platform. Users are being deprived of using more than one type of crypto currency to transact is the current scenario.

III. PROPOSED SYSTEM

This proposed system aims at a platform where the health & fitness related views and concerns will be posted. The users of this application will gain the knowledge shared by others. The main objective of this application is to provide health related awareness among the users of this application. Here are some more features about this applications:

a) This application creates a virtual community for likeminded people with respect to health and fitness to communicate, improve and inspire those around them.

b) To achieve reward based system using crypto currency.
c) Creation of own currency called “GSSS Coin”.
d) This website aims at bringing total health of a being to finger tips.

IV. METHODOLOGY

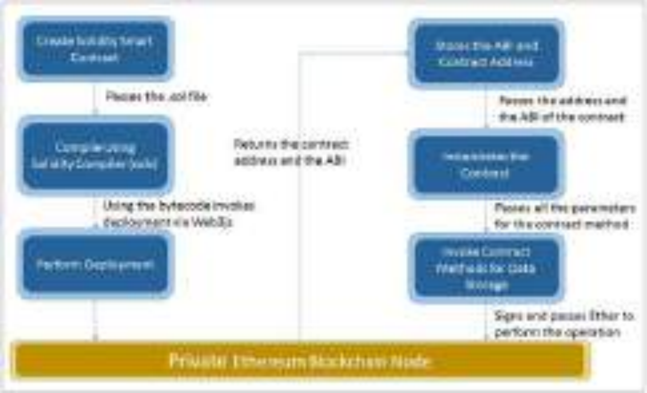

Create Solidity Smart Contract
Passes the .sol file
Compile using Solidity Compiler (solc)
Using the bytecode invokes deployment via Web3js
Perform Deployment
Returns the contract address and the ABI
Stores the ABI and Contract Address
Passes the address and the ABI of the contract
Instantiate the Contract
Passes all the parameters for the contract method
Invoke Contract Methods for Data Storage
Signs and passes Ether to perform the operation
Private Ethereum Blockchain Node

Fig 1 : Work Flow diagram of Private Block-chain Network using Solidity.

When the rewards points are to be given, we concentrate on making it the most randomly generated token and distribute using Ethereum. In the above seen diagram, the private blockchain node is explained. The code and customization is done in a solidity file and sent for execution, the .sol file is broken into ether byte-codes and then deployed into the ethereum dedicated node. This creates a contract for the sender and the receiver for accountancy of the transaction. This contract is stored for further references and records. Before the final step of transaction the contract is passed for digital signs of the party participating and then performs the operation of addition or deduction within the node.

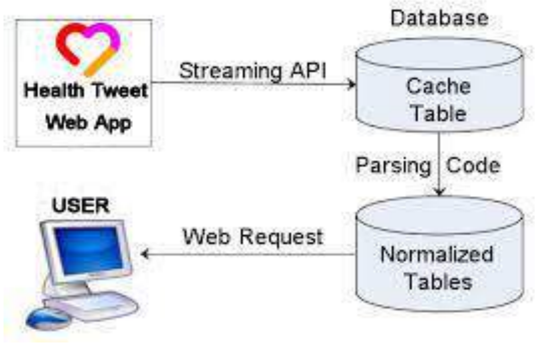


Fig 2 : Work Flow Diagram of Web application.

Health tweet web application architecture is well defined as shown in the above figure fig no[2]. The APIs and plug-ins fetch data from the cache repository and processes it and inserts the obtained data into page and tables. This normalized table has the requested data and is sent to the user for computation or as an output.

Overview of how the whole system works

Any interested user can create an account by providing valid credentials. Users can share their thoughts, tips, videos related to health. It may be about physical or mental health. By putting out interesting, unique and beneficial posts they can earn more number of likes and followers.

By reaching the specified milestones they earn rewards in the form of GSSS coins .GSSS coins are the crypto currency which we have developed using solidity. Users have a shop page in which they can put health related products into the cart. It may be products, pdf, vouchers all related to health. They can complete the transaction by paying for the products using BTC or GSSS coins.

This way our platform helps people to get motivated to share health related thoughts and tips which will be beneficial for all the other users by providing rewards in the form of crypto currency which will yield them health related products. This is a cycle of giving out thoughts about health and receiving back health in the form of few products. Health tweet operates on the Ethereum blockchain using X11 Algorithm to deploy smart contracts for health -related services.

Few important aspects in the system:

GSSS Coin

The creation of our own crypto currency was done using solidity coding on the Ethereum platform. These coins are distributed as the rewards to the users on reaching the milestones, in a way motivating them to achieve the milestone. Users can use this crypto currency to make a purchase. This enables a sense of own world with own currencies. As we stress on bringing like-minded crowd on a single stage, this serves as one of the main key feature we can achieve in building that environment.

X11 algorithm

A reward based system is technically a proof-of-work concept. This proof-of-work and rigging of the GSSSC happens using the x11 algorithm. We compared all the randomness generating algorithm, keeping security in mind and we ended up with this reliable algorithm. 11 chained hash functions of this algorithm provides the expected output. Dealing with currencies we have to be sure about security and encrypting methods, x11 algorithm provides us the solution for both.

Wall of health

A society is well defined only with existence of good communication and interaction. Through our application we give that space for the users to share thoughts, read inspiring stories and even share videos of best practice. The most important and beneficial fact about this feature is being able to connect with a doctor and follow their account for a routine check on one's self.

Video Stream

An interesting feature of our application is the video portal. Users can post, like and view the videos of their choice to contribute to their health practice. This method is found to be effectively motivating viewers into following routine in order to gain reward. Through this video stream concept we focus on proof-of-work. The reward based system in our approach deals with proof-of-work by each user. if the user completes the milestone of the task expected, only then the GSSSC will be awarded.

Products, books and vouchers

As an application we are not confined only to being the provider of platform for likeminded users, but we also provide a portal exclusive for purchase of products available on the portal, request for access for the PDFs based on their interest and services like delivery of the product is also an exciting add on to our application.

V. CONCLUSION AND FUTURE WORK

The system provides a platform for the like-minded people in order to share thoughts, tips through which they can earn likes can be earned, followers in turn it yields them the reward points. Using the reward point, purchase of health related products can happen. Usage of the block chain in social media networking is one that is likely to experience the ramifications of the disruption. The capability to earn small amounts of crypto currencies for the rightful behavior will pull the users towards the platform that promotes health where increase in contributions to the platform would get a payback. Time spent on social networking sites and each post made will earn a tiny amount of Bitcoin which would in turn promote the health sector.

In the future the implemented GSSS coin can be deployed for some value using which transactions in the digital world can be made. Also Mobile app can be developed for better usage of the system.

REFERENCES

[1] Wiesław Wolny, "Knowledge Gained from Twitter Data", 2016 Federated Conference on Computer Science and Information Systems (FedCSIS)

[2] Mauro Isaja, John Soldatos , "Distributed Ledger Technology for Decentralization of Manufacturing Processes", IEEE 2018.

[3] Harry Halpin, Marta Piekarska, "Introduction to Security and Privacy on the Blockchain", 2017 IEEE European Symposium on Security and Privacy Workshops

[4] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends", 2017 IEEE 6th International Congress on Big Data.

[5] Massimo Di Pierro, "The Blockchain", IEEE 2017.

[6] Jesse Yli-Huumo, Deokyoon Ko, Sujin Choi, Sooyong Park, Kari Smolander., "Where Is Current Research on Blockchain Technology? —A Systematic Review", October 3, 2016 Yli-Huumo et al.

[7] Sachchidanand Singh, Nirmala Singh., "Blockchain: Future of Financial and Cyber Security", 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I).

[8] Morgen, Peck., "Reinforcing the Links of the Blockchain.", November 2017 IEEE.

[9] Da-Wei Liu., "Study on the Factors of Customer's Loyalty in E-Business World.", 2007 International Conference on Wireless Communications, Networking and Mobile Computing.

[10] Michael T. Capizzi, Rick Ferguson., "Loyalty trends for the twenty-first century", Emerald Group Publishing Limited 2005.

[11] Ting Chen ∗ †, Xiaoqi Li†, Xiapu Luo†‡, Xiaosong Zhang ∗ , "Under-Optimized Smart Contracts Devour Your Money" 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER).

[12] Pinyaphat Tasatanattakool, Chian Techapanupreeda, "Blockchain: Challenges and Applications", 2018 International Conference on Information Networking (ICOIN).

[13] Tareq Ahram, Arman Sargolzaei, Saman Sargolzaei, Jeff Daniels, and Ben Amaba, "Blockchain Technology

Innovations", 2017 IEEE Technology & Engineering Management Conference(TEMSCON).

[14] Tomaso Aste and Paolo Tasca, Tiziana Di Matteo, "Blockchain Technologies: The Foreseeable Impact on Society and Industry", Computer (Volume: 50 , Issue: 9 , 2017) IEEE Computer Society .

[15] Ryan Henry , Amir Herzberg and Aniket Kate. "Blockchain Access Privacy: Challenges and Directions", IEEE Security & Privacy (Volume: 16 , Issue: 4 , July/August 2018).

[16] Stefan K. Johansen "A comprehensive literature review on the Blockchain as a technological enabler for innovation" , Mannhiem University.

[17] Josep Lluis de la Rosa, Victor Torres-Padrosa, Andres el-Fakdi, Denisa Gibovic, Horynak, O, Lutz Maicher, and Miralles, "A survey of Blockvhain technologies for open innovation", TECNIO Centre EASY, Universitat de Girona.

[18] Supriya Thakur Aras and Vrushali Kulkarni, "Blockchain and Its Applications- Detailed Survey", International Journal of Computer Application (0975-8887).

[19] Stefan steebacher and Ronny Schuritz, "The Blockchain Technology as an Enabler of Service Systems", Karlsruhe Institute of Technology.

[20] Rishav Chatterjee, Rajdeep Chatterjee, "An Overview of the Emerging Technology: Blockchain" , 2017 International Conference on Computational Intelligence and Networks

Advanced Face Recognizition Based Door Locking System Using Raspberry PI 3

Suresh Kumar [1], Ravindra k [2], Shridhar Duttaragi [3], NaveenKumar M K [4], Swathi S Prabhu [5]
Asst. professor[1] & 8 th Semester B.E Student[2, 3, 4, 5]
Computer Science Department, VKIT, Bengaluru,India
sureshkatte89@gmail.com [1], naveenkumarmk23@gmail.com [2], shridharsd4@gmail.com [3],
swathiprabhu97@gmail.com [4], ravindrakeshavamurthy18@gmail.com [5]

ABSTRACT

This paper deals with the idea of secure locking automation utilizing IOT for door unlocking system to provide essential security to our homes, bank lockers and related control operations and security caution through the GSM module .It uses an image capturing technique in an embedded system based on raspberry pi server system. RPi (Raspberry pi) controls the video camera for catching it for turning on a relay for door unlocking. The module contains a secured face recognizer for automatic door unlocking. The camera catches the facial picture and compares it with the image which is stored in the database .If the picture is found in the database then the door lock opens otherwise it will produce a SMS that an unknown person is trying to gain access and along with that we are implementing Home automation which will be very helpful for providing support to disabled people and fulfill their needs in home and thus they lead a normal life. This module controls the home appliances with a very ease of installation and it is user friendly.

Keywords

Facial Recognition; Internet of Things(IOT);Image matching; Sensor System; Digital camera; Raspberry Pi 3;Person Identification, Android, Wi-Fi module.

I. INTRODUCTION

In these modern times, home security is the need of the hour for the development of society as a whole which in turn will help make our cities smart, so the concept of facial recognition to gain access of the house is an idea which is used to make our place of living more secure. A facial recognition system is a system which captures facial images and verifies the identity of a person using a digital camera.

Face is detected automatically by using face detection technique and the entire face recognition is completed without touching with any hardware. Face detection is the first step of the face recognition system. The performance of the entire face recognition system is influenced by the reliability of the face detection. By using face detection, it can identify only the facial part of an image regardless of the background of this image.

A facial acknowledgment framework is a framework which gets facial pictures and confirms the character of a man using a propelled camera . It is an application fit for distinguishing or checking a man from a computerized picture. One approach to do this is by looking at chose facial components from the picture and a face database.

Home automation is become more beneficial because of its safety and security. Nowadays, home automation became more advance and precise to monitor all the home appliances. In this paper, we used a Wi-Fi wireless technology to monitor the device. An android application is installed in a mobile device i.e. android smart phone and it has inbuilt switch interface of all the appliances separately in it. Through which all the respective devices can be control and monitor individually.

The Wi-Fi module receives the command from mobile phone and passes to relay circuit. As per the given signal from the user, the relay circuit switched ON/OFF the respective devices. The main purpose of using Wi-Fi wireless technology is to provide a greater extent to range and better feasibility.

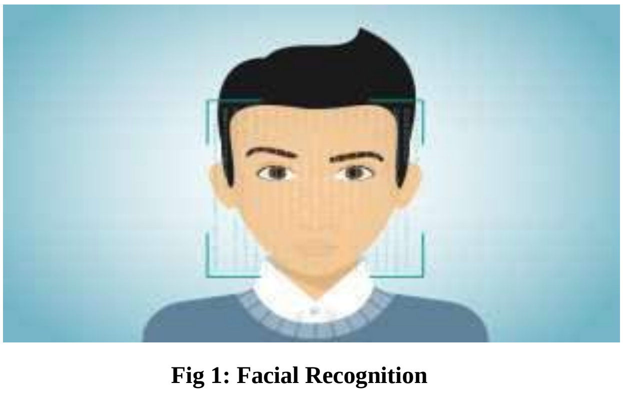


Fig 1: Facial Recognition

Internet of things: The internet of things, additionally called the internet of articles, refers to a remote system between items. The term IOT has come to portray various advances and research teach that empower the web to connect into this present reality of physical articles.

There are various technologies that enable IOT:

- RFID and near-field communication

- Optical tags and quick response codes
- Bluetooth low energy
- ZigBee

In this work we utilized raspberry pi 3 is a single board PC created in the United Kingdom by the raspberry pi establishment. Raspberry pi has many generations. What we are using here is pi3.It replaced pi2 model b in February 2016. Raspberry pi3 is believed to be approximately 80%faster than

RPi2 in parallelized task. . Price is in between US 20$ to 35$. It has architecture of ARM v8(64/32 bit), Broadcom BCM 2837 System on chip used along with the CPU of 1.2GHz 64/32 bit quad core ARM cortex A53. The memory of Raspberry pi 3 is 1GB and the storage is in the micro SDHC slot. It has an additional feature of Wi-Fi and Bluetooth as compared to other versions of Raspberry Pi.

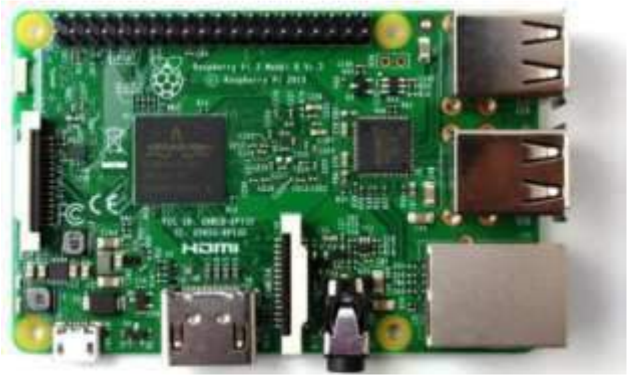


Fig 2: Raspberry Pi 3

IBM Blue mix is an open standard cloud based platform as a service (PaaS) developed by IBM. It supports several programming languages and services as well as integrated DevOps to build, run, deploy and manage all types of applications on the cloud. Bluemix depends on Cloud Foundry open technology and keeps running on Soft Layer infrastructure. Bluemix supports several programming languages including Java, Node.js, PHP, Swift, Python, Ruby, etc.

• **Public Cloud:** An open cloud is one in which the framework and other computational assets that it contains are made accessible to the overall population over the Web. It is possessed by a cloud supplier offering cloud administrations and, by definition, is outside to an association.

• **Private Cloud**: A private cloud a restrictive system or a server farm that provisions facilitated administrations to a predetermined number of individuals. It might be overseen either by the association or an outsider, and might be facilitated inside the association's server farm or outside of it.

• **Community Cloud**: A people group cloud is fairly like a private cloud; however the foundation and computational assets are shared by a few associations that have regular protection, security, and administrative contemplations, instead of for the restrictive utilization of a solitary association.

•**Hybrid Cloud**: A cross breed cloud is a piece of at least two mists (private, group, or open) that stay extraordinary elements yet are bound together by institutionalized or restrictive innovation that empowers interoperability.

Cloud computing uses three conveyance models by which diverse sorts of administrations are conveyed to the end client. The three conveyance models are the SaaS, PaaS and IaaS which give framework assets, application stage and programming as administrations to the purchaser.

- **Software-as-a-Service** - SAAS is characterized as a product appropriation display in which applications are facilitated by a seller or specialist co-op and made accessible to clients over a system. Otherwise called "on request" programming, it is the most develop kind of Cloud computing on account of its high adaptability, demonstrated bolster administrations, improved versatility, lessened client support, and diminished cost due to their multi-principle designs. It is a model of programming arrangement whereby at least one application and the computational assets to run them are given to use on request. Its primary intention is to decrease the aggregate cost of equipment and programming advancement, upkeep, and operations. Security arrangements are completed predominantly by the cloud supplier. The cloud supporter does not oversee or control the hidden cloud framework or individual applications, with the exception of inclination choices and constrained regulatory application settings.
- **Platform-as-a-Service** - PAAS gives foundation on which programming engineers can fabricate new applications or amplify existing applications without requiring the need to (buy advancement, QA, or generation server framework. It is a model of programming arrangement where the figuring stage is given as an on-request benefit whereupon applications can be created and sent. Its primary reason for existing is to lessen the cost and unpredictability of purchasing, loading, and dealing with the fundamental equipment and programming segments of the stage, including any required program and database advancement instruments. The cloud supporter has control over applications and application environment settings of the stage. Security arrangements are part between the cloud supplier and the cloud endorser.
- **Infrastructure-as-a-Service** - Infrastructure-as-a-Service (IaaS) is a model of programming organization whereby the fundamental figuring foundation of servers, programming, and system gear is given as an on-request benefit whereupon a stage to create and execute applications can be built up. Its fundamental reason for existing is to abstain from buying, lodging, and dealing with the essential equipment and programming framework segments, and rather get those assets as virtualized items controllable by means of an administration interface. The cloud supporter for the most part has expansive flexibility to pick the working framework and improvement environment to be facilitated Security arrangements past the essential foundation are done predominantly by the cloud endorser.

The cloud stage is utilized as a part of this venture is IBM Blue mix. It is an open standard, cloud based stage for building, overseeing and running utilizations of different types (web, versatile, huge information, and new brilliant gadgets, so on).

A. HOME AUTOMATION:

Home Automation is a unique system that can control and establish communication between nearly all aspects of your house Home Automation is a term used to describe the working together of all household amenities and appliances. For example, a centrally microcontroller panel can have the capability to control everything from heating, air conditioning, security system, lighting and overall electrical appliances. Home automation can include controlling aspects of our home remotely through a computer or any mobile equipment, programming electronics devices to respond automatically to some conditions or scenarios or centralizing the control of a variety of appliances in our home into a single control center. For example, Control of lights in and around our house from one central location so there is no need to get out of to that place or go to downstairs if we forgot to turn OFF or ON any appliances just we can control

II. EXISTING SYSTEM

Most doors are controlled by persons with the use of keys, security cards, password or pattern to open the door. The aim of this paper is to help users for improvement of the door security of sensitive locations by using face detection and recognition. Face is a complex multidimensional structure and needs good computing techniques for detection and recognition. This paper is comprised mainly of three subsystems: namely face detection, face recognition and automatic door access control. Face detection is the process of detecting the region of face in an image. The face is detected by using the Haar Cascade method and face recognition is implemented by using the Haar Cascade. If a face is recognized, it is known, else it is unknown. The door will open automatically for the known person due to the command of the microcontroller. On the other hand, alarm will ring for the unknown person.

- In the present scenario the crimes are increasing exponentially , arising a need of security.
- Security can also be described as a condition so that one can develop and progress freely and with a faith that no harm may be done.
- Visually impaired people are more immune to such crimes.
- Camera is now enormously being used and with the development of its content that is used in various applications. One of such is visual surveillance that is in flagship demand in today's market.

III. LITERATURE REVIEW

In today's fast paced and ever changing world security is one of the basic needs of our lives. Use of technology in the field of security plays an important role in increasing the security as well as reducing the manpower efforts.

S. Suresh Kumar et al [1] The growth of the internet has given rise to a number of open online learning platform, enabling access to learning materials by millions of individuals regardless of age or education background collective intelligence. These massive open online courses have replaced traditional institutional learning environments, such as physical attendance of lectures, with globally accessible learning via the web.

S. Suresh Kumar et al [2] Smart education with cutting edge technology is trending in E-learning platform. The design of e-learning system exist with lot of challenges and need for a strong, flexible and robust system which is available to user. Database-per-service is a per0formance oriented cloud based service model, provides flexible access to user anywhere-anytime without setting up any physical hardware or setting any software configuration.

Y. Januzaj. et al. [3] proposed real time access control for face recognition using, Raspberry pi instead of GSM services and relay. The limitation of the work was it couldn't control the background light situation and ambient light conditions.

H.Lwin.et al.[4] has proposed a door lock access system which consists of three subsystems: to be specific face recognition ,face detection, and automated door access control. Face recognition is actualized by using the PCA (Principal Component Analysis). The door will open itself for the known person in command of the microcontroller and caution will ring for the unknown person. Demerit of this system is input images are taken through a web camera continuously until the 'stop camera' button is pressed. Somebody is required at the location to check unauthorized person's images or status of the system and take further appropriate action. Personal computer (PC) is associated with the microcontroller, The entire system will not work if PC is crashed or Non-Function..

M. Chowdhury.et al. [5] had implemented security system where if any person came at the door it was notified to the home owner via e-mail and twitter then the user could see the person standing at the door using camera from remote location. The image of the person got captured and sent to twitter and e-mail. They stated that user couldn't control the door remotely. They had concluded that this system was useful for preventing unauthorized access. The limitation of this work was that the alert generated was sent to the mail and twitter account but if the user didn't have internet connection on his/her phone, he/she couldn't check the mail and couldn't recognise that any unauthorized person was trying to access the door.

G.senthilkumar.et.al. [6] proposed a work on Embedded Image Capturing System Using Raspberry Pi. In this work, they captured the image and compared it with the database but the limitation was the system couldn't work properly in the ambient light condition.

M. Carikci et al. [7] proposed a work on A Face Recognition System based on Eigen face method in which they used Eigen method for face recognition and Euclidean distance method to compare the image of the person concerned with the images in the database. It was very efficient and fast method and also gave high accuracy.

S. Jogdand.et.al [8] proposed a work on Implementation of Automated Door Accessing System with Face Design and Recognition in which they used Viola Jones method for face detection and PCA (Principal Component Analysis) for the comparison of images. The limitation of this work was that it is not robust and the efficiency is less.

U. Sowmiya.et al [9]. Developed to connect any door with internet. In this system user also implemented PIR sensor and camera. PIR sensor used for detecting person and camera used for capturing the video of the person who comes at the door. The video was sent through 3g dongle to authorized person. They had also discussed some advantages of this system. They had concluded use of this system in banks, hospitals etc. But their proposed model didn't provide the facility of sending messages to the authorized people.

J. Kartik et al [10]. Have proposed two systems are proposed, one is based on GSM technology and other uses a web camera to detect the intruder. The first security system uses a web camera, installed in house premises, which is operated by software installed on the PC and it uses the Internet for communication. The camera identifies movement of any intruder before the camera measurements or camera range. The product imparts to the planned client through Internet arrange and in the meantime, it gives a sound alarm. The second security system is SMS based and utilizes GSM innovation to send the SMS to the owner.

IV. DESIGN

The proposed works are as follows:

- Interfacing of camera to capture live face images.
- Create a database of authorized person if they exist.
- Capturing current image, save it and compare with the database image.

- Interface GSM module to send alert to authorized person while unlocking the locked door in the form of SMS and CALL.
- The project can also be used for surveillance. For instance, it can capture the images of unidentified individuals and store it which can later be used to determine the impostors who tried to gain illegitimate access.
- Interface relay as an output.

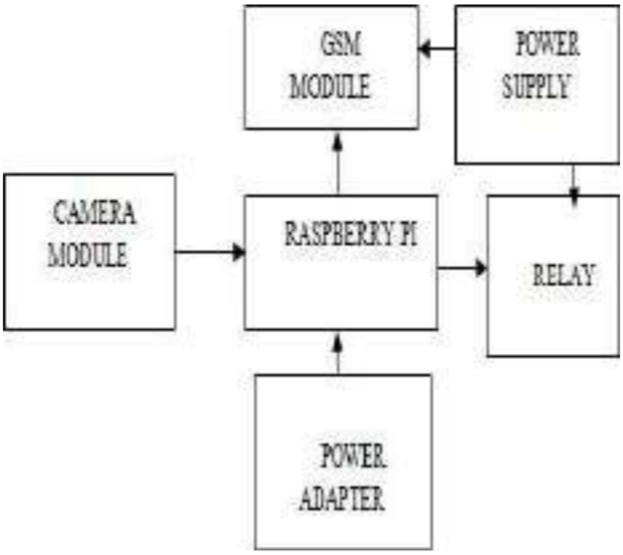


Fig 3: Block diagram of “Raspberry pi based face recognition system for door unlocking”.

The system will works in two different parts. The first part is for capturing and creating a database by storing the image. And the second one is to compare the image with the stored images in the database .For feature extraction we will use Eigen faces methodology and Euclidian distances will used for recognition of the face.

- **Camera module:** Camera module is pi camera interfacing to the raspberry pi module. It is used to capture images and send the clicked images to the raspberry pi module. Camera contains LEDs and flashes to handle that light condition that is not explicitly supplied by the environment and these light conditions are known as ambient light conditions.
- **Raspberry pi module:** raspberry pi 3 module is a small computer board. When an image is taken by raspberry pi it is compared with database image. For the first time when we capture an image to Create a database raspberry pi module captures many images to create a database in the system and this database is compared with the live captured images. After comparing the two images, based on whether the output is positive or negative it gives commands to GSM module.
- **GSM Module:** GSM module is used to send a message to the authorized people based on the output. If the output is positive “Information matched Access granted” message will be sent to the authorized people, otherwise in case of unauthorized access it will send an “Access denied. Some unknown person is trying to unlock the door”. Message to the certified users of the system.

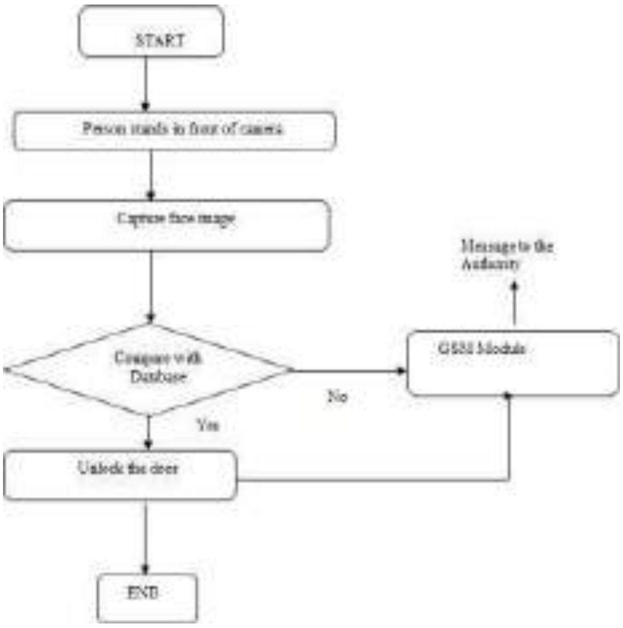


Fig 4: Flowchart of Image capturing and database comparison

V. METHODOLOGY

A. Proposed System Implementation

- Interfacing of camera to capture live face images
- Create a database of authorized person if they exist
- Capturing current image, save it and compare with the database image.
- Interface GSM module to send alert to authorized person while unlocking the locked door in the form of SMS and CALL
- The project can also be used for surveillance. For instance, it can capture the images of unidentified individuals and store it which can later be used to determine the impostors who tried to gain illegitimate access.
- Interface relay as on output

VI. RESULT ANALYSIS

The result in the creation of real time database are recorded. The real time database is created by using python. While executing it produces 30 images of each subject. . Likewise, databases should be created for at least 10 individuals and it creates each image size of about 100*130 pixels of height and width..along with will provide greater advantages like it decrease our energy costs, it improves home security. In addition, it is very convenient to use and will improve the comfort of our home.

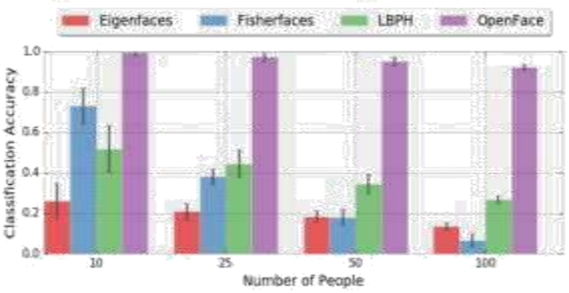


Figure 5 : Classification accuracy

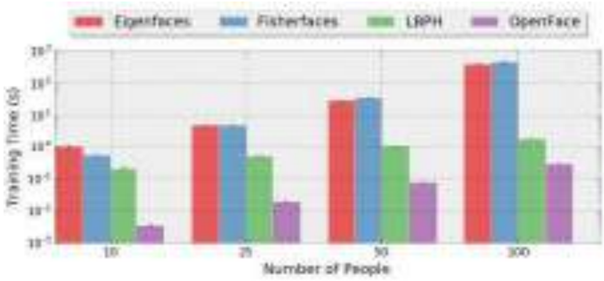


Figure 6 : Training times

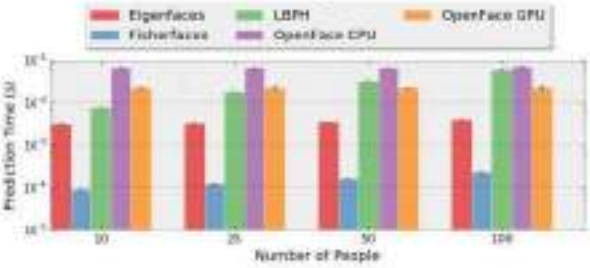


Figure 7: Per-image prediction time

VII. CONCLUSION

The arrangement of a facial recognition system using raspberry pi can make the system littler, lighter and work successfully utilizing lower control use, so it is more convenient than the pc- based face recognition system. It is open source software on Linux. Also, send a security alert message to the authorized person utilities. We are also providing power backup for the smooth and continuous functioning of the system in case of power failure. The power bank is used to charge the Raspberry Pi so there is less chance to slow down the system.

This development scheme is cheap, fast, and highly reliable and Raspberry pi takes less power and provides enough flexibility to suit the requirement of different people.

B. ACKNOWLEDGMENT

We would like to sincerely thank our project guide **Mr. S Suresh Kumar, Asst.Professor,** department of CSE**,** VKIT for his invaluable guidance, constant assistance, support and constructive suggestions for the betterment of the prefect, without which this Project Paper would have not been possible.

VIII. REFERENCES

[1] S. Suresh kumar, Dr Mallikarjuna Shastry P M, 2017, Database-pre-service for E-LEARNING System with Micro-Service Architecture.

[2] S.Suresh kumar, Dr Mallikarjuna Shastry P M, 2018, Analysis of Student Engagement and Cource Completion in Massive Open Online Courses.

[3] Januzaj, Y., Luna, A., Ramaj, V. 2015 Real time access control based on Facial Recognition.

[4] Lwin, H., Khaing, A., Tun, H. 2015.Automic door access system using face recognition.

[5] Chowdhury, M., Nooman, S. 2013. Access Control of Door and Home Security by Raspberry Pi through Internet.

[6] Senthikumar, G., Gopalkrishnan, K., Sathish Kumar, V. 2014 Embedded Image Capturing System Using Raspberry Pi System.

[7] Çarıkçı, M., , Özen, F. 2012 A Face Recognition System Based on Eigen faces Method.

[8] Jogdand, S., Karanjkar, M. 2015Implementation of
Automated Door Accessing System with Face Design and Recognition.

[9] Sowmiya, U., shafiq mansoor, J. 2015 Raspberry pi based home door security through 3g dongle.

[10] Kartik J. Srimadhavan V. 2013 SMS Alert and Embedded Network Video Monitoring Terminal.

[11] Sahani, M., Nanda, C., Sahu, A., Pattnaik, B. 2015Web Based Online Embedded Door Access Control and Home Security System Based on Face Recognition,.

[12] Mulla,,M., Patil, R. 2015 .Facial Image Based Security System using PCA.

[13] Gubbi, Jayavardhana,. (2013) Internet of Things (IOT): A vision, architectural elements, and future directions. Future Generation Computer Systems 29.7: 1645-1660.

[14] Raspberry Pi Foundation [Online] http://www.raspberrypi.org/downloads/.

Consciousness About Mobile App Permissions

Harish Gowda[1], Sanjana Gowda N C[2], Lakshmikanth Gowda M[3], Tanuja[4], Nikith G S[5]
Asst professor[4]
Computer Science Department
BGS Institution of Technology
BG Nagar,India
harishgowda8549@gmail.com
sanjanagowdanc@gmail.com[1], lakshmikanthgowdam@gmail.com2, tanuja212@gmail.com3, nikithniki555@gmail.com4

***Abstract*-In almost all the mobile apps after downloading it will ask some of the permissions to allow after clicking for the allow button only we can able to access any thing in that application otherwise we are unable to use all the features of that application. To be honest the apps only way to download the app is by allowing to "all" permissions since there is no option for partial permissions. Many popular android apps including Facebook messenger, WhatsApp, Skype, Twitter, Camera, Share it, Instagram get user permission after the installation. These permissions include recording with the phone audio and video at any time, calling contacts without additional permissions and modifying the USB storage contents. Lack of knowledge and awareness about permissions to the people may cause significant negative consequences. This research evaluates effectiveness of a demo app with visual ques to increase permissions awareness and avoid negative consequences. By using our new technique we can overcome from the mobile app permissions and some disadvantages.**

***Keywords*—Mobile applications, Permissions, Android, awareness, education, empirical study, Mobile app permission disadvantage.**

I. Introduction

Mobile apps are playing an increasingly important role in our daily lives. Permissions in Android apps is an either or proposition . Agree with the permission request or we cant able to use all the features of that particular application .There is no middle ground. Many popular Android apps including Facebook Messenger , WhatsApp, Skype, Twitter , Share it and Instagram get user permissions during installation .These permissions include recording with the phone audio and video at any time , calling contacts without additional permissions ,and modifying the USB storage contents.

EX : True caller mobile application will be working like this only by copying all the contacts information to their database once they had been allow the contact permission.

The 2018 smart phone market share shows Android at 74. 15%, Apple at 23. 28%,

Windows at 0.29%, KAIOS at 0.96%, Samsung at 0.29% and 0.42% for all others. Apple and Blackberry review permissions prior to store approval. Because of the significantly large Android market share and because of its –take-it-or-leave-it permissions structure focuses only on Android permissions .

Many apps request more permissions than needed by the app; free apps are infamous in requesting more than what they need for the app. Some apps are " Specifically designed to collect and expose users personal information to cyber criminals or other nation states"," UC Browser asks for fine location that is a precise point from where a person is searching for information .I understand that(ride hailing app) Uber or(food delivery app)SWIGGY needs to know because they offer a service .Why does UC Browser need to know the precise location for a search, "said NADKARNI .

Our smartphones have a lot of sensitive data including personal information, bank account information, and client-information. A cybercriminal or a nation state that can purchase user's sensitive data from an app provider installed on the smart phone can cause significant damage, and often without their knowledge.

Many people own cell phones .For example, in the USA 96% of the population own cell phones; the percentage is even higher(98%)among youth ages 18-29.A recent USA study found usage behaviour of smartphones includes 62% to lookup health information, 57% for online banking, 44% to find a place to live ,43% to look for a job ,40% for government services, 30% to take classes or educational content , and 18% to submit job applications .These usage statistics are reminders of how much sensitive personal data is involved in our routine activities.

Many of us buy insurance to protect our smart phones from theft or loss. But theft and loss are not the only way to loose data. Permissions falling to the wrong hands can cause data loss; in some case's permissions falling in the wrong hands is more damaging than a lost or stolen phone. Because data loss through permissions is often unbeknown to the owner unlike the physical device stolen or lost. Increasing awareness about the risk of permissions may help to reduces unintended consequences of sensitive data stored in our smart phones.

Various apps, as mentioned above are providing many facilities to people in free of cost as they are very useful for people directly. But indirectly these apps are gaining their personal information silently. They are using their personal information for their benefits.

We are providing an idea, to get rid of these data mining in our phones by a simple process.

We propose an awareness to people, not to get fooled by these kind of free apps which we are using in our daily life.

In this paper we study the research question: Can permission risk awareness reduce android app user's risk?

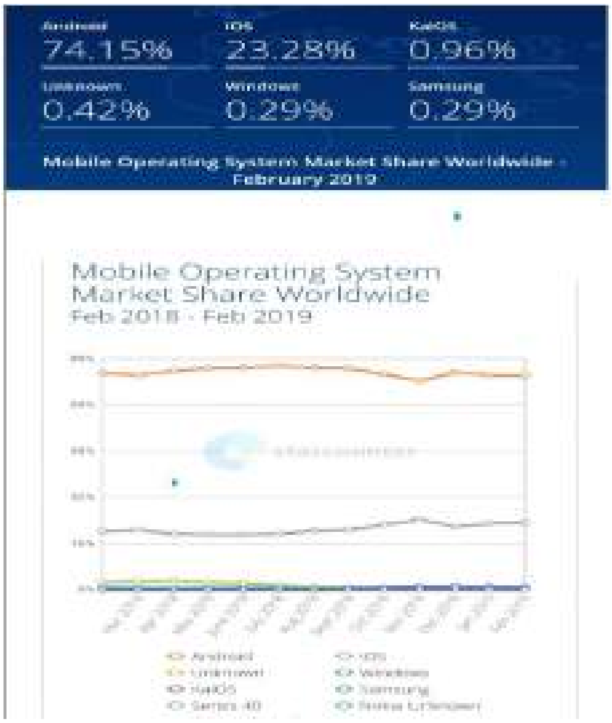


Fig 1: Mobile Operating System Market Share Worldwide Feb2018-Feb 2019

2 THEORETICAL BACKGROUND :

Prior research confirms that Android permission warnings are often ignored and do not help most users make correct security decisions. Most users are complacent in following security warnings. Some studies suggest awareness model to help organizations monitor security awareness positions while others recommend attention and behavioural changes. Safety risk awareness has been found to provide only temporary relief; further investigation is needed to see how this phenomenon apply to smart phones. Smart phone user characteristics were also found unique when applied to organizational mandates and desktop procedures. When smart phone security steps are mandated by the organization some users rather switch their phones off instead of complying with security steps; a study found that 39% of users were found to turn off their smart phone instead of doing additional security steps required by the organization .Un like computer users smart phone users are found to dislike re-authentication .

In inter organizational data exchange trust and risk were found to influence intention to use. When users trust the device or firm they tend to overlook warning signs. In our study we adopted three constructs from including risk, trusting beliefs and intention to use mobile apps.

While requesting the app permissions is totally on the discretion of the developer a multitude of permissions is usually requested may vary from generic permission like accessing your application information to more privacy invasive permissions like accessing the camera and personal information. The permission that is requested by most of the apps are shown in the fig 2.

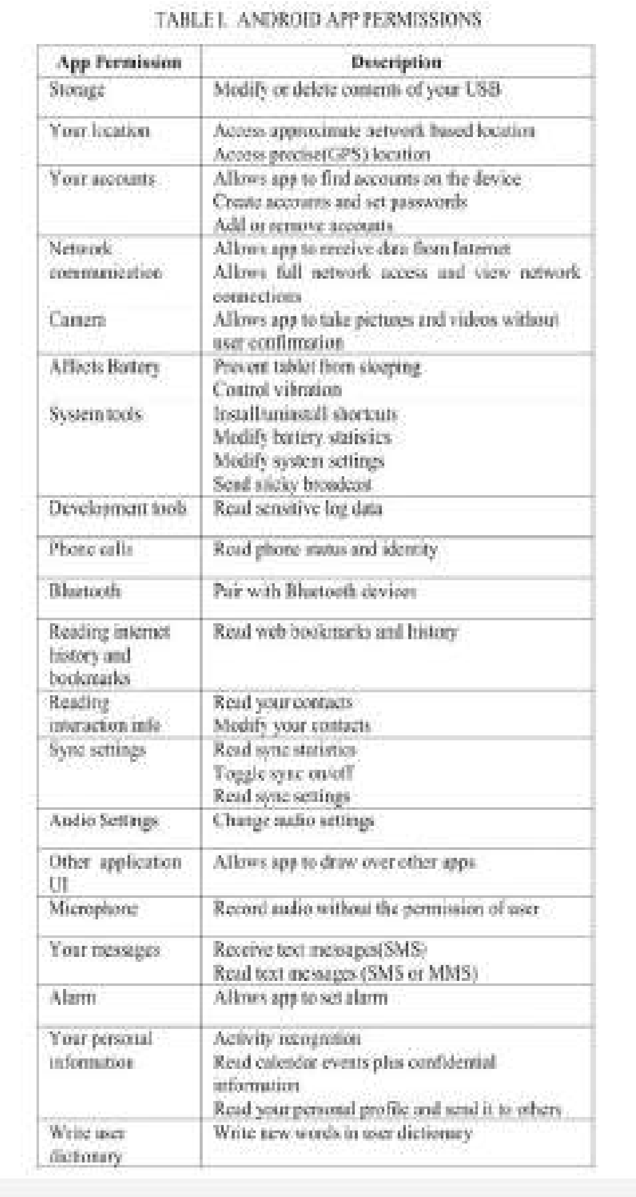


TABLE I. ANDROID APP PERMISSIONS

App Permission	Description
Storage	Modify or delete contents of your USB
Your location	Access approximate network based location Access precise(GPS) location
Your accounts	Allows app to find accounts on the device Create accounts and set passwords Add or remove accounts
Network communication	Allows app to receive data from Internet Allows full network access and view network connections
Camera	Allows app to take pictures and videos without user confirmation
Affects Battery	Prevent tablet from sleeping Control vibration
System tools	Install/uninstall shortcuts Modify battery statistics Modify system settings Send sticky broadcast
Development tools	Read sensitive log data
Phone calls	Read phone status and identity
Bluetooth	Pair with Bluetooth devices
Reading internet history and bookmarks	Read web bookmarks and history
Reading interaction info	Read your contacts Modify your contacts
Sync settings	Read sync statistics Toggle sync on/off Read sync settings
Audio Settings	Change audio settings
Other application UI	Allows app to draw over other apps
Microphone	Record audio without the permission of user
Your messages	Receive text messages(SMS) Read text messages (SMS or MMS)
Alarm	Allows app to set alarm
Your personal information	Activity recognition Read calendar events plus confidential information Read your personal profile and send it to others
Write user dictionary	Write new words in user dictionary

Fig 2: Table Android App Permissions.

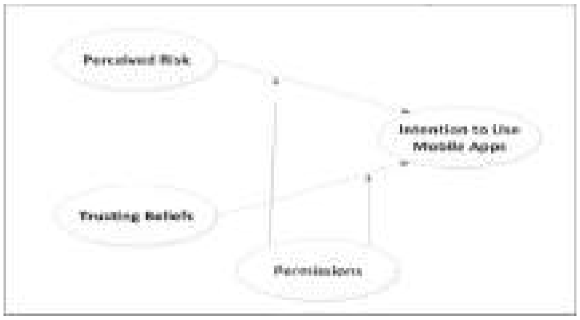


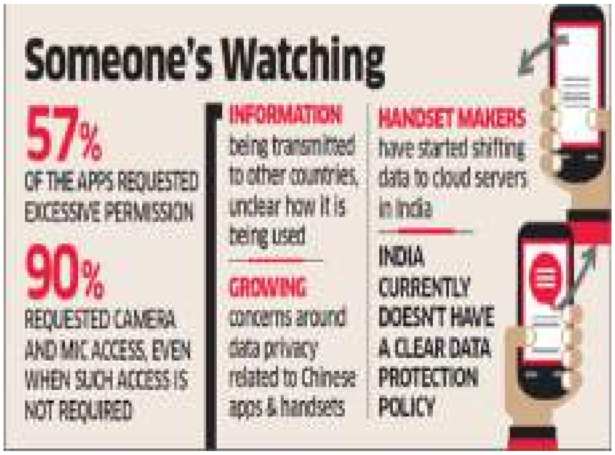
Fig 3: Theoretical framework for Intention to Use Mobile Apps.

This study focuses on the impact of risk and trusting belief on intention to use or install a mobile app. The theoretical model is shown in fig 3.

3.RESEARCH :

The study, commissioned by The Economic Times in the second week of January, reviewed the permissions sought and data shared by these apps among themselves or with third parties outside India. It also covered the various permissions sought by the apps to access features on user's phones such as contacts camera, microphone, sensors, location and text messages.

Given the proliferation of Chinese apps in India, the study focused specifically on the privacy aspects of mobile apps-the so called "Dangerous permissions" being taken by the apps and the data being shared with external parties. Social platform TIKTOK, and UC Browser owned by Chinese ecommerce giant Alibaba have hundreds of millions of user's accessing these apps every day. UC Browser has over 130million of its global 430 million users in India, according to the company.



Someone's Watching

57% OF THE APPS REQUESTED EXCESSIVE PERMISSION

90% REQUESTED CAMERA AND MIC ACCESS, EVEN WHEN SUCH ACCESS IS NOT REQUIRED

INFORMATION being transmitted to other countries, unclear how it is being used

GROWING concerns around data privacy related to Chinese apps & handsets

HANDSET MAKERS have started shifting data to cloud servers in India

INDIA CURRENTLY DOESN'T HAVE A CLEAR DATA PROTECTION POLICY

Fig 4: Survey regarding data protection.

The study found that on an average, these apps transfer data to around seven outside agencies, with69% of the data being transferred to the US. TIKTOK sends data to China Telecom; Vigo Video to TENCENT; Beauty Plus to MEITU; and QQ and UC Browser to its parent owned by Alibaba.

Xiaomi said it was moving local users' data to the cloud infrastructure of Amazon Web Services and Microsoft Azure located in India from servers in Singapore and the US. It will become the first major smartphone maker to initiate such a migration amid the ongoing debate on information security.

One Plus said that it will also relocate servers that house data of Indian consumers to India.

Vivo said to allay consumer fears on data privacy and security on their phones, it would be willing to comply with any regulations set by the government. The company, however, said it doesn't deal with consumer data and has no intent to monetize it, like apps do.

There is no privacy law in India today whereas in the US, there is some legal requirement and in Europe, (there) is the stringent GDPR REGIM.

4. EXPERIMENTS CONDUCTED :

A RANDOM sample of 1021 Android apps was taken into consideration. During the study, various categories of Android apps like Games, Tools. Entertainment Social& Communication, Music and Video, Personalization Productivity, Photography ,Education and eBook and life style were an ANALYZED . The dataset included the category of App and the lists of permissions that this app requested during installation. The apps have been classified in to two categories depending on the permissions requested. The first category includes those apps that request Generic Permissions and those that request Privacy invasive Permissions. Figure5 shows the percentage of apps that requested for generic permissions like the Audio settings, SYNC settings, wallpaper, reading internet history and soon. From Figure5, t can be seen that more than 95% of the apps request. Network Communication. 72% request Storage and 42 % requested Your app info permission.

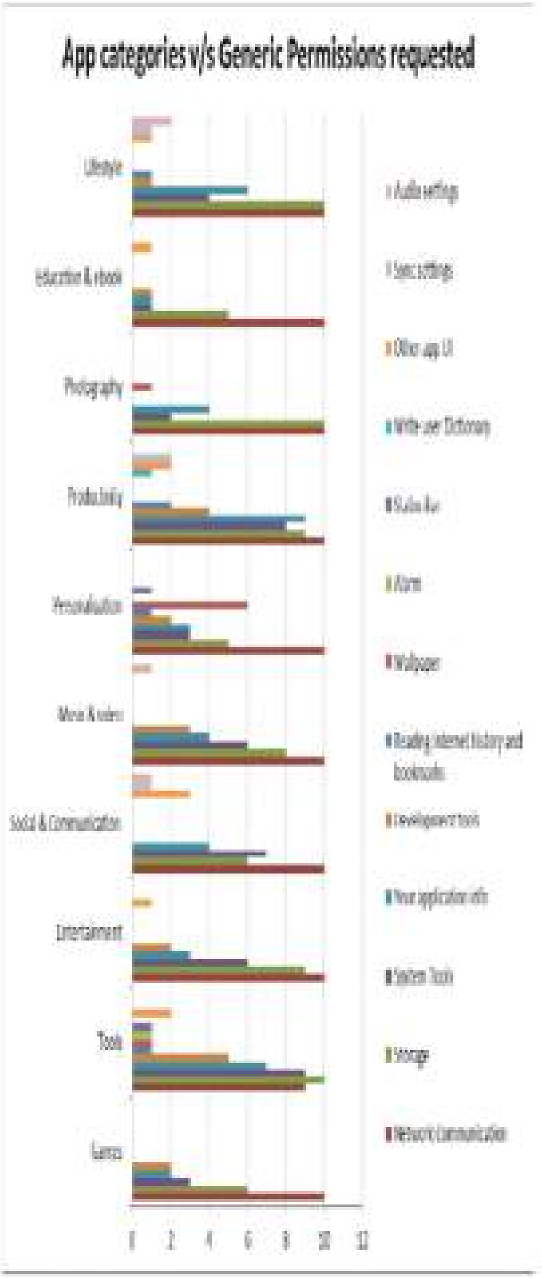


Fig 5: Percentage of apps requesting Generic permissions.

A similar study was conducted for the Apps requesting privacy invasive permissions like access to your account on the device, your messages, your personal information and even access to your camera, location and phone calls. As shown in Figure6, most of the apps requested access to the sensitive information as well .As indicated in the Figure, 48% of the apps requested access to the phone calls, 38% requested Your location. Thus there are several apps that request the Privacy invasive permissions.

It was observed that most of the apps demand large number of permissions from the user. On an average 71% apps were known to affect the battery, 99% access the Network communication, 78% affect the storage, 49% access the system tools, 43% affect the user's application info, 48% access phone calls, 38% access the user's location, 24% access Bluetooth, 25% access the account information of the user and 26% access the Camera.

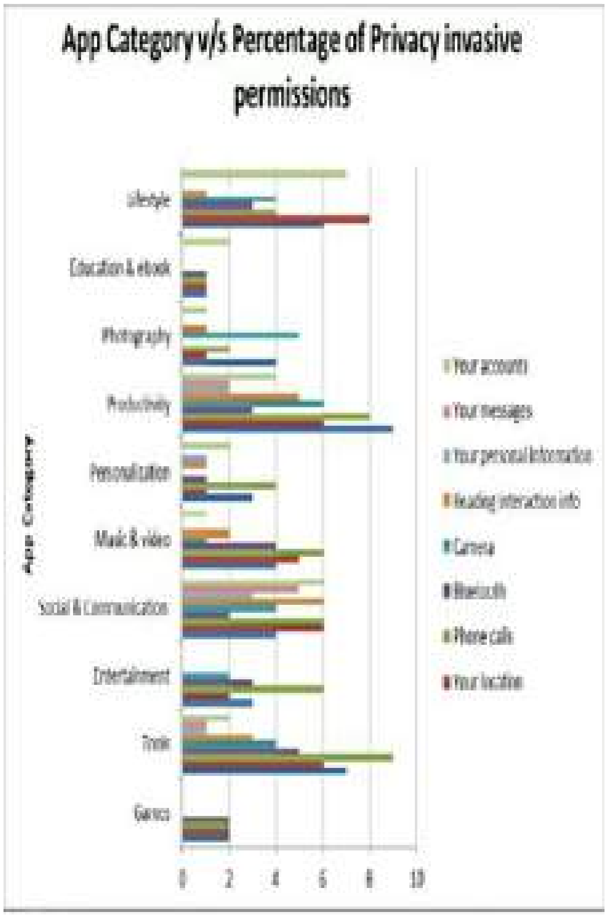


Fig 6: Percentage of apps requesting Privacy invasive permissions.

The good news is that some of the permissions are still requesting in lesser numbers! These include 20% for development tools, 19% for reading interaction info, 5% read internet history and bookmarks ,8% access wallpaper, 7% access personal information, 8% access messages and rest of the permissions requested are very less in numbers.

On the basis of GPPP model the user can calculate the extent to which the app is privacy invasive .And thus decide whether to install the app or not .For instance if an app requests 6 permissions that fall in to PP category and 4 permissions that fall in to GP category ,then the app is privacy invasive and the user must take a look at all the permissions requested before installing the app.

5. CONCLUSION :

Though most of the apps request a multitude of generic as well as privacy invasive permissions, a conscious decision on the part of the user is essential before installing the apps. The user should inadvertently read the permissions requested and their implications there of before granting access. This can help the user in preventing the revelation of personal information. The GPPP model provides a classification of the apps on the basis of the app permissions. This model can be used to aid the user in making a judicious decision whether to install a particular app or not. Before giving any permission to any of the application we should aware of the above disadvantages and there is an another way of using the mobile application without giving any permission by using another application called bouncer.

6.REFERENCE :

[1] Mobile Technology Fact Sheet, Pew Research Center, http: / / www. pcwintemet org/ fact sheets/ mobile technology fact sheet

[2] Why does this Android app need o many Permissions?, http: / / www. lifehacker.com/ 5991099/ why-does-this-app need-so-many permissions.

[3] Mobile operating system Wikipedia en. wikipedia. or g/ wiki / Mobile_ operating system Statista, The Statistics portal http: / / www. statista. com/ statistics 281106number-of-android app downloads-from-google-play

[4] The Economic Times

[5] IEEE Paper - How privacy invasive android apps are?

[6] IEEE Paper - Mining android apps to recommend permission

A Survey On Intelligent Answering System Using Similarity Model Based Learning

Dr. Yogish H K[1], Rachel Britto[2], Shwetha S[3],Deepak N[4]
Head of the Department, Professor[1],Student[2,3,4]
Computer Science & Engineering,
Sapthagiri College of Engineering,Bangalore, India
hodcse@sapthagiri.edu.in[1],brittorachel2697@gmail.com[2],
shwetha2017@gmail.com[3],deepaknsurya7@gmail.com[4]

ABSTRACT

Machine Learning (ML) is an artificial intelligence (AI) area that is a set of statistical techniques to solve problems. ML techniques can be applied to a wide range of unlimited problems-vision-based research, fraud detection, price prediction, and even natural language processing (NLP). A smart, user-friendly automatic response system is developed with the ability to detect and answer questions in English. There are many response systems that use the concept of natural language processing to answer the questions, but they are not so accurate in finding the right answer. Specific predefined queries with specific format are also required. Users need to ask the questions in the given format only for such systems. Users can enter the queries as they wish in the proposed system. No specific format is required. If the query fails to match any predefined query, the user will be suggested the best matched query. The domain expert keeps the answers to such questions in a database. The best matched response searched from the database is returned to the user during the retrieval of answers. A template matching technique is used to perform this match. Thus the system is more efficient and accurate from all the perspectives.

INTRODUCTION

In the modern world, the internet technology has become so much advanced that the way of interaction has changed across the entire application domain. Natural language is being used for communication in all the fields. When a user types a question on the internet, the user is flooded with relevant and irrelevant data and its time consuming to sort it out and find the information relevant to the question which triggers a need for Automated Intelligent Question Answering System for navigating to meaningful data. This is a challengingtask which can be achieved by using Natural Language Processing (NLP) and Information Retrieval (IR) techniques. NLP helps a computer to better understand the language used by user to communicate.

The skills required to build a smart response system include tokenization, parsing, speech tagging parts, question classification, query construction, sentence understanding, document retrieval, keyword ranking, classification, response extraction and validation. There are two important domains in which the QA system can be implemented: Open domain and closed domain. Open domain deals with questions in all domains. Closed domain deals with questions only in a particular domain and a database is maintained for it.

NLP is a hard problem which deals with understanding the human language not only words but also how the words team up to for a meaning. The different NLP techniques are: Deep analytics, Machine translation, Named entity extraction, Co-reference resolution, Automatic summarization, Sentiment analysis. IR is used to search and retrieve knowledge based information from the database.

LITERATURE SURVEY

In [1] a survey is conducted on various types of system for answering questions. In the field of Information Retrieval (IR), the questioning answering system is an important research area. Different fields of research are combined in this system like Natural Language Processing (NLP), Artificial Intelligence (AI), Information Retrieval (IR), and Information Extraction (IE). It is mainly categorized into two types based on the availability of resources: open domain and closed domain. The search engines like Google comes under the open domain where the information is retrieved from the World Wide Web. This is usually done by using keywords matching and frequencies of accessing documents. It is a time consuming process for the user to search for the relevant answer from the whole bunch of documents/data obtained as a search result. The QA is a rapidly growing research field. It reduces the time for extraction of data as well provides an exact or nearest to exact answer. This time is also dependent on the domain on which the system is used. The system used on closed domain fetches more accurate result when compared to system used under open domain. The search space for the system is less in closed domain when compared to the internet. The QA system has been implemented for different languages like Korean, Japanese, and English etc. The applications of the QA system are extracting information from document, online examination system, document management, Language learning, human and computer interaction, classification of document and many more.

In [2] an Intelligent QA system is designed for Arabic language. The main aim of this system was that it must be able to respond to the queries automatically and the answer retrieved should be accurate with respect to the query. The system answers more sophisticated questions that require some sort of temporal inference. It builds its database with different forms of question-answer pairs entered by the user and answered by the system. This will result in increased speed and accuracy. This system makes use of Information Retrieval (IR), Information Extraction (IE), and Natural Language Processing techniques. The user enters a query in spoken language and gets the answer in a word or a sentence format. Search Engines usually lack in conceptual knowledge of world and depends on probability theories and bivalent logic. It does not provide precession in the extracted answer. The retrieved snippets are not good for big data analysis and web information. But the QA system will provide only the requested answer. This system will first look up the database for the answer if it's not available then it will opt for WWW and consequently the result will be entered on to the database. If a specific question is not found in database, the question will be analyzed by the module of question processing and the answer will be extracted from the web by the module of answer extraction. Usually QA system will work on the concept of matching keywords, but this does not address the issue of how to conduct extensive question analysis and understanding of natural language. This proposed QA is said to have some degree of semantic understanding and can generate answers autonomously.

In [3] a system is developed which takes fact database as input and has the capability to answer questions of wide range of complexities. It implements a machine which will convert the input into a computable format so as to retrieve an answer for the question asked. The fact database consists of input facts and questions. The implemented system is evaluated based on whether it could answer simple deduction questions which produces yes/no as the answer and also on its capability to relate facts and answer a complex question. The tasks used for this evaluation are Dynamic Memory Networks (DMN), which passes the bAbi tasks of Facebook and uses this model to evaluate the data set released for the ARISTO challenge of Allen Institute, which contains increasingly difficult science questions. The knowledge base for this system is the Facebook bAbi tasks. Developing a system which gives consistent result for both simple and complex questions is a challenging task. BAbi tasks consist of 5 cases. Case 1 requires only single facts to answer the question. Case 2 requires multiple facts to be considered in order to conclude on an answer. Case 3 decides the answer on the basis of keywords and also takes into account the position of the keyword in the question. Case 4 deals with the entire yes/no questions. Case 5 deals with counting cases like questions with "how many?" criteria. The system takes into account all the relevant facts necessary to conclude on the final answer. To implement the above methods the system should have the capability to understand the facts given Neural networks has shown reliable average performance in this field but the performance is lowered as complex tasks such as bAbi are considered. A neural network has made a good progress in the field of image and text classification. Recently, performance has increased for complex tasks due to the addition of memory and attention components to neural networks.

In [4] the major role of the Questioning Answering System is to simulate answer required by the user. People usually use browsers to access information of the web. Blogs and Forums are a source of dynamic information sharing. Due to the widespread use of smart phones and apps for a huge number of purposes and need of information in short span of time, a mobile based questioning answering system is developed which provides personal assistance in learning and also for providing information for the users stored on the computers. It is used as a surface to fulfill the requirement of the users with related content. it uses natural language for the communication purpose and displays the optimized result. The mobile QA system will be taking input in natural language and analyses it and match it with the information stored in knowledge base and display result. Amazon book reviews, 20newsgroup and Yahoo datasets are used to build the knowledge base. The answers are stored as content - specific clusters in the knowledge base and display the output as snippets. Use Sentiment Analysis to decrease the vocabulary gap between user query and the retrieved response. The results of the proposed interface are evaluated using standard metrics such as Precision, Recall, F1-Score, Inverse Precision and Inverse Recall to return the relevant response. The open domain system is capable of answering dynamic data irrespective of domain nature, but the closed domain uses datasets for retrieving the answer. The learner who is in a need of static data will use the closed domain QA system like tourism, medical health, historical data etc. Recently, many efforts have focused on question answering systems based on social media networks for getting the precise information.

In [5] a research is made on the Questioning Answering Systems and the techniques used for development. As the information and communication technology has increased very much, the necessity and importance for the question answering system has increased rapidly. Complex assessment techniques are necessary for its development. A system must first analyze the question to answer a question and search for one or more possible answers and present it in a user - friendly form. The systems usually use graphical methods for representing knowledge. The knowledge gained is converted to a model and then the model is searched for answers. The Nodes Of Knowledge (NOK) method is one such example. The QA system is capable of answering factual questions by referring to the collection of documents by combining Information Retrieval and Natural Language Processing Techniques. There are two prominent fields in research of the QA system: resource and evaluation. Kupiec-a system used trivial method for collecting questions and Internet encyclopedia as source for answer. A set of multiple choice questions makes the evaluation easy. The evaluation of questions with no answers and questions with many answers is a complex process which is still under research. The six criteria for assessment of a QA system are: relevance, correctness, conciseness, completeness, coherence and justification.

In [6] an intelligent answering system is proposed based on the research in the field of Artificial Intelligence (AI). It aims at delivering concise information which may contain the answer for the question asked. This system is a solution for reducing the time consumed in searching the relevant answers from the data bombarded on the users. AI's goal is to develop a system that has intelligent behavior i.e. it has the ability to solve different problems, learn and understand languages. This system consists of two primary parts: knowledge base and inference engine. Usually the knowledge base is made up of factual and experimental data and the inference engine is used to determine the answers. The system designed is for closed domain and aims at retrieving the exact answer for the query rather than bombarding the user with a set of documents as done by search engine. The closed domain QA system requires extensive use of natural language processing and a lot of research to be done on factual data, definition lists, and paragraphs and cross lingual questions. When a question is given only the relevant part of the knowledge base is searched for thus reducing the search space. Knowledge base is represented as a rule i.e. it contains an IF and a THEN condition. The systems with this kind of representations are called as rule based systems.

In [7] a system is developed which takes fact database as input and has the capability to answer questions of wide range of complexities. It implements a machine which will convert the input into a computable format so as to retrieve an answer for the question asked. The fact database consists of input facts and questions. The implemented system is evaluated based on whether it could answer simple deduction questions which produces yes/no as the answer and also on its capability to relate facts and answer a complex question. The tasks used for this evaluation are Dynamic Memory Networks (DMN), which passes the bAbi tasks of Facebook and uses this model to evaluate the data set released for the ARISTO challenge of Allen Institute, which contains increasingly difficult science questions. The knowledge base for this system is the Facebook bAbi tasks. Developing a system which gives consistent result for both simple and complex questions is a challenging task. BAbi tasks consist of 5 cases. Case 1 requires only single facts to answer the question. Case 2 requires multiple facts to be considered in order to conclude on an answer. Case 3 decides the answer on the basis of keywords and also takes into account the position of the keyword in the question. Case 4 deals with the entire yes/no questions. Case 5 deals with counting cases like questions with “how many?” criteria. The system takes into account all the relevant facts necessary to conclude on the final answer. To implement the above methods the system should have the capability to understand the facts given Neural networks has shown reliable average performance in this field but the performance is lowered as complex tasks such as bAbi are considered. A neural network has made a good progress in the field of image and text classification. Recently, performance has increased for complex tasks due to the addition of memory and attention components to neural networks.

CONCLUSION

This paper emphasizes on various question answering systems and question answering techniques. There are many question answering systems which return answers to the questions entered by the user but fail to return appropriate and accurate answers. It is a challenging task to develop a system which would understand natural language questions correctly and provide exact answers to the user. The quality of answer returned by closed domain question answering system is high because the system is restricted to a particular domain, whereas the quality of answer returned by open domain system is low. The performance of a system can be improved to return exact results by retrieving answers from knowledge base. The knowledge based system returns the specific answer to the user instead of a document.QA system will help all users to retrieve exact information easily.

REFERENCES

[1] “A survey on types of Question Answering system”A. Chandra Obula Reddy 1, Dr. K. Madhavi 2 1 Research Scholar, Department of Computer Science & Engineering, JNTUA, Ananthapuramu - 515002, A.P., India. 2 Associate Professor, Department of Computer Science & Engineering, JNT University, Ananthapuramu - 515002, A.P., India.IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 6, Ver. IV (Nov.-Dec. 2017), PP 19-23.

[2] Waheeb Ahmeda1, AjushaDasana, BabuAnto ” Developing an Intelligent Question Answering System” Kannur University, Kannur, Kerala 670567, India Received: 03 January 2017; Accepted: 28 March 2017; Published: 08 November 2017.

[3] VinayChandragiri “Intelligent Question Answering System” Department of Computer Science and Engineering, IIT Guwahati Guwahati, Assam, 781039, India ISSN:0975-9646. VinayChandragiri/ (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (5) , 2016, 2231-2234.

[4] J. Tomljanović, M. Pavlići M “Intelligent Question – Answering Systems: Review of research” AšenbrenerKatić Polytechnic of Rijeka Vukovarska 58, Rijeka, Croatia. I.J. Information Engineering and Electronic Business, 2018, 1, 16-23 Published Online January 2018 in MECS.

[5] KarpagamK ”A Mobile based Intelligent Question Answering System for Education Domain” Department of Computer Applications, Dr.Mahalingam College of Engineering and Technology, Pollachi, 642003, India. Saradha A Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, 638316, India. Received: 21 July 2017; Accepted: 22 September 2017; Published: 08 January 2018.

[6] V.S.Babanne1 Dr.S.T.Patil2, D.J.Joshi3 “Intellegent Question Answering System” 1 Research Scholar, Vishwakarma Institute of Technology, Pune, India.2 Professor, Vishwakarma Institute of Technology, Pune, India 3 Asst.Professor, Vishwakarma Institute of Technology, Pune, India. International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2016 199 ISSN 2229-5518.

[7] A. Roshdiand and A. Roohparvar, “Review: Information Retrieval Techniques and Applications” International Journal of Computer Networks and Communications Security, September 2015.

[8] ShubhangiRathod, SharvariGovilkar, “Survey of various POS tagging techniques for Indian regional languages”,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2525-2529.

[9] Sunil A. Khillare,Bharat A. Shelke, C. NamrataMahender, “Comparative Study On Question Answering Systems And Techniques”, International Journal Of Advanced Research In Computer Science And Software Engineering, Volume 4, Issue 11, November 2017.

[10] K, P. C. Reghuraj, “A Natural Language Question Answering System in Malayalam Using Domain Dependent Document Collection as Repository”, International Journal of Computational Linguistics and Natural Language Processing Vol 3 Issue 3 March 2015, ISSN 2279 – 0756.

[11] Zhiguo Gong1 and Mei Pou Chan Faculty of Science and Technology, University of Macau, Macao, PRC 2 Macao Polytechnic Institute, Macao, PRC.

L.I.F.E
(Light facility using IOT For Ethical tribal community)

Durga Prasad K, Dept Of Master of Computer Applications, VTU Center for Post graduate Studies,

Mysuru, karnataka, INDIA

Abstract— **L.I.F.E is a project undertaken primarily to provide a light facility and monitor the installed system for Tribal community, basically resides at the forest. The project also aims to alert these tribal people in case of wildfire anywhere near the location (within 5 km) where these people reside.**

Keywords-
LM35- Temperature sensor
LDR- Light Dependent Resistor
LEO - low earth orbit

I. Introduction

India is a country where about 8.6% of its population people belong to Tribal community according to 2011 census. Most of them are evidently located in Andhra Pradesh, Chhattisgarh, Gujarat, Jharkhand, Madhya Pradesh, Maharashtra, Odisha, West Bengal, and some North-Eastern states and the Andaman and Nicobar Islands. Many of these community people still lack from obtaining power supply (Electricity), as most of them still reside in some parts of a forest where it's strenuous to provide electric supply by the Government for the people of these tribal community because of the following reasons:

- It is arduous to cut down the tree in the forest and bring up the electric lines.
- Most of the times the forest surface are undulation hence it is difficult to setup network tower also electric grid.
- In case of short circuit, it may result in a wildfire.

The below table shows the details about of tribal population of India according to 2011 census

State code	**India/State/UT**	***Total Population***	***Scheduled Tribe Population (Percentage)***
	India	1,028,610,328	84,326,240 (8.2)
01	Jammu& Kashmir	10,143,700	1,105,979 (10.9)
02	Himachal Pradesh	6,077,900	244,587 (4.0)
03	Punjab	24,358,999	-
04	Chandigarh	900,635	-
05	Uttaranchal	8,489,349	256,129 (3.0)
06	Haryana	21,144,564	-
07	Delhi	13,850,507	-
08	Rajasthan	56,507,188	7,097,706 (12.6)
09	Uttar Pradesh	166,197,921	107,963 (0.1)
10	Bihar	82,998,509	758,351 (0.9)
11	Sikkim	540,851	111,405 (20.6)
12	Arunachal Pradesh	1,097,968	705,158 (64.2)
13	Nagaland	1,990,036	1,774,026 (89.1)
14	Manipur	2,166,788	741,141 (34.2)
15	Mizoram	888,573	839,310 (94.5)
16	Tripura	3,199,203	993,426 (**31.1**)
17	Meghalaya	2,318,822	1,992,862 (85.9)
18	Assam	26,655,528	3,308,570 (12.4)
19	West Bengal	80,176,197	4,406,794 (5.5)
20	Jharkhand	26,945,829	7,087,068 (26.3)
21	Orissa	36,804,660	8,145,081 (22.1)
22	Chhattisgarh	20,833,803	6,616,596 (31.8)
23	Madhya Pradesh	60,348,023	12,233,474 (20.3)
24	Gujarat	50,671,017	7,481,160 (14.8)
25	Daman & Diu	158,204	13,997 (8.8)
26	Dadra & Nagar Haveli	220,490	137,225 (62.2)
27	Maharashtra	96,878,627	8,577,276 (8.9)
28	Andhra Pradesh	76,210,007	5,024,104 (6.6)
28	Karnataka	52,850,562	3,463,986 (6.6)
30	Goa	1,347,668	566 (0.0)
31	Lakshadweep	60,650	57,321 (94.5)
32	Kerala	31,841,374	364,189 (1.1)
33	Tamil Nadu	62,405,679	651,321 (1.0)
34	Pondicherry	974,345	-
35	Andaman& Nicobar Islands	356,152	29,469 (8.3)

Table 1 : Total Population, Scheduled Tribes and their Proportions to the total population (2001).

This paper aims to install renewable power source and also helps in monitoring the systems as most of the tribal community people are literates and not able to handle the

issues that may crop up within the system.

Not only in India, but this project can also be implemented in many African, Asian and American countries where many tribal people lodged and isolate themselves from rest of society and locate in woodland, especially in African countries where such Indigenous peoples who are leading their lives without the availability of power supply.. ABOUT L.I.F.E

The main agenda of L.I.F.E is to provide electric supply for remotely located tribal people, especially who reside in woodlands where it is arduous to bring up the electric lines and also these tribal folks are illiterate and unable to overcome an issue that arises within an installed system. The project makes the task effortless for tribal people that they need not bother about switching on or switching off the light. It is accomplished remotely by a firm member who is monitoring the system remotely with the help of IoT technology. The project also aims to install the following technologies as follows:

- Installing inverters/batteries along with solar panels to charge and store power supply.
- Placing temperature sensors within 5km radius of tribal community to monitor wildfire within this zone.
- Installing light monitoring sensor so that, if natural light intensity drops down below a threshold, the organization member shall switch on the light remotely.

II. TECHNOLOGIES USED IMPLEMENTATION

Here are some of the technologies listed below for the implementation of proposed project :

- ➢ BOLT IOT Module
- ➢ LDR
- ➢ LM35
- ➢ Inverters/Batteries with solar panels
- ➢ Satellite Internet

1. BOLT IOT MODULE

BOLT is an Internet of Things platform which provides Hardware + Software + Cloud facility that enables the user to build IoT products and projects. Using BOLT, users can control and monitor devices from any part of the world. [2]

Functionalities of BOLT IOT Module:

- Station mode in which it can connect to Wi-Fi networks.
- When not connected to any Wi-Fi network it hosts its own Wi-Fi hotspot to which users can connect.
- Commands an ESP8266 to do all GPIO and UART based tasks.
- Bolt comes with a Wi-Fi/GSM chip and a cloud platform which helps you connect your devices and sensors to the Internet.
- With Bolt Cloud, you can control and monitor them over the internet.
- Runs at a frequency of 80Mhz.

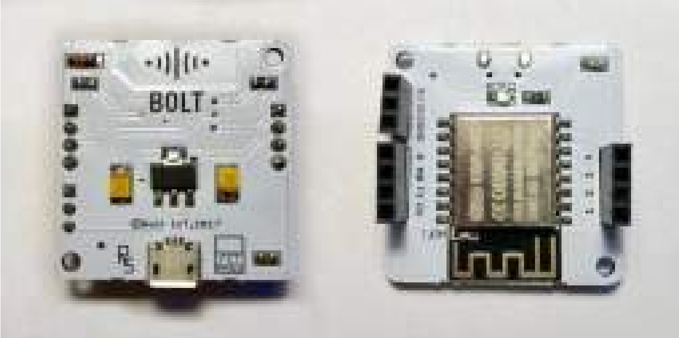


Figure 1: BOLT IOT Wi-Fi Module

2. LDR (LIGHT DEPENDENT RESISTOR)

An LDR is a component that has a (variable) resistance that changes with the light intensity that falls upon it. This allows them to be used in light sensing circuits. In other words, it is a temperature sensor used to sense a change in light intensity. [3]

Figure 2: Light Dependent Resistor

3. LM35 (TEMPERATURE SENSOR)

LM35 is a temperature sensor is a device which is designed specifically to measure the hotness or coldness of an object. LM35 is a precision IC temperature sensor with its output proportional to the temperature (in °C). [4]

Features of LM35:

- Its temperature can be measured more accurately than with a thermostat.

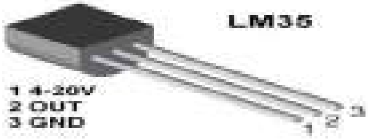


Figure 3: LM35

4. Inverters/Batteries with solar panels

Inverters / Batteries along with solar panels are installed at a dwell where the tribal community resides as these are the only source to store the electric power generated by solar panels and utilized later after sunlight fades off.

There may be a query that may arise as solar panels could generate only enough power during the dry season, what is a solution when it comes to wet seasons?

The alternate source of generating electric power is small scale hydroelectric power generation, as any tribal community would always reside near the water source available, with the availability of water source it is possible to generate electric power.[5]

5. Satellite Internet

Satellite Internet access refers to Internet access provided through satellites. In other words, it is a telecommunications network provided by orbital communication stations. Signals from these satellites allow a user with a dish to have a high-speed internet connection.

Satellite Internet access is generally provided either through low earth orbit (LEO) satellites or geostationary satellites. Signals of geostationary satellite usually are not accessible in some polar regions of the world.[6]

Advantages of using satellite internet:

- Avoiding cut down of trees to bring up network towers within a threshold bandwidth range in the forest to establish internet connection at the tribal dwell.

III. IMPLEMENTATION

As explained previously there are some steps to implement the preferred project. The steps are as follow:

- Installing solar panels around or at the tribal dwell.
- Installing the mini hydroelectric power plant at the nearby water source.
- Placing batteries/inverters, bolt IoT module, LDR sensor and also fire alert siren at the location where tribal folks
- The operating temperature range is from -55°C to 150°C.
reside.
- Installing an LM35 temperature sensor within a 5 km radius.
- Establishing a connection with L.I.F.E Organization using dish/antenna using satellite internet.
- Once all the connection is done, using sensors data firm members can monitor the tribal community location where the system is installed.
- **LM35 sensor**: In case of wildfire within a range. The temperature reaches the maximum threshold and will alert the L.I.F.E organization system and firm members will turn on the siren at a tribal dwell which alerts them.
- **LDR**: In case of natural light intensity reaches minimum threshold during some monsoon season based on sensor readings L.I.F.E organization members get notified and they will switch on the lights.
- **IMPLEMENTING THE CONDITION MONITORING:** With the help of this L.I.F.E organization members can also monitor the overall installed technology and if any issue occurs within an installed system they can fix it remotely if possible.[7]

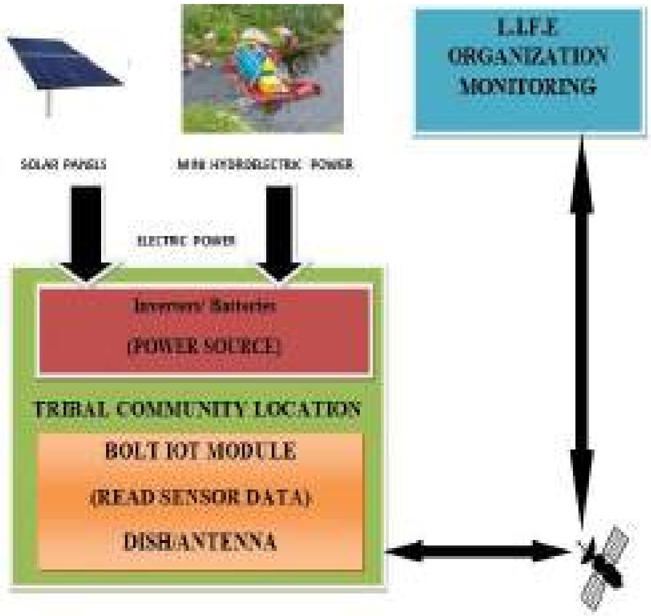


Figure 4: L.I.F.E System Architecture Diagram

- figure 4 shows the overall architecture of a proposed system. Here the figure describes how communication takes place between components installed within a single tribal community system.
- figure 5 shows how a single L.I.F.E firm can monitor 2 or more tribal community dwell at a time. It can be done because each bolt IoT module are identified with different device name and identity.

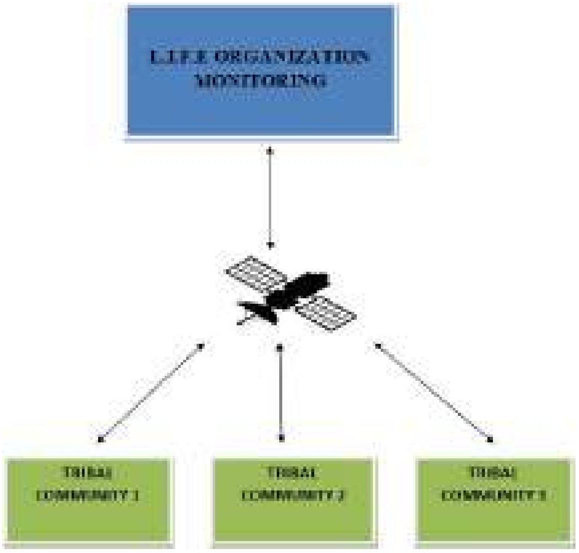


Figure 4: L.I.F.E System monitoring 2 or more tribal community at a time

IV. Advantages of L.I.F.E

- Thousand of trees could be saved from being cut down to bring up the electric lines.
- No need to set up the transmission towers which are based on the electromagnetic waves, which over prolonged usage have adverse impacts on animals and birds as well as on other fauna.
- Satellite dish/antenna will redirect its signal to the direction where the satellite is located at lower orbit and minimize its effects on birds which are affected by electromagnetic waves.
- Many tribal's communities across the world can get electric power i.e.., Light facility who isolate themselves from the rest of society.
- Save tribal people from a wildfire.

Inverters / Batteries along with solar panels are installed at a dwell where the tribal community resides as these are the only source to store the electric power generated by solar panels and utilized later after sunlight fades off.

There may be a query that may arise as solar panels could generate only enough power during the dry season, what is a solution when it comes to wet seasons?

The alternate source of generating electric power is small scale hydroelectric power generation, as any tribal community would always reside near the water source available, with the availability of water source it is possible to generate electric power.[5]

V. Conclusion

Any technology being developed or implemented must not affect our environment especially installing this technology at forest must not affect the habitats. The proposed project minimize such effect to the maximum extent possible and also provide the light facility to tribal people.

References

[1] https://en.wikipedia.org/wiki/List_of_indigenous_peoples
[2] BOLT IOT https://docs.boltiot.com/docs/so-what-is-bolt-iot
[3] Kitronik https://www.kitronik.co.uk/blog/how-an-ldr-light-dependent-resistor-works/
[4] https://wiki.eprolabs.com/index.php?title=Temperature_Sensor_LM35
[5] SIZES OF HYDROELECTRIC POWER PLANTS https://www.energy.gov/eere/water/types-hydropower-plants
[6] Satellite Internet Access https://www.techopedia.com/definition/25271/satellite-internet-access, https://en.wikipedia.org/wiki/Satellite_Internet_accesshttps://dzone.com/articles/benefits-of-industrial-iot-in-condition- monitoring

Automated Smart Shopping Cart Using IOT Based on RFID

[1]Zarina K M, [2]Shaheen Mulla, [3]Farheen kavare, [4]Amreen Darga, [5]Asif Badeghar
Department of Computer Science and Engineering ,SIET, Vijayapur
zarin1100@gmail.com[1], shaheensmulla033@gmail.com[2], farheenkavare29@gmail.com[3]
amreendarga27@gmail.com[4], asifbadeghar65@gmail.com[5]

Abstract*— A supermarket is a place where customers buy their daily using products. There is need to calculate how many products are purchased and accordingly the bill is generated. People purchase variety of items and put them into trolley & go to billing counter for payments. During special offers and festivals, at the billing counter each customer has to wait in the long queue, which is time consuming. So to reduce time we have proposed an "AUTOMATED SMART SHOPPING CART USING IOT BASED ON RFID". In this system, each product is embedded with RFID tag when the product is placed into the cart, product details are read by the RFID reader automatically using Raspberry pi3 . The proposed system suggests implementing smart cart using touch-screen LCD which shows product details and total bill is send to main server using ZigBee technology, each product can be traced and located using product navigation feature.

Index Terms— *Raspberry pi, RFID, Touch Screen LCD, ZigBee*

I. INTRODUCTION

In metro cities purchasing and shopping at supermarkets is a daily activity. We have seen long queues for payment of the bill at marts on holidays and weekends. When there are special offers and discount the crowd is also even more. Customers will purchase products and put it into the trolley. When the customers are done with the shopping, they need to go to billing counter for payment. At the billing counter the cashier will use bar code reader to scan each product every time and prepares the bill. This existing process consumes more time and will creates long queues at billing counters. The drawback of the existing procedure is that each customer has to wait in the queue of the checkout counter and while billing if the customer's budget exceeds then he has to remove some products, which requires some more time to remove the product from the cart.

To overcome the existing drawback we proposed a new methodology Automated Smart Shopping Cart using IoT Based on RFID. It includes RFID reader which reads the RFID tag from the product. The captured information will be displayed on the touch screen LCD . In order to make the customer aware of each product, everytime the item is scanned the price will be displayed on the touchscreen LCD and also the brief information about the product will be displayed. Each product will be traced and located using navigation feature, which helps the customers to find the products easily and saves time. The purpose this project is to provide an automatic billing system where the customer can pay the bill through online payment. The proposed system avoids long standing queue and also saves customer time. The information is send to the master computer using ZigBee.

II. PROBLEM STATEMENT

In big cities purchasing and shopping at supermarkets is a daily activity for every person. Customers waste their time waiting in long queue at the billing counter. To overcome this problem the present technique is based on RFID and ZIGBEE. This present technique provides status updation of complete listings of the products purchased, automatic bill generation, online transaction for billing , management to the central system, tracing of product by navigation feature.

III. RELATED WORK

It's a long standing challenge to develop an smart shopping cart using IoT that can ease the users to have an effective shopping with less time consumption. Certain proposals [1, 3, 4, 5,7] have been made to develop the shopping cart using RFID readers. Aishwarya et al in 2018 Survey [1]This Proposal is to develop an automated trolley using RFID technology.This approach makes billing process simple and also payment can be done online an

provides a way to reduce the time taken for payment but they do not provide security for the items that are being sold.. Ms. Dhanashree et al [2]Proposed wireless billing trolley includes barcode reader and ZIGBEE (NRF24L01) module . The barcode reader scans all the products when they placed in the trolley. The document of the objects offered is saved in the micro controller along with their costs in addition to the overall expenditure. Then this facts display on our lcd screen and that lcd screen restore on our trolley which will show the product prizes and also total bill and every trolley has its own trolley number. At last employee get an itemized bill from all trolley and after printing bill all data will be erased. Goddndnla Aruna on IOT with RFID tag. RFID tag can be attached to each product which when placed into a cart can be automatically read by the RFID reader. instead of manually scanned by a laborer. The future research will focus on improving the current system, for example to have higher efficiency and how to improve the communication efficiency for security properties.

Hyder Ali Hingoliwala, et al[4] Proposed a **smart cart using a LCD** which will show the product details using RFID. The LCD displays an overview of the entire products The system has a centralized billing system because of the customer need not wait in long queues. In future it focuses on adding a new feature that is navigation feature where each product can be traced and located.Pritha N1 et al [5] Proposed a system in which When the product is placed into a smart cart, the product detail is automatically read by the cart equipped with an RFID reader. Hence, billing is made from the shopping cart itself . Also, expiry date of the product is displayed and the damaged products can be identified with respect to its weight. And enhancing the shopping experiences and security issues.

Rajesh Nayak, et al [6] Proposed **The Automated Shopping Trolley is a Smart Trolley** which consist of Barcode Scanners, Arduino, GSM module, Weighing Sensor in it. The scanned items will be stored into the database and thereby can generate the total bill for items purchased. The weighing sensor checks the weight of the products. These all modules are integrated into an embedded system.. and in future it focuses to make this system more robust. This can be done with respect to billing to browse the offers, deals and facility of billing payments online can be used to make cart more advance and will increase the customers experience.

Manikandan Tet al [7] Developed a smart trolley in which the customers have to place the products in trolley after scanning and when done with shopping amount will be displayed in the trolley. The payments can be done either through ATM cards or through pre-recharged customer card provided by the shop. To ensure the security for preventing theft and has facility for cautioning the users who unknowingly add the product without scanning into the trolley. Future advancement is to use high frequency RFID reader which can read multiple tags simultaneously. To avoid smart card and GSM a Mobile application can be developed.

Most of these methods include the barcode reader to scan the product which is very time consuming and customers have to wait in the long queue and multiple products cannot be scanned at the same time. Our approach of shopping cart differs from other adaptive approaches by using the RFID reader to scan the RFID tag and displays the details on the touch screen LCD and also the product is traced and located using the navigation .

IV. PROPOSED SYSTEM

In the existing system, bar code reader is used to scan the product details where the customers have to wait in long queue for generating the bill. Sometimes, the bar codes would have been damaged and that particular product could not be scanned by a barcode scanner leading to confusion. Also, each and every product has to be scanned manually. In order to solve the previously identified problem,save costumers time and help the retailers .we are going to propose a system, which is helpful for both customers as well as shop owner.

The proposed methodology shows automated billing for customers in supermarket. RFID reader scans the RFID tag from the product and those readings will be displayed on the Touch screen LCD. The location of particular product is displayed on the LCD in the navigation feature. The option of payment is provided by online payment methods.

To design the system the different kinds of modules are combined together. The main modules that are used are touch-screen LCD, microcontroller. RFID Tags, RFID Reader, RF Transmitter, RF Receiver, ZigBee and EEPROM. All these modules are individually programmed and then combined together.

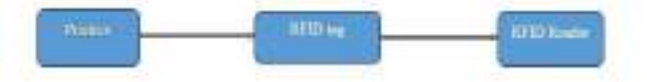
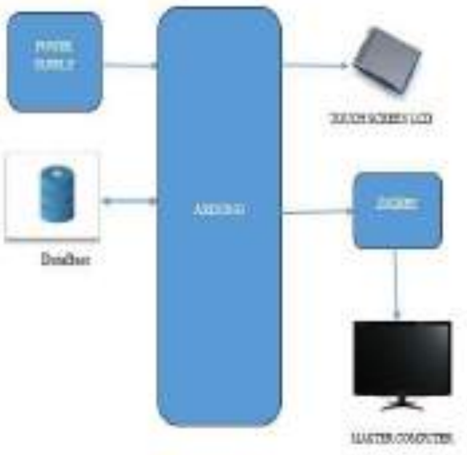


Fig :3.1 . RFID scanning process



First the system is initialized and then it is ready to use for customer. If customer wants to purchase any product then he/she has to place the product in the trolley. Every product in the shop contains RFID tag on it. When the product falls in the trolley the RFID reader reads the RFID tag placed on the product. This RFID reader is connected to the microprocessor. Microprocessor crosschecks the information from the RFID reader and information in the memory of microprocessor. If the information is matched then the cost of product, name of product, expire date of product and the total bill will be displayed on touch screen LCD. If user wants to remove any product the amount will be deducted from the total bill. After the payment is done, the Cart must get reset. The Payment procedure of billing can also be done online. RFID reader and ZigBee Transreceiver is implemented on each cart. ZigBee transfers the information to the main server which is in the range. This main server has its own cloud from that owner can access the information from the cloud irrespective of the time with the help of user ID and password. This is the concept of Internet of thing (IOT).

V. IMPLEMENTATION AND RESULTS

- Raspberry pi3 is used as the microcontroller . The project is implemented using python. RFID reader is used to read the information of the product.

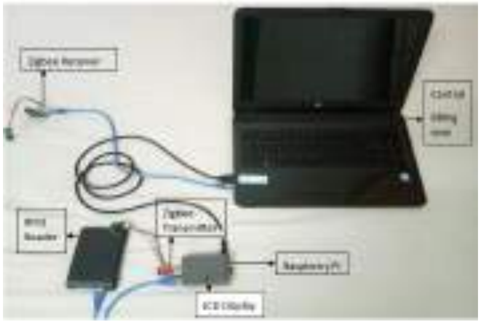


Fig 5.1 Hardware circuit diagram

Whenever the products get added into the cart its information like product name, cost will be displayed on the touch screen LCD. If product is removed from the trolley, then cost of respective product will be deducted from total amount. After purchasing is done the total bill will be displayed and the user can pay through online payment method. The same bill will be communicated to the main server through ZIGBEE.

Fig5.2: Installion of OS (raspbian) in raspberry pi 3

Fig 5.2: RFID Reader and raspberry pi setup

Fig 5.3: Scanning RFID tag using RFID reader

VI. CONCLUSION

Automated Smart shopping trolley system creates an automated central billing system in malls. By using the ZigBee, the product information is directly sent to billing system. So that customers no need to wait in a long queue. It is trustworthy, highly dependable and time efficiency. The proposed smart shopping trolley system will reduce the customers time in searching the location of the product. The customer just types the name of the product he/she want to purchase ,on the touch screen LCD. The trolley will automatically guide them to the location of the product.

REFERENCES

[1] S. Aishwarya, D. Gomathi Shankari, R. Ilakkiya, S. Prasanth,Ms.S. SriHeera, Survey on Smart Trolley System based on IOT in Journal of Recent Activities in Infrastructure Science Volume 3 Issue 1 , MAT Journals 2018 .

[2] Ms. Dhanashree R. Ajagekar, Ms Rajashree B. Tapase, Ms. Amruta B. Suryawanshi. Ms. Sangeeta A. Virkar,2018 Wireless Billing Trolley in Journal of Control and Instrumentation Engineering Volume 4 Issue 1, MAT Journals 2018.

[3] GoddndnlaArunaKumari, GuduruMallikarjuna, K.Nagalatha, High Security System for Smart Shopping using IOT with RFID Tag in International Journal of Research Volume 7, Issue VIII, August/2018.

[4] Hyder Ali Hingoliwala, KalyaniBedre, ShrutiDeshmukh,Akshata Bhosale,2018 Smart Follower Sensing Shopping Cart using Centralized Billing System in International Journal on Recent and Innovation Trends in Computing and Communication. Volume: 6 Issue: 2, February 2018.

[5] Pritha N1, Sahana,] SelvinStephy, Shiny Rose , Unnamalai S1, Smart Trolley System for Automated Billing using

RFID and IoT in International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 04 | Apr-2018 .

[6] Rajesh Nayak, Ravi S Raikar, Yogendra, Vishwas, Automated Trolley for Shopping in International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 5, Issue 6, June 2017.

[7] Manikandan T, Mohammed Aejaz M.A, Nithin Krishna N.M, Mohan Kumar A.P, ManigandanR, RFID based Advanced Shopping Trolley for Super Market in Journal of Chemical and Pharmaceutical Sciences Special Issue 8: June 2017

Fault Tolerance Control in Industry of Metal Crack Detection & Segrigator Using Image Processing

Prof. Geetha. M[1], Bhoomika. K. S[2]
Assistant Professor[1], Student[2]
Information Science Engineering,
Nagarjuna College of Engineering and Technology, Karnataka,
Bangalore, India
geethaneha5@gmail.com[1],bhoomikasuresh33@gmail.com[2]

Abstract - This is the system for automatic crack detection is presented. The system is capable of analysing metal parts by means of a laser excitation system and a thermographic camera. The laser creates thermal gradients inside the part under inspection, and the thermal camera observes how heat diffuses inside the part. Cracks can be automatically detected by using computer vision algorithms specifically developed for this task, which might be capable of measuring and classifying heat profiles. Different algorithms must be developed for rugged and smooth metal parts, since the reaction to laser excitation is rather different. The detection algorithms can be tested on several sequences and showed very good detection performance also with cracks of very small size, having a width.

INTRODUCTION :

Quality inspection at the end of a production line is an important stage in industry, especially for high-performance components. Parts undergoing strong mechanical and thermal stress should be carefully checked, since small defects can affect performance and reliability of a component. Crack detection is one of the most common checks to be performed, because cracks are a common source of failure, and they affect a high number of different productions.

For metallic parts, crack detection is still performed exploiting a technique called “magnetic particle inspection” (MPI): the part to be analyzed is first washed, then put into a magnetic field and finally covered with magnetic particles, either in the form of a dry powder or more frequently, in a wet suspension. Cracks are easily detected because they cause leaks in the magnetic flux; such leaks are highlighted by the particles, which can be inspected by means of a UV light. The whole process is very complex and needs to be done manually; it is also extremely time-consuming, because parts need to be cleaned, magnetized, covered with particles, inspected, de-magnetized and cleaned again. Moreover, magnetic particles and their carrier are a source of pollution, and should be properly processed after use.

Given the complexity of MPI, a method for simplifying the process of crack detection and making it automatic is highly desirable: investigation on this topic is the aim of the ThermoBot project. The main idea is to exploit thermography instead of magnetic particles to detect cracks, and to apply this method to parts made of non-metallic materials, like carbon fiber. Inspection is performed by means of a laser and a far infrared (FIR) camera (also called thermal camera or thermocamera), that observes how the heat carried by the laser diffuses inside the part since cracks cause alterations on the heat flux, these can be exploited to detect cracks.

Methods based on image analysis have also been exploited in the literature, ranging from detection of welding defects in pipelines to concrete surface analysis and the protection of cultural heritage. Thermographic image analysis systems have recently been proposed for performing in-situ non-destructive inspections during thermomechanical fatigue tests; the system showed a high sensitivity, being able to detect cracks smaller than 500 µm. The system proposed is slightly different from the others discussed above as it is meant to inspect different types of materials during fatigue tests, and detect the cracks as soon as they appear.

Material Segregator:

In recent times, garbage disposal has become a huge cause for concern in the world. A voluminous amount of waste that is generated is disposed by means which have an adverse effect on the environment. The common method of disposal of the waste is by unplanned and uncontrolled open dumping at the landfill sites. This method is injurious to human health, plant and animal life. This harmful method of waste disposal can generate liquid leachate which contaminate surface and ground waters; can harbour disease vectors which spread harmful diseases; can degrade aesthetic value of the natural environment and it is an unavailing use of land resources. In India, rag pickers play an important role in the recycling of urban solid waste. Rag pickers and conservancy staff have higher morbidity due to infections of skin, respiratory, gastrointestinal tract and multisystem allergic disorders, in addition to a high prevalence of bites of rodents, dogs and other vermin. Dependency on the rag-pickers can be diminished if segregation takes place at the source of municipal waste generation. The economic value of the waste generated is not realised unless it is recycled completely. Several advancements in technology has also allowed the refuse to be processed into useful entities such as Waste to Energy, where the waste can be used to generate synthetic gas (syngas) made up of carbon monoxide and hydrogen. The gas is then burnt to produce electricity and steam; Waste to Fuel, where the waste can be utilized to generate bio fuels. When the waste is segregated into basic streams such as wet, dry and metallic, the waste has a higher potential of recovery, and consequently, recycled and reused. The wet waste fraction is often converted either into compost or methane-gas or both. Compost can replace demand for chemical fertilisers, and biogas can be used as a source of energy. The metallic waste could be reused or recycled. Even though there are large scale industrial waste segregators present, it is always much better to segregate the waste at the source itself. The benefits of doing so are that a higher quality of the material is retained for recycling which means that more value could be recovered from the waste. The occupational hazard for waste workers is reduced. Also, the segregated waste could be directly sent to the recycling and processing plant instead of sending it to the segregation plant then to the recycling plant. Currently there is no system of segregation of glass, plastic and metallic wastes at an industry. The purpose of this project is the realization of a compact, low cost and user friendly segregation system for urban households and scrap shops to streamline the waste management process.

PROPOSED METHOD:

The block diagram shown in Figure 1 represents the automated waste material segregator where three types of materials are segregated namely Metal, Glass and Plastic. The controller used is Arduino UNO. An object is placed on the conveyor which runs on a motor of 12v, 1A which is connected through the motor driver and is programmed to run in clockwise direction by the Arduino.

The object is placed on the conveyor, depending on the output of inductive sensor and capacitive sensor the motor driver drives the motor.

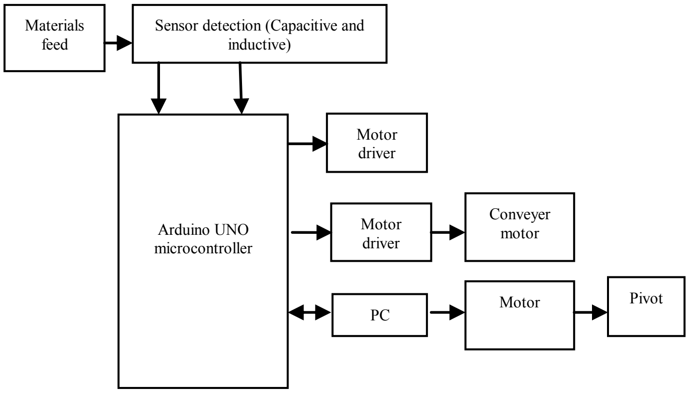

Materials feed
Sensor detection (Capacitive and inductive)
Arduino UNO microcontroller
Motor driver
Motor driver
Conveyer motor
PC
Motor
Pivot

Figure 1: Block diagram of proposed system

Arduino Uno Act as a microcontroller, Arduino Uno is based on the ATmega328. It has 14 digital input/output pins, 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, and a reset button. The board can be programmed with Arduino Software (IDE). The board can operate on an external supply from 6 to 20 volts. If using more than 12V, the voltage regulator may overheat and damage the board. The recommended range is 7 to 12 volts. The ATmega328 has 32 KB flash memory. It also has 2 KB of SRAM and 1 KB of EEPROM. Inductive proximity sensor Inductive proximity sensors operate under the electrical principle of inductance. Inductance is the phenomenon where a fluctuating current, which by definition has a magnetic component, induces an electromotive force (emf) in a target object. To amplify a device's inductance effect, a sensor manufacturer twists wire into a tight coil and runs a current through it. An inductive proximity sensor has four components; the coil, oscillator, detection circuit and output circuit. The oscillator generates a fluctuating magnetic field the shape of a doughnut around the winding of the coil that locates in the device's sensing face. Inductive Proximity Sensors being contactless sensors can be used for position sensing, speed measurement, counting, etc. They can be used in extreme conditions, such as oily, dusty, corrosive environment. Their application ranges from Automobile Industries to Steel Industries, from CNC/NC machines to material handling equipment, process automation, conveyor systems, and packaging machines.

Capacitive proximity sensor Capacitive proximity sensors use the face or surface of the sensor as one plate of a capacitor, and the surface of a conductive or dielectric target object as the other. The capacitance varies inversely with the distance between capacitor plates in this arrangement, and a certain value can be set to trigger target detection. The sensing surface of a capacitive sensor is formed by two concentrically shaped metal electrodes of an unwound capacitor. When an object nears the sensing surface it enters the electrostatic field of the electrodes and changes the capacitance in an oscillator circuit. As a result, the oscillator begins oscillating. The trigger circuit reads the oscillators amplitude and when it reaches a specific level the output state of the sensor changes. As the target moves away from the sensor the oscillator's amplitude decreases, switching the sensor output back to its original state.

Result:

1. Identifies metal and non-metal products
2. Identifies cracks on the metal products
3. Edges and curves can be identified using camera
4. Separator will separate cracked and non cracked metal products
5. Automatic counter to count the number of cracked and non cracked metal products

References:

[1] A. Gachagan, A. McNab, P. Reynolds "Analysis of ultrasonic wave propagation in metallic pipe structures using finite element modelling techniques." Ultrasonics Symposium, 2004 IEEE, 2:938-941, 204.

[2] T.P. Theodoulidis, S.M. Panas, E.E. Kriezis, "Eddy current detection of crack orientation using elliptical excitation", Science, Measurement and Technology, IEE Proceedings, vol.141, no.1, pp.41-47, Jan. 1994.

[3] P. Xu; K. Shida, "Eddy current sensor with a novel probe for crack position detection", Industrial Technology, 2008. ICIT 2008. IEEE International Conference on, pp.1-6, 21-24 April 2008 doi: 10.1109/ICIT.2008.4608445.

[4] G.Y. Tian, A. Sophian, D. Taylor, J. Rudlin, "Multiple sensors on pulsed eddy-current detection for 3-D subsurface crack assessment", Sensors Journal, IEEE, vol.5, no.1, pp.90-96, Feb. 2005. doi: 10.1109/JSEN.2004.839129.

An Effective Way of Web News Mining With The Use of New Features

Murali M Student [1], Shruthi B Gowda [2] , Assistant Professor, Computer Science And Engineering, Vivekananda Institute Of Technology, Bangalore India. murali.muniyan@yahoo.com [1], shruthi.b.gowda@gmail.com [2]

Abstract— **Internet users widely use the web based applications for information exchange. But it is very challenging in the modern world to process the huge amount of data such as web new, advertisements of product. The Internet users use the web applications to get the updated information which need huge computation in terms of space and time leading to draining up of battery power of the user's devices. This problem could be overcome by mining or extracting the specific data based on the user's behaviour and the information gathered from different sources. A solution to this problem can be by developing a web based application which refine and extract news information using new features and such as geolocation and time information as well as shows a comparative study on three different mining techniques.**

The application can run on different devices including Laptops, Smartphones and Tablets devices as well as the application can retrieve information features accordingly. Then this obtained information could be used as a basis for starting or as input for the data-mining techniques including K-Nearest-Neighbor (k-NN), decision tree and deep-learning recurrent neural network such as Long Short-Term Memory (LSTM). These techniques are implemented separately and they are compared in terms of time/space complexity and classification accuracy. The obtained results showed that the mining accuracy via k-NN is the worst one with 85% and takes huge time, while the mining accuracy through using LSTM is the best one and its accuracy around 94%, when location information is used.

Index Terms— Web News Mining, KNN Algorithm, Decision Tree Algorithm, Long Short Term Memory Algorithm

I.INTRODUCTION

The use of web applications by the users has increased from the last decade for different types of e-services like billing, communications, etc. Web applications are well known platform which are widely used by the users for information exchange and to get up to date information [6]. But processing of huge information like web news, videos, images and advertisements is a challenging task. The processing of these huge data may reduce the performance of the system in terms of time and space, which leads to complexity in the system. It will also drain the battery of the user's devices. And also due to heavy amount of data accessing through the network will result in network traffic [7]. Hence it will reduce the performance of the system.

Datamining is an approach by which we can overcome this type of problem. Datamining is a process of extraction of specific data from a large volume or a heap of data based on some criteria. In this process some patterns is selected in large number of information records by using a set of methods at the intersection of statistics method, database-engine, and machine learning algorithms. The mining some data may need pre-processing or post-processing of data, however the result of the datamining mainly depends on the specified feature based on which the data is extracted. For example in web mining the features such as user's behaviour, frequently visited contents, users search history etc., may be considered. The use of this technique might reduce the content and retain only the required data from large data or Big-data.

This article is mainly focused in mining the web contents based on features like geolocation of the user. When the user uses the web application, the application without any action or interaction with the user it extracts specific content from the large amount of news data based on the location of the user. The process will not learn any user's behaviour instead it retrieves only reduced amount of data or specific content automatically with respect to the location and time of the user. The web application will first retrieve the location and time information from the user's device. Then it will make use of that information for mining the web contents and reduce the data or number of records from the content. By this approach only the contents which is related to user's location is retained from the huge data. Further the main objective of the article is to reduce time and space complexity by the use of new features such as user's geolocation and time and to make a comparative study on three well known classification techniques KNN algorithm, Decision tree and Deep Learning Recurrent Neural Network.

II. LITERATURE SURVEY

Mazhar Iqbal Rana , Shehzad Khalid and Muhammad Usman Akbar proposed News Classification Based On Their Headlines: A Review [1] on 2014. In this article, the process of classification of news headings are made by using machine learning techniques. At first step the unstructured data gathered is pre-processed and converted to a structured data which does headlines tokenization, removal of diacritics, stop words, frequent words, etc. Then the news headlines are indexed. After that the highly specific features are selected from the headlines using Boolean weighted, Class frequency threshold holding and Information gain. Then the headlines are classified using k-NN, Naïve Bayes, Artificial Neural Network and Decision tree techniques. Finally it is observed that different classification scenarios and algorithms perform differently depending on news and data gathered.

Sukhpal Kaur and Er. Mamoon Rashid proposed Web News Mining using Back Propagation Neural Network and Clustering using K-Means Algorithm in Big Data [2] on 2016. It focuses on research to manage huge amount of news like BBC news channel data gathered from the internet. To manage the large web news data, a clustering based K-means and Back Propagation Neural Network algorithm for classification is proposed. K-means algorithm is used to make the cluster and BPNN algorithm is used to classification of news that check the false error and rejection rate of system and tests the accuracy where the news are categorized as Movies, Sports, Politics and Normal.

Syafruddin Syarif, Anwar and Dewiani proposed Trending Topic Prediction by Optimizing K-Nearest Neighbor Algorithm [3] on 2017. The K-Nearest Neighbor (KNN) method is used in prediction of the trending topic based onthe membership distance of a class. The research was conducted based on the news and conversation taken from the online and social media related to Makassar City Government with 393.667 raw data, in which the preprocessing was thencarried out to determine the trending and non-trending conversations, producing 2007 trained and tested data. Based on the analysis it shows that highest accuracy value of 81.13%, in 60% data tested.

Husna Sarirah Husin, James A. Thom and Xiuzhen Zhang proposed News Recommendation Based on Web Usage and Web Content Mining [4] on 2013. It mainly focuses on web usage on web content mining techniques using which, the news articles are recommended to the users. The use of web server log of a Malaysian newspaper website, Berita Harian and IP address of the user is collected. A time series analysis is made on the server log to predict each different user even if they make use of the same system based on their session. And using the gathered data, finally the news content is recommended to the users.

III. EXISTING SYSTEM

The news content in our daily life is rapidly increasing resulting in a large set of data. The processing of such huge amount of data is very essential to retain only the required and important news from the large set of data. In current system, Back Propagation Neural Network (BPNN) has been developed for mining from the big news portal such as BBC news website. During the training phase, the records of the news are likely will be the input of the network. Once, the network is learned, then in the test phase, when any news is inserted into the website, it will be classified according to the trained classes. The class-examples are entertainment news, health news, business news and etc. In such manner, when the users visited the website, the number of the big news records will be reduced accordingly. Some of the advantages of the existing system are- classification of large data set into different categories, required data is selected rather than browsing all the data and reduction is time complexity due to reduced set of data. And the disadvantages faced are variation in the accuracy of the result obtained, need of further enhancement in precision and accuracy of the algorithm and improvement in terms of time and space complexity.

IV. PROPOSED SYSTEM

The news mining from the large amount of data is improved in terms by using new approach. This is achieved by using the geolocation of the user and applying the k-NN (K Nearest Neighbor), decision tree and ANN (Artificial Neural Network) techniques. The news document is collected and then pre-processed before classification. Then the data is represented in workable format for classification. The represented news data is then classified in the next step as shown in the below figure 1. Finally the system outputs the classified news data which will be viewed by the user. This approach can extract specific data from large and dynamically changing news data. The resulted content will be very high accuracy and precision when compared with traditional classification techniques and also will reduce the time and space complexity compared with the existing system.

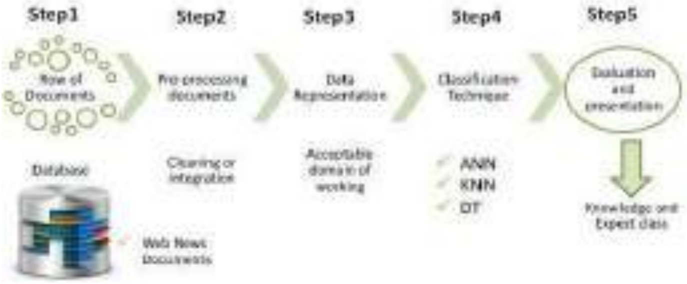


Figure 1: Flow diagram of the proposed system

V. CLASSIFICATION TECHNIQUES

1. *K-NN (K-Nearest Neighbor) Algorithm*

The k-nearest neighbors algorithm (k-NN) is anon-parametric method used for classification and regression. K-NN Algorithm is one of the simplest and easy to understand algorithm which is used for classification [8]. This algorithm is used in finding the relative distance between the different news gathered. The algorithm is started by getting a set of documents from the training dataset as well as passes the documents through the pre-processing step. The model of the K-NN algorithm works in the following ways:

1) Calculation of the Euclidian distance of the training dataset and the test dataset gives as input.
2) Then the calculated distance of each document is sorted in ascending order.
3) The k nearest document is selected based on the k smallest value.
4) Then the corresponding class or category of the testb document is predic

2. *Decision Tree Technique*

The decision tree is a powerful and popular technique used for prediction and classification [11]. The Decision tree is working by given a set of training objects and their attribute values, and then it is trying to determine the target attribute value of the new examples. When the technique is applied to a set of training objects, a tree is constructed, and then the tree could be easily translated to a set of rules. The procedure of the technique is as the following:

1) In the first step, the entropy value of the training documents set is computed.
2) Then the probability information value is calculated for each attribute.
3) In the next step, the gain value is calculated for each attribute.
4) Finally, the attribute with the biggest gain value among all the attributes is selected as a root of the tree. This means that the process will continue until the leaf node will be reached.
5) Once, the decision tree is constructed, then the tree is translated to a set of rules.
6) Finally, new web news documents from the test data set can be used as inputs to the model tree. Then the tree predicts the class or category of the new news.

3. Artificial Neural Network Technique

The web news text is a sequence of words, where the words of the news may be dependable on to each other. Hence the web news documents are considered to be sequential. Equally, there are several strategies to build neural networks to learn and to classify document text including supervised neural network strategy (such as BPNN) and recurrent neural network strategy (such as long short-term memory 'LSTM' network). The LSTM is a deep-learning algorithm for the purpose of text classification. In this study, the LSTM is used to learn and classify web news text. This is because LSTM can be used to learn long-term dependencies between the sequences of text data [10]. The LSTM algorithm is designed as in the following steps for web news mining:

1) In the first step, the pre-processed web news documents are converted into the numeric values and stored in a vector.
2) In the next step, to equalize the length of the documents, the process of padding to the right of the documents should be applied.
3) As a requirement to build a neural network, there should be the initialization of the sequence input, hidden and output layers. As well as the input and output size should be configured.
4) Then the network should be designed to connect all the nodes in all the layers, accordingly.
5) In the next step, the training parameters should be configured including the number of epochs, learning rate, absolute error, activation functions for the node of all the layers and set the equations of the updating weight and bias.
6) Then the process of the training data sets is started. The training process depends on the number of epochs, initialize absolute error and the number of dataset documents.
7) Once, the network model has been training and learned, and then the model will be ready to test and predict/classify any new coming documents (web news documents in test dataset).

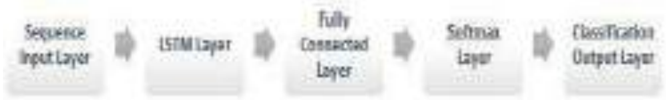


Figure 2: Flow Diagram of LSTM Layer

As it can be seen in figure 2, the main layers of the network are the sequence input layer and LSTM layer [9]. The sequence input layer inputs set of words of the web news documents into the network and the LSTM layer then learns long-term dependencies between the words of the sequence data (web news document). However, due to having multi-classes in web news mining, the Softmax layer is used to introduce the probabilities for each class to the output layer from its network inputs.

VI. CONCLUSION AND FUTURE WORK

The use of location and time based information was implemented in the proposed system to overcome the problems faced by huge data processing. With the use of new feature such as location and time based information the content of the web data was reduced thus reducing the space and time complexity. Further the traditional classification techniques were applied along with this to get more accurate result with best classification. The implementation and analyzation were done by applying three classification techniques to find the more accurate and well classified content. From the comparison of the result the Long Short Term Memory (LSTM) algorithm is found to give more accurate result, as the words of the news text had long term relationship in them. Thus a technique to provide more accurate and best classification is predicted.

It is observed that the application of these new features in web mining could provide sufficient data without any interaction with the user's behavior. And from the literature survey it is proven that the use of this type of feature such as location and based information is not implemented along with the classification techniques in web mining. Thus it gives us the best and more accurate result with the use of these new features.

From the results, it shows that the processing of huge data is made simpler by use of location and time information. And thus only required data is provided to the user that is processed based only on the location. In future consideration, the system could be made more efficient by applying combination of the existing system and the proposed system, where the system makes use of location information at the beginning sage and further reads about the user activities. This process could provide more accurate and clear results as it uses both the method. At beginning stage with the use of the location information the number of records of the documents may be reduced. Later on studying the user activities the system can understand the topic or the content in which the user is showing more interest. Hence it could only retain sufficient data with very less number of records and only the contents on which the user is interested. This might also increase the accuracy rate of the result, as extracted less number of records based on the interest of the user.

ACKNOWLEDGMENT

I am very thankful to Mrs. Shruthi B Gowda who supported and guided for completion of this Paper. I am very much grateful to the entire CSE department of Vivekananda Institute of Technology, Bangalore for giving me all facilities and work environment which enable me to complete my task. I am also thankful to all researchers which I have mentioned in reference list. I am also thankful to all my family members and friends for supporting to complete my paper.

REFERENCES

[1] Mazhar Iqbal Rana, Shehzad Khalid, Muhammad Usman Akbar, “News Classification Based On Their Headlines: A Review” *IEEE 17th International Multi-Topic Conference (INMIC)* 2014.

[2] Kaur, Sukhpal and Rashid, Er Mamoon, "Web news mining using back propagation neural network and clustering using K-Means algorithm in Big data," *Indian Journal of Science and Technology*, vol. 9, no. 41, 2016.

[3] Syafruddin Syarif, Anwar, Dewiani, “Trending Topic Prediction by Optimizing K-Nearest Neighbor Algorithm”, *5th IEEE International Conference on Broadband Network & Multimedia Technology (IC-BNMT 2013)* 2017.

[4] Husna Sarirah Husin, James A. Thom, Xiuzhen Zhang, “News Recommendation Based on Web Usage and Web Content Mining” *IEEE 29th International Conference on Data Engineering Workshops (ICDEW)* 2013.

[5] Chenbin Li, Guohua Zhan, Zhihua Li “News Text Classification Based on Improved Bi-LSTM-CNN” *9th International Conference on Information Technology in Medicine and Education*, 2018.

[6] Jayamalini, K and Ponnavaikko, M, "Research on web data mining concepts, techniques and applications," in *Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017 International Conference on*, 2017

[7] Terzi, Duygu Sinanc and Terzi, Ramazan and Sagiroglu, Seref, "Big data analytics for network anomaly detection from netflow data," in *Computer Science and Engineering (UBMK), 2017 International Conference on*, 2017.

[8] Chaurasia, Vikas and Pal, Saurabh, "A Novel Approach for Breast Cancer Detection Using Data Mining Techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 1, p. 17, January 2017.

[9] Hochreiter, Sepp and Schmidhuber, Jugen, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735--1780, 1997.

[10] Murdoch, W James and Szlam, Arthur, "Automatic rule extraction from long short term memory networks," *arXiv preprint arXiv:1702.02540*, 2017.

Shaikhina, Torgyn and Lowe, Dave and Daga, Sunil and Briggs, David and Higgins, Robert and Khovanova, Natasha, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," *Biomedical Signal Processing and Control*, p. 7, February .

HoneyPass: A Shoulder Surfing Resistant Graphical Authentication System Using Honeypot

Arun Kumar S [1], Rashika R [2], Renu R [3], Ramya R [4]
8 th Semester B.E Student[1, 2, 3, 4]
Sapthagiri College of Engineering, Bengaluru
arunkumars@sapthagiri.edu.in [1], rashikarajaraman@gmail.com [2], renurudramurthybt@gmail.com [3], ramyahai01@gmail.com[4]

ABSTRACT

In today's modern world, securing the organization's data has become a major concern. To provide security, the most widely recognized authentication methods are credentials, OTP, LTP etc. These methods are more prone to Brute Force Attack, Shoulder Surfing Attack, and Dictionary Attack. Shoulder Surfing Attack (SSA) is a data theft approach used to obtain the personal identification numbers or passwords by looking over the user's shoulder or by external recording devices and video capturing devices. Since SSA occurs in a benevolent way, it goes unnoticed most of the times. It is one of the simple and easy methods for hackers to steal one's sensitive information. The hacker has to simply peek in while the user types in the password without any much effort involved. Therefore, this phenomenon is widely unknown to people all over the world. Textual passwords are a ubiquitous part of digital age. Web applications/mobile applications demand a strong password with at least one capital letter and a special letter. People tend to give easy passwords in order to remember them which can be easily shoulder surfed. To overcome this, graphical password techniques are used to provide a more secure password. In the graphical authentication system, the users click on target images from a challenge set for authentication. Various graphical systems have been proposed over the years which are shown to be more secure when compared to other authentication systems. In this paper, a shoulder surfing resistant graphical authentication system is implemented using honeypot concept.

Keywords: Shoulder Surfing Attack, Textual Password, Graphical Authentication System, Honey Pot Concept

1. INTRODUCTION

Security plays a vital role in any organisation. Data protection is one of the main challenges faced in any business environment. In order to protect any resources, the companies undertake various security measures. However, security has become a worldwide problem as websites have become an integral part of everyone's life [1]. The uncompromising security issues that has to be addressed in websites occur during the user authentication phase. In today's computer world, authentication is very important in order to keep the unauthorized users from accessing the protected resources. Authentication is a process that allows a device to verify the identity of a person who connects to a network resource. In order to keep the users' data private, authentication mechanism is used wherein the user types in the username and password to access their private account. However, people performing authentication process in public results in shoulder surfing attack. [2]

Shoulder Surfing attack is a direct observation approach where the shoulder surfer steals the user's Personal Identification Number (PIN), passwords by looking over his shoulder. [2,3] It commonly happens in public transports while the victim is commuting which involves a smart phone in almost all cases. A good example is shoulder surfing at ATMs, a crime in which a suspect watch over the victim's shoulder as he punches in his PIN number. The ATM screen asks for another transaction when the customers complete theirs. Some customers fail to notice the prompt and walk away leaving it on the screen. In this way, the thief enters the stolen PIN and pretends to be the user. But the phenomenon of shoulder surfing is not widely known. [4] Users tend to use the strategies such as hiding the device screen, shielding the device with their hand etc. However, by observing, one cannot get a hold with most of the victim's detailed biodata such as information about his relationships, sexual preferences, interests, hobbies, and login data. Hence, the damage shoulder surfing can cause is widely unknown. [5]

Textual password approach is used tremendously all over for authentication. During the authentication phase, websites demand strong passwords with at least six to eight characters comprising of uppercase and lowercase alphabets, numbers, and special characters. Such passwords are believed to prevent brute force attacks. [6] A password cannot be remembered if its strength is more. In [7] today's digital age, websites play a major role in one's life. People are part of an enormous amount of such websites with each containing the authentication phase where the user validates him by entering a password to access their private data. In order to remember all such passwords, the user tends to choose the same password for multiple websites which makes the password unprotected for the hackers to break. A more complex password is shoulder surfing resistant. Thus, these passwords can be easily revealed if the shoulder surfer peeks or uses video recording devices [7].

Graphical authentication systems are used in order to overcome the disadvantages of textual password systems. Here, images are used as the password instead of a string of characters. These graphical passwords are expected to be stronger and safer than textual passwords. [8] Several studies prove that a human brain has a better ability to memorize and recollect images easily when compared to a string of characters. Since it is easy to recollect the password, the user need not choose the same password for multiple websites. It makes it hard for the assailants to break the password if the user prefers to use a strong graphical password. This, in turn, increases the security level during the authentication phase. A strong graphical authentication system not only safeguards the password from brute force and dictionary attacks, but also from shoulder surfing attacks. Since shoulder surfing can create damage to the user during authentication, a strong graphical password is preferred over a textual password according to the studies conducted [9, 10].

1.1 OVERVIEW OF AUTHENTICATION SCHEMES

Password-based authentication schemes have been most commonly used on many smart devices when compared to other authentication schemes. The lower complexities in implementation, computation, processing requirements and so forth have led to the use of a password-based authentication system. Again, text-based passwords are more commonly used when compared to other existing authentication systems. However, various vulnerabilities were discovered by several cryptanalysts in text-based systems like brute force attack, guessing attack, dictionary attack, social engineering attack etc. In smart phones, the tiny screen size imposes some more constraints such as limited password length, implementation of easier authentication systems to increase performance etc. Moreover, the small on-screen keyboard makes typing inefficient and less precise. Consequently, the users tend to use a smaller password which makes it even more vulnerable. Since the size of smart devices is getting smaller and smaller; few authentication systems cannot be implemented in it due to its size [11].

The invention of graphical password authentication systems was triggered by the well-known limitations of textual password authentication systems. The graphical authentication systems have been generally categorized into draw metrics, loci-metrics, and search metrics systems. In draw metrics-based systems, the users will have to recall and reproduce the predefined pattern on a canvas to use the system. In loci-metrics-based systems, the users will have to recall and select the previously defined points in an image in order to log in to the system. In search metrics-based systems, the users will have to choose the predefined target images from the displayed challenge set. During the login phase, the system throws in with entirely the same images or with a few different images which were displayed to the user during the registration phase. The selection of correct target images will let the user access the system. Shoulder surfing has always been a problem in these systems because of the use of the graphical interface [12].

Many authentication systems have been evolved over the years. Today, biometric authentication system holds a prominent place as many users utilize them over textual or graphical based authentication systems.

[13] However, one study showed that for mobile authentication, 70% of the users preferred PIN or android graphical pattern even though they are more prone to attacks. The users tend to opt the textual-based method as they don't care about the security but the ease with which they can simply get over with the login phase. Thus, knowing this fact, the attacker will try to break into those systems which use textual or graphical based systems. Besides, biometrics wouldn't be the one used for authentication if the users give more priority to the ease of use when compared to other technologies. Biometrics also lacks privacy, reliability, and security. Thus, the existence of PIN and pattern approaches is present even in the overexposure of biometrics [14].

1.2 HONEYPOT CONCEPT

In computer terminology, Honey pot is a computer security mechanism set for detecting and counteracting attempts from unauthorized users. Honeypot is designed by the system just to get attracted by the attackers and intruders. The main function of the honeypot is to portray itself as the possible intent for the attackers mainly server to collect information and to report the defenders about the attempts to access the honeypot by intruders. Honeypots are mostly used by cybersecurity research companies and enterprises, to examine and defend their system from being attacked from potential threats. Honeypot is an important implementation tool for business associations and cybersecurity researchers to defend their systems from advanced threat actors.

The working operation of a honeypot mainly deals with computer applications that replicate the functioning of a system, diverse services, and pretend to viewed as the network part. When an intruder tries to log in to the system, the admin will be notified about the threat immediately and the log is generated for all the entries. The intruder becomes successful in logging to the system and stealing information but here honeypot is able to fool the intruder by providing the fake data. The intruder remains unaware about this fool act by then the attacker will be charged for legal actions by the official. So, by this, it is possible to protect our system. Researchers doubt that few cybercriminals tend to apply the concept of honeypot to collect information intelligence about researchers, fake their identity as a lure and mislead by spreading wrong information.

With respect to the design and classification, honeypots are divided into two types: Production and Research. Production honeypots are usually deployed within an organizational environment to protect the organization. They protect the system by giving regular alerts to administrators. Research honeypots are used to gather intelligence on the threats and inspect the hacker activity well in advance and learn how to prevent the systems from attackers and progress. Honeypots can also be classified as low-interaction, medium interaction or high-interaction honeypots. A low-interaction honeypot replicates only the services which are often requested by the attackers and hence they are less risky and easily maintainable. A medium-interaction honeypot involves solving more complex attacks by providing a better illusion of operation systems. A high-interaction honeypot gives practical experience to the attackers by imitating the activities of production systems and representing an ample amount of information.

2. LITERATURE SURVEY

In [15], a dynamic pin is used as a password so that it becomes difficult for the attacker to break even if he observes while the user types in the password. This system requires less memory and dynamically changes the PIN of the device. Four digits of date and time are used as a password. Different formats such as h1:h2:m1:m2, m1:m2:h1:h2, h2:m1: h1:m2 can be used based on the user's preference. The system cannot be taken down by the brute force as the PIN changes from time to time. Although this system is a good solution for shoulder surfing attack because of the dynamic PIN generation, the shoulder surfer can easily deduce the password if this method gets universally accepted.

Pass matrix [16] protects the user suffering from shoulder surfing in public places through a one-time login indicator. The login indicator which is generated randomly during each phase for pass images will be unused after the session ends. Better security is provided by the login indicator in opposition to shoulder surfing attack because a dynamic pointer is used by the user to identify the location of their password rather than selecting the password directly. In the pass matrix, a part of every image is used as a password from a sequence of n images. In this, the first square is located in the first image and second square in the second image and so on. The user chooses one grid from each image instead of choosing 'n' grid in the same image. The Cued Click Point (CCP) helps the user to remember and recall their password. If the user clicks on an improper password area within the picture the login will be failed. However, the disadvantage is the hacker can deduce the password through concealed cameras.

The randomized keyboard [12] expects the user to type in something which is incorporated with an augmented reality wearable device. The user can see the keys on the randomized keyboard through augmented reality device which is commercially feasible. Different keyboard layout is made visible to the shoulder surfer wherein he cannot deduce the actual keyboard pattern. It is important to make sure that the keystrokes done by the user cannot be easily identified by the shoulder surfer. Even if he does so, the different keyboard pattern misleads the shoulder surfer from knowing the actual password. An algorithm called Individual Key Randomization (IKR) is used to randomize the keys on the keyboard. An algorithm called Row Shifting (RS) is used to shuffle the keys row wise whereas Column Shifting (CS) is used to shuffle the keys column wise. This method overcomes the disadvantages of having a shoulder surfer peak in while the user types in the password. The above three algorithms help the user to efficiently type in the password by misleading the shoulder surfer. However, the user should always wear the augmented reality devices or glasses. [12]

This [17] technique consists of two phases namely registration and authentication phase. During the registration phase, the user is expected to enter his valid username and select images from the given set as his password. Every image is associated with three-digit code where this code has to be entered by the user to choose his image along with direction and the same has to be remembered by the user for the entire process. During the authentication phase, the user is expected to identify the password images and the random code associated with the images. However, for every authentication session the images will be randomized. This technique uses indirect selection such as choosing the image next to password image called the subordinate image. This subordinate image is decided based on the direction chosen initially during the registration process. The correct identification of the subordinate image for every password image from the given set leads to successful authentication else it directs the user to start the whole process again from the beginning. [17]

The proposed ColorPass [18] technique follows the concept of partially observed attacker model where the user can view only the response provided by the system but not the challenge values. Here, the user chooses four pin colors. In the login procedure firstly, the user has to enter his login id and then when the system authenticates the login id it will generate the feature table on the system that throws some challenge values in the range 1 to 10 to the user. The feature table can be selected depending upon the challenge values and further the color pin has to be selected depending upon the feature table that exactly indicates the colour cell. The digit in the color cell has to be identified and submitted as a response to the given challenge by the user. The login process will be completed only after responding similarly to all the other remaining three given challenges. The response given by the user will be evaluated by the system which then the system finally decides if the user is a legitimate user or not. However, this system does not work for fully observable attacker model [18].

In [21], a concept based on merging images called hybrid images is used wherein this technology simply fools the eyes of the shoulder surfer. The core idea is on the simple observation of the variation in the distance between the screen and the user with that of the screen and the shoulder surfer. The user views the screen from the lesser distance when compared to the shoulder surfer who is at least 0.9 meters away from the screen. Taking this into account, a hybrid keypad is implemented. The keypad consists of numbers with each button being the combination of two digits. The shoulder surfer is misled since the button is totally viewed differently by him with varied layouts. Consequently, the extraction of user's PIN becomes difficult. The shuffling of digits is performed in every authentication. This helps in knowing the spatial arrangement of the digits pressed. The hybrid keypad consists of two keypads. One keypad is viewed only by the user called user's keypad and the other which is visible to the shoulder surfer called shoulder surfer's keypad to confuse him. This hybrid keypad is created by using low pass and high pass filter parameters. This filtering helps in creating two images. The spatial frequency of both the keypads varies which differentiates the keypad layouts. The algorithm used called visibility algorithm helps to find the minimum safety distance from the user's keypad to the shoulder surfer. Therefore, a false PIN layout is created in order to protect the shoulder surfing attack. The disadvantages are - it's too complicated when compared to fingerprint scanner authentication scheme and the third-party apps already use a shuffling scheme which is nearly as complex as Illusion PIN concept [21].

This paper [22] presents a more secured pattern-key based password authentication system where these grid points forms the pattern and these grid points only point to the location of number in an integer matrix. The pattern key being the first level is followed by the secret key and then the dummy values at the last. During the Registration phase, user will be given a 5*5 block grid numbered from 1 to 25. Firstly, the user types in the location number of the pattern. Then in addition to the pattern, the user registers a key for numbers 0 to 9. A key value ranging from "0 to 9" maps to any integers or characters or to any special characters. Followed by this, the user needs to type in the number of dummy values in this phase where dummy values precede as well as succeed the real password values. These dummy values are named as left and right dummy values. During Login phase, after entering the username a next screen appears containing 5*5 grid block will appear with randomly generated numbers. The pattern choose in the registration phase has to be remembered by the user to map the key values to the selected password values along with left and right dummy values and then enter the password. If the password matches it authenticates the user and is able to log in successfully else fails in the login process. The disadvantages are- three secured features are time consuming and remembering a lot of key values during authentication may frustrate the user [22].

In [23], pictures are used as password as the human brain has a capacity to remember hundred images with detail. At the beginning, users are exposed to 50 images out of 70 images wherein each character are assigned to an image. The shoulder surfer cannot easily identify which character is assigned to which image. These images will be from the random art gallery. Here, the user chooses images that are difficult to relate and are colourful. So, for every 10 characters the user selects a picture that signifies a character. The user’s pass images are these 10 images. Also, the user should enter the username. The login phase takes 5 columns and 14 rows of the 70 images. The images displayed during login phase help him to recognize the password character. The account will be blocked after 3 trials. Therefore, [23] provides a secure medium for an authentication system. The disadvantages are – the system is very complicated since it consists of complex images and difficult to remember many images.

3. PROBLEM STATEMENT

Nowadays, there is an increasing usage in web services so the personal business email can be sent to the user when the user gets accessed to the user's personal information, load files and photos to albums in the cloud or cancel the transaction. Users might not be able to secure the password from the unknown while logging into these facilities. People hack on the whole authentication either by direct supervision or with the help of external devices such as video cameras or a surveillance camera or from the reflected image on the window. Once the intruder acquires the password, it is possible for them to get the personal accounts and that would eventually end in stealing one's information. The following are the problems:

1. The issue of securing password in public in order to reduce shoulder surfing.
2. The issue of creating secure password efficiently as compared to textual password.
3. The issue of searching the exact image during the login phase becomes tedious.
4. The issue of memorizing the password images at the time of authentication.
5. The issue of having finite usage in authentication applicable only to some devices.

3.1 ASSUMPTIONS

In this paper, we not only explore about the shoulder surfing and the preferences of the user and the hacker may guess the correct password but also, the following assumptions.

1. The client and the server communication are secured by SSL so that the information would not be leaked to the attacker while transmitting.
2. The authentication system between the client and server devices is reliable.
3. The screen and the keyboard of the system are difficult to protect, but an OTC which is sent to the e-mail and the coordinates selected during the registration phase can be protected.
4. Users should register their account in a secure place where there are no observers or the cameras present.

4. OVERVIEW OF HONEYPASS

HoneyPass is composed of the following components (see Figure 1):

a) Grid Formation Module
b) One-Time-Code Generation Module
c) Horizontal and Vertical Slider Control Module
d) Transmission Module
e) Password Verification Module
f) Honey Pot Module

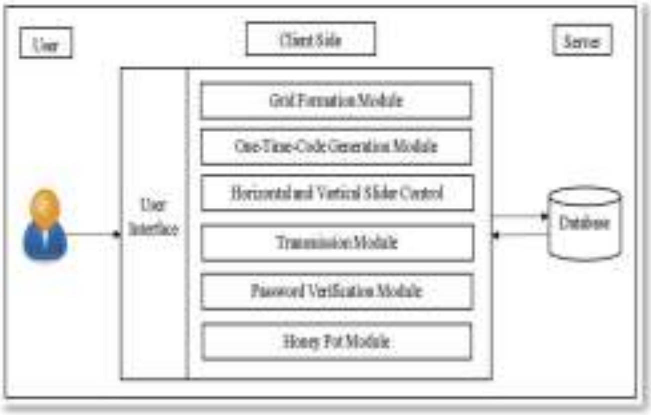
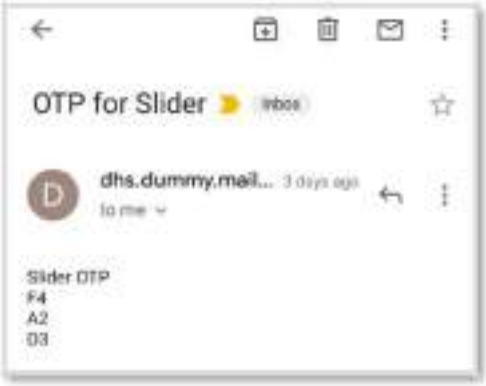


Figure 1: Overview of the HoneyPass System

Grid Formation Module: This module formulates grids by dividing each image into tiny squares. From the set of formed squares, the user gets to choose one square as a password known as pass-square. The system provides three images divided evenly into a 6x8 grid as shown in the figure. The user needs to select one pass-square from each grid. The larger the image divided, the higher the risk for a brute force attack. However, the formed squares cannot be extremely smaller that makes it difficult for the user to recognize the password image. This difficulty, in turn, might increase the complexity of the user interface. As a result, a 6x8 grid system is selected to make it more user-friendly without having to decrease the concentration of the grid.

Figure 2: Three 6x8 image grids

One-Time-Code Generation Module: This module generates One-Time-Code (OTC). The OTC consists of three alphanumeric pair. Each pair is a combination of one alphabet from A to F and one digit from 1 to 8. (For example, the three pairs can be C5, B3, E1). The system generates all three pairs randomly. The user gets the OTC through the mail. Each time the user logs in to the system, a new OTC sent to the user. The motto is to keep the password secret from the shoulder surfers by generating a new OTC each time. The pass-squares can be known if the shoulder surfer gets to know the OTC. However, gaining access to OTC is a lot harder than imagined.



OTP for Slider

dhs.dummy.mail... 3 days ago
to me

Slider OTP
F4
A2
D3

Figure 3: Obtained OTC

Horizontal and Vertical Slider Control Module: The system consists of two sliders: horizontal slider and vertical slider. As shown in the figure, the horizontal slider includes alphabets from A to F, and the vertical slider includes numbers from 1 to 8. The user can shift one character at a time with the help of arrows. The arrows help the character move in all directions, i.e., up, down, right, left. From the obtained OTC, the horizontal slider is shifted to the respective pass-square's column, and the vertical slider is shifted to the pass-square's row. Therefore, the slider implicitly leads to the user's pass-square location.

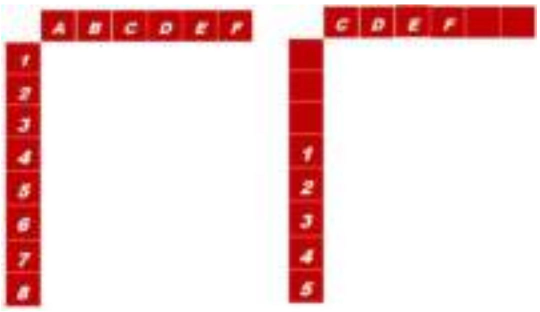
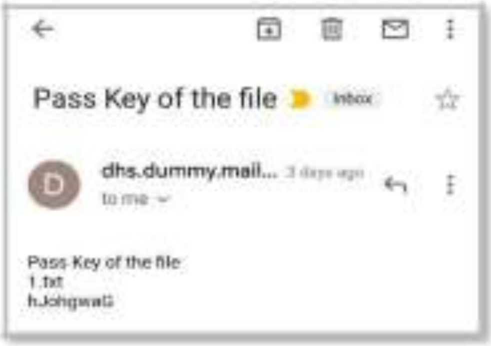


Figure 4: Before Sliding and After Sliding

Password Verification Module: This module verifies the password entered by the user in the form of slider movement. To get authenticated, it is important to align the slider's character with the pass-square. This alignment is performed with the help of obtained OTC. Once the user locates the slider character to the respective column and row, the user will be able to login to the system. The details of how the slider is aligned to the pass-square will be discussed in the next section.

Transmission Module: This module transmits information from the client to the authentication server. In this case, the pass-squares are transmitted to check for authenticity. The authentication server stores the user's pass-squares in the database. These pass-squares are matched with the ones the user submits. Once the pass-squares match, the user will be able to login to the system. SSL (Secure Socket Layer) protocol protects the exchanged information from being intercepted and stolen.

Honey Pot Module: Once the user logs in to the system, he can upload and download files from the cloud. To download any file, the user requires to enter the passkey. This passkey which is known only to him is sent to the user's mail when he uploads the file which is shown in the figure. When the hacker tries to gain access to the user's private file, the system limits the login attempts to three. However, a duplicate file is downloaded when a wrong passkey is entered after three attempts.



Pass Key of the file

dhs.dummy.mail... 3 days ago
to me

Pass Key of the file
1.txt
hJohgwat3

Figure 5: Obtained Passkey

4.1 HONEYPASS: THE PROPOSED SYSTEM

The HoneyPass's authentication consists of two phases such as registration phase and login phase. The description is as given below:

4.1.1 *Registration Phase*

In this phase, during the creation of an account, the user requires to enter all the required information such as user id, username, password, valid email-id, etc. Once all this information is submitted, three random images appear in three consecutive pages divided into 6x8 grid wherein the user has to select one coordinate image square from each page as the graphical password to get authenticated in the login phase. The three coordinates selected will be concatenated together to generate a hash code and the same will be stored in the database for reference.

4.1.2 *Authentication Phase*

During login phase, registered user logs in to the system by using his authenticated user id and password, if both matches one-time-code (OTC) will be received by the user immediately to his email id. OTC contains the random pair of horizontal and vertical slider coordinate points for all the three images. Using Coordinates received in OTC the slider has to be adjusted to match with the coordinates chosen by the user during the password setting phase. Once the hash code created while registration matches with the generated hash code, user will be successful in logging in to the system and enter in to the home page else, process ends and login page will be displayed again.

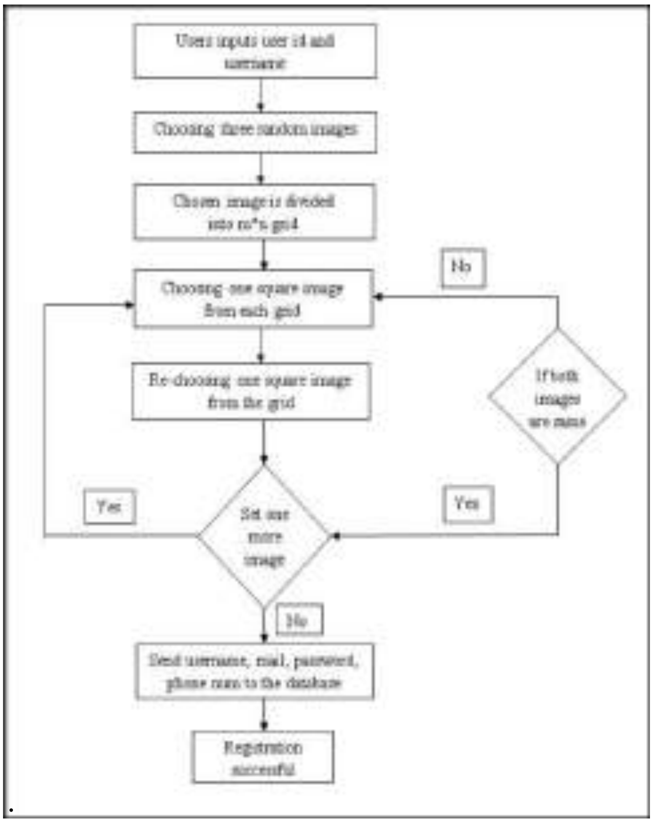

Users inputs user id and username
Choosing three random images
Chosen image is divided into m*n grid
Choosing one square image from each grid
Re-choosing one square image from the grid
If both images are same
No
Yes
Set one more image
Yes
No
Send username, mail, password, phone num to the database
Registration successful

Figure 6: Flowchart of Registration Phase.

Figure 6 is the flowchart of registration phase and the following describes the **registration** steps in detail-

1. The user inputs his credentials such as user id and user name and requests for registration.
2. Server randomly chooses three images and breaks it into an m*n grid format.
3. The user selects one square image from each grid and those coordinates have to be remembered by the user.
4. Once the pass-square has been chosen from each grid by the user, the password string is created and the hashtag is generated.
5. Hence, the registration process becomes successful and all the details about the user such as user id, user name, email id, and phone number are automatically stored in the database for future reference.
6. A confirmation message about the successful registration is displayed to the user.

Figure 7 is the flowchart of authentication phase and the following describes the **authentication** steps in detail-

1. The user needs to enter the same user id that was registered during the registration phase.
2. User requests for OTC to receive a One-Time Code to his mail.
3. Once the OTC is received, the server again chooses three random images and displays it to the user for authentication.
4. The user has to slide the coordinate points with the help of the slider and locate to the correct coordinate points using the OTC.
5. The password hence created generates a hash code.
6. The comparison is made between the currently generated hash code and the existing hash code during password setting.
7. If they both are equal user will be given an entry to the home page else error messages will be displayed.

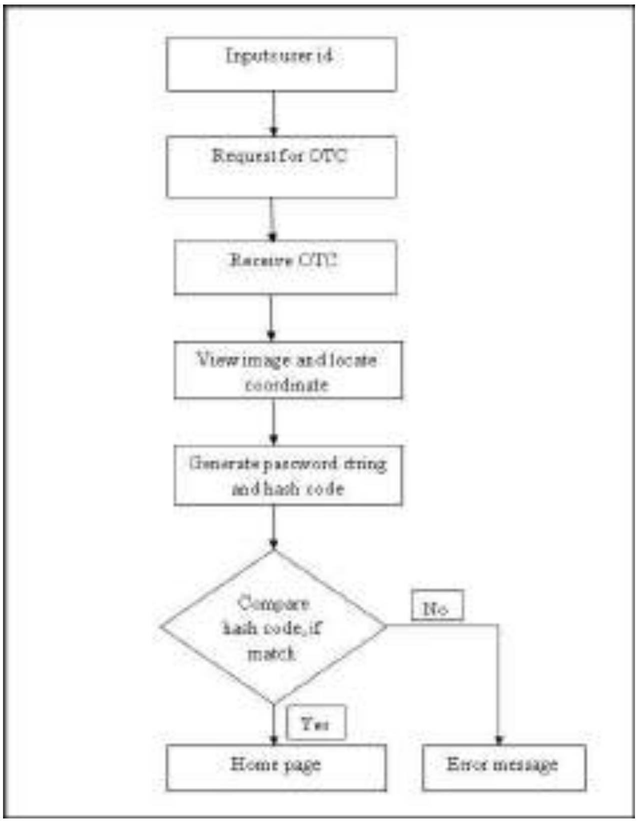
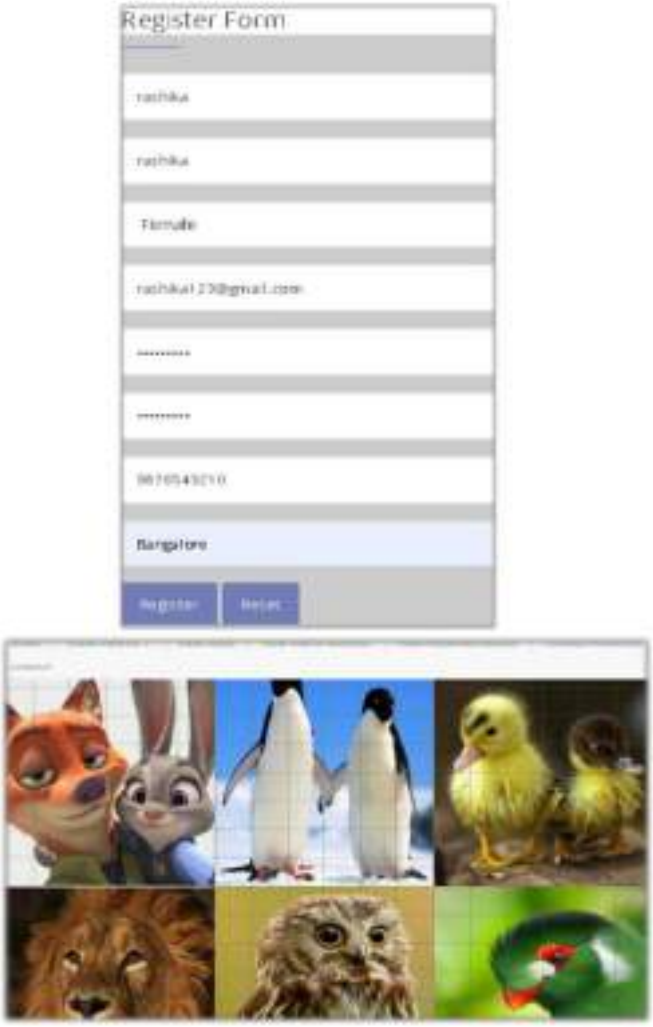

Inputs user id
Request for OTC
Receive OTC
View image and locate coordinate
Generate password string and hash code
Compare hash code, if match
No
Yes
Home page
Error message

Figure 7: Flowchart of Authentication Phase.

5. IMPLEMENTATION

The HoneyPass prototype was built using the IDE Eclipse Galileo. The client side is designed using Java and Servlet, JSP to implement the functions that check for registered users and matching password, the creation of password images, grid formation, one-time-code delivery, and the horizontal and vertical slider control system. The server side is implemented using Tomcat 7.0 and My-SQL 5.0. The driver used for JDBC connection is Type 4 - Driver. The registered users and pass-squares stored in the database are fetched during the authentication phase. The stored details help in password verification.

In our implementation, the database is uploaded with a set of predefined password images. However, the system supports users' upload action. Each image is divided into a 6x8 grid system wherein the user selects three images as a password during the registration phase. From each of the three images, one square is chosen as a password. The actions performed during the registration phase is shown in the figure 8 and figure 9.



Register Form

rashika

rashika

Female

rashika123@gmail.com

••••••••••

••••••••••

9876543210

Bangalore

Figure 8: (a) User's registration form. (b) Set of predefined password images stored in the database.

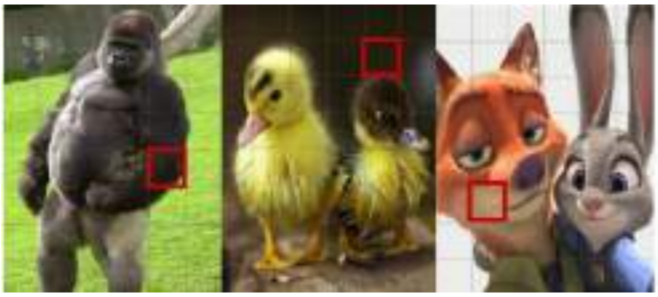
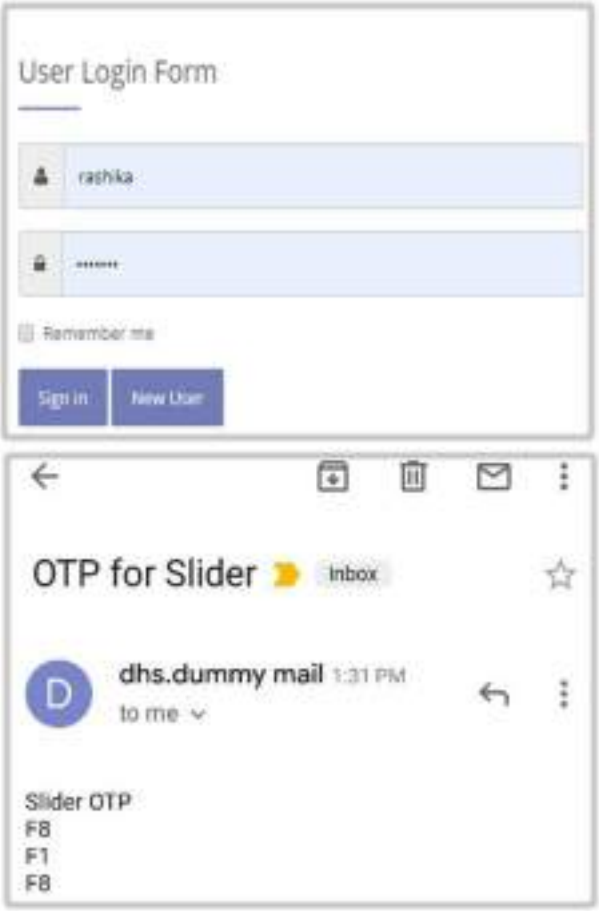


Image 1 (col=5, row=5), Image 2 (col=5, row=2) Image 3 (col=2, row=6)
Figure 9: Selected Pass-squares.

The first step in the login phase is to get the one-time-code as shown in the figure 10. This code is sent to the user's email as soon as he requests for it. To protect against shoulder surfers, the user can check the obtained OTC from his smartphone. The next step is to match the coordinates of the pass-squares to the obtained OTC. The user has to match coordinates of the first image with the first alphanumeric pair in OTC, coordinates of the second image with the second alphanumeric pair in OTC and coordinates of the third image with the third alphanumeric pair in OTC. The match actions are performed using the horizontal and the vertical slider as shown in figure 11.



User Login Form

rashika

•••••••

□ Remember me

OTP for Slider

dhs.dummy mail 1:31 PM

to me

Slider OTP
F8
F1
F8

Figure 10: (a) Login Form (b) Received OTC that has to be matched with the pass-squares coordinates

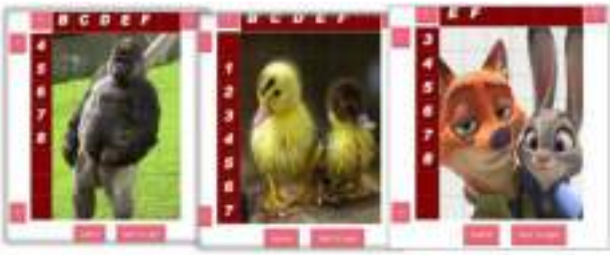
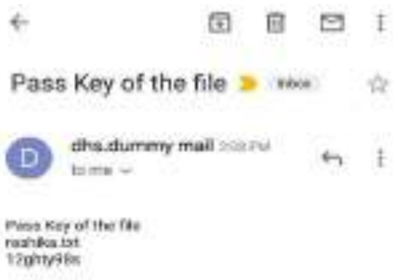


Figure 11: Match action using the horizontal and the vertical slider

Once the user logs in to the system, he can upload and download files to/from the system. To download the uploaded file, the user requires the passkey sent to his mail as shown in the figure 12. The user has to enter the received passkey to download the file (figure 13). Even though the shoulder surfer steals the OTC, he will not be able to download the file. The honeypot concept prevents the shoulder surfer from accessing the file. When he tries to gain access to the user's private file, the system limits the login attempts to three. However, a duplicate file is downloaded when a wrong passkey is entered after three attempts.



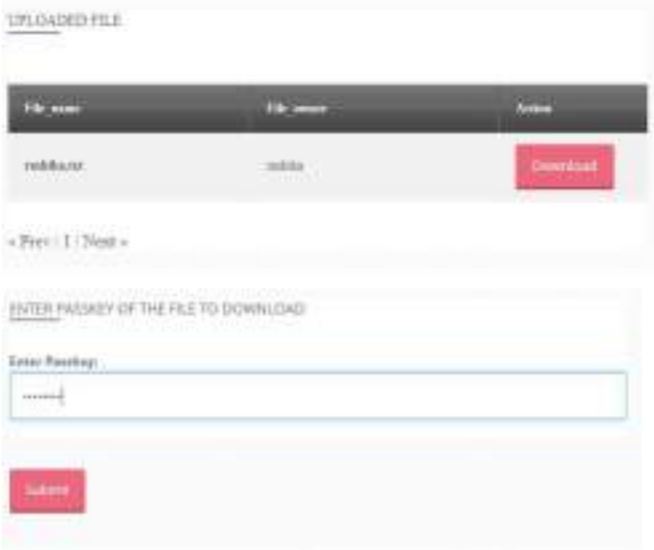
Pass Key of the file

dhs.dummy mail

to me

Pass Key of the file
rashika.txt
12ghty96s

Figure 12: Obtained Passkey



UPLOADED FILE

File_name	File_owner	Action
rashika.txt	rashika	Download

« Prev | 1 | Next »

ENTER PASSKEY OF THE FILE TO DOWNLOAD

Enter Passkey:

Figure 13: (a) Uploaded file by the user (b) Passkey is entered to download the file

6. CONCLUSION

In this paper, the cause for shoulder surfing attack and the prevention methods is put forth. An attempt that has been made to contemplate the significance of various graphical authentication systems that have been proposed over the years to overcome shoulder surfing attacks. The methods to overcome the disadvantages of textual passwords are presented. The system's advantages and disadvantages that have been surveyed are presented for each paper. The need for graphical authentication system is emphasized. Implementation of the honeypot is addressed here to secure the system from counteracting attempts of unauthorized users to steal the information. Like any other graphical authentication system, HoneyPass is also vulnerable to random guessing attacks but it is strongly resistant to any form of shoulder surfing attacks i.e. either direct observation or with the help of external devices. This approach will help various research analysts to move forward with the graphical authentication system who was unfortunate about the textual password system and their drawbacks.

7. REFERENCES

[1] Mayuri Gawandi, Saloni Pate, Pokhara Snehal, Prof.Said S.K: A Survey on Resisting Shoulder Surfing Attack Using Graphical Password. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 10, ISSN: 2278 – 1323, October 2017.

[2] Aishwarya N. Sonar, Purva D. Suryavanshi, Pratiksha R. Navarkle, Prof. Vijay N. Kukre: Survey on Graphical Password Authentication Techniques. International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 02 | Feb-2018.

[3] MalinEiband,MohamedKhamis, EmanuelvonZezschwitz, HeinrichHussmann, FlorianAlt: Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. The CHI Conference on Human Factors in Computing Systems (CHI 2017), At Denver, CO, USA,2017.

[4] K. Divyapriya, Dr.P. Prabhu: Image Based Authentication Using Illusion Pin for Shoulder Surfing Attack. International Journal of Pure and Applied Mathematics Volume 119 No. 7 2018, 835-840,2017.

[5] Choi, D., Choi, C., & Su, X: Invisible Secure Keypad Solution Resilient against Shoulder Surfing Attacks. International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS). doi:10.1109/imis.2016.77, 2017.

[6] Miss. Priyanka Nimbalkar, Miss. YashashriPachpute, Mr. Nishiket Bansode, Prof. Vaishali Bhorde: A survey on shoulder surfing resistant graphical authentication system. Open Access International Journal of Science and Engineering, Volume 2, ISSN (Online) 2456-3293, December 2017.

[7] Vijayakumari Rodda, Gangadhar Rao Kancherla, Basaveswara Rao Bobba: Shoulder-Surfing Resistant Graphical Password System for Cloud. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, pp. 6091-6096, Number 16 2017.

[8] J. Thirupathi: A Comprehensive Survey on Graphical Passwords and shoulder surfing resistant technique analysis. IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015.

[9] Dhanashree Chaudhari: A Survey on Shoulder Surfing Resistant Text Based Graphical Password Schemes. International Journal of Science and

Research (IJSR) ISSN (Online): 2319-7064, Volume 4 Issue 11, November 2015.

[10] Monali Pawar, Prof. G.S Mate, Soni Sharma, Sonam Gole, Snehal Patil: A Survey Paper on Authentication for Shoulder Surfing Resistance for Graphical Password using Cued Click Point (CCP). International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 1, January 2017.

[11] M. S. A. Noman Ranak, Saiful Azad, Nur Nadiah Hanim Binti Mohd Nor, Kamal Z. Zamil: Press touch code: A finger press-based screen size independent authentication scheme for smart devices. PLoS ONE,12(10): e0186940, October 30, 2017.

[12] Peng Foong Ho, Yvonne Hwei-Syn Kam, Mee ChinWee, Yu Nam Chong and Lip Yee Por: Preventing Shoulder-Surfing Attack with the Concept of Concealing the Password Objects' Information. Scientific World Journal, Volume 2014.

[13] Su, X, Wang, B, Zhang, X, Wang, Y, & Choi, D: User biometric information-based secure method for smart devices. Concurrency and Computation: Practice and Experience, 30(3), e4150. doi:10.1002/cpe.4150, 2017.

[14] John T. Davin, Adam J. Aviv, Flynn Wolf, Ravi Kuber: Baseline Measurements of Shoulder Surfing Analysis and Comparability for Smartphone Unlock Authentication. CHI 2017, Denver, CO, USA, May 6–11, 2017.

[15] Yogadinesh S, R. Sathishkumar, Akash L, Aakash V, Kishore Kumar K, Harichander S: Counterfeit shoulder surfing attack using random pin. International Journal of Pure and Applied Mathematics, Volume 118 No. 22, 2018.

[16] M. Kannadasan, J. Amarnadha reddy, K. Venkat Raman: Shoulder Surfing Resistant Graphical Authentication System. International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017.

[17] Swale Saeed and M Sarosh Umar: PassNeighbor: A Shoulder Surfing Resistant Scheme. International Conference on Next generation Computing Technologies Dehradun, October 2016.

[18] Gopika Anil, Chippy, Mary John, Pradeep P Mathew: Color Combo: An authentication mathod against shoulder surfing attack. International Journal of Computer Science and Information Technology Research, Vol.4, Issue 2, pp:(142-147), Month: April-June 2016.

[19] Andrew Lim Chee Yeung, Bryan Lee Weng, Wai, Cheng Hao Fung, Fiza Mughal, Vahab Iranmanesh: Graphical password:Shoulder-surfing Resistant using Falsification. 9 th Malaysian Software Engineering International Conference, Dec 2015.

[20] Shruthi V: CRASH-Cued Recall Based Authentication Resistant to Shoulder Surfing Attack. Online Internatinal Conference on Green Engineering and Technologies Kerala, 2015.

[21] Athanasios Papadopoulos, Toan Nguyen, Emre Durmus: llusionPIN: Shoulder-Surfing Resistant Authentication Using Hybrid Images. IEEE Transactions on Information Forensics and Security,2017.

[22] M Hamza Zaki, Adil Husain,m Sarosh Umar, Muneer H Khan: Secure Pattern-Key Based Password authentication Scheme. International Conference on Multimedia,Signal Processing and Communication Technologies,2017.

[23] Elham Darbanian and Gh. Dastghaiby fard,: A Graphical Password against Spyware and Shoulder Surfing Attacks. Malaysian Software Engineering International Conference, Dec,2015

Smart Parking Management System For Vehicles

Prof. Shridevi Kembhavi[1], Ibtesam Sanglikar[2], Misba Rangrez[3], Nuzhed Gulbarga[4], Sayyeda Humera Quadri[5]

Department of Computer Science & Engineering, SECAB I.E.T Vijayapura, Belgaum

shridevik.cse@secab.org[1] , ibtesamsanglikar@gmail.com[2], mkrangrej@gmail.com[3],
nuzhatgulbarga024@gmail.com[4], shumeraquadri@gmail.com[5]

Abstract*— It is difficult to find the parking space in a overcrowded area or the parking space available, during the summit hours and its always unbearable. It is common for drivers to keep circling within a parking lot for a parking space. Leaving negative influence on traffic congestion, climate changes, air pollution. Most of the existing automated parking systems are bit confusing for unfamiliar users. Every vehicle proprietor is worried about the security of the vehicle parked in the parking lot. In this paper, we try to diminish the driver's cause of stress by proposing a new Smart Parking Management System for Vehicle(SPMSV) which provides real-time information for detecting parking lots and reservation, e-payment solutions with integration of Wireless Sensor Network (WSN), Radio Frequency Identification (RFID), Adhoc Network, and Internet of Things (IoT).

Index Terms—*WSN (Wireless Sensor Network), RFID, IoT, SPMSV, ad hoc network*

I. INTRODUCTION

According to the recent reports of the United Nations (UN), 54% of the world's population live in urban areas, and this proportion is expected to raise to 66% by 2050. A direct implication is the increase of traffic congestion, which brings other issues with it, such as air pollution and parking problems. Already now, searching for an available parking spot can be considered more than a daily overhead, since it accounts for more than 30% of traffic congestion in urban areas. Currently, the common method of finding a parking space is manual where the driver usually finds a space in the street through luck and experience. This process takes time and effort and may lead to the worst case of failing to find any parking space if the driver is driving in a city with high vehicle density.

Theft of vehicles is a serious problem in large parking lot when compared to search for the space to park the vehicles, as per the statistical report more than 100000 vehicles are been stolen every year in Canada. Extents of traffic in the parking lot, inability to remember and locate parked space of vehicle are some of the major issues associated with current multilevel parking systems. and identifying whether the lot is full or not at the door-level. Solving this problem means a better service can be offered by malls and shopping complexes and the average quantity of time drained by the public in parking garages can be reduced drastically. Some of the technologies can be adopted to solve such problems using IoT, Wireless sensor network, Vehicular ad hoc network etc. Things which are connected to each other via internet are termed as "Internet of Things". Sensors, actuators, RFID tags could be the things in Internet of Things which can be supervised remotely [1].Vehicular Ad hoc Networks (VANETs), is trending in the industry due to their advance and broad usage of wireless communication technologies, as the manufacturers and telecommunication industry are trying to adapt onboard unit (OBU) device that allows vehicles to communicate using the roadside infrastructure to improve the safety for the drivers[2]. Group of sensor nodes forms a wireless sensor network which can be self-organized to establish an ad hoc network via the wireless communication module equipped on the nodes.Taking the advantages of sensing and wireless communication, wireless sensor networks have already found many civil and army applications, such as smart home, intelligent health-care, wild environmental monitoring, battle surveillance, etc [3].

Nowadays, IoT and RFID are invading our daily live and activities which are transforming assets into smart objects, allowing information exchange among them and make decisions without or with minimal human intervention. Hence to determine parking lots status regarding vacancy\occupied using inexpensive and efficient WSN Adopt RFID to identify vehicle registration numbers (cars plates numbers) and its relevant information such as parking lot No, parking period, parking fee and assigned password for security purpose [4].

In this paper, we propose a method to minimize the drivers' hassle and inconvenience, we propose a new smart vehicle parking management system using wireless sensor network.

Fig 1.1: Existing parking system

II. RELATED WORK

There are various solutions that have been proposed to address SPS issues Using the image processing technique the parking system is automated by displaying the available lot in the given parking area. Authorization card will be given to individual user, which carries the vehicle number and user details [5].

Another approach is to use wireless sensor node (combination of Ultrasonic Sensor and Wi-Fi technology) to detect the movement of the car at the entry and exit level of parking area [6]. The real-time information of the slot vacancy can be dynamically sent to android mobile application to accomplish this they have used a sensor node at the entry and exit points respectively instead of using sensor per slot. Experimental results have proven that, the system requires less cost for implementation, minimize the human intervention.

A convenient way for reservation of parking slot is to put forward an application where user can view different parking space to umpire availability of free space and book it for specific time period [7].

To overcome the disadvantages of existing vision based target parking position Fresh light stripe projection based free parking space identification method is used[8]. By evaluating the three-dimensional information, alternating points, pivot, and opposite-site reference point are identified. Results are effective to find the position during the dark conditions and also the black surface of the vehicle.

A novel solution to defeat the parking problems is proposed by providing a Mobile Application for IoT based Smart Parking System. The basic logic involved is, when vehicle enters through the gate, it is directed to a exact parking slot[9]. The Ultrasonic Sensors which are deployed on the slot detects the vehicle and send the data to the Arduino processing unit for further processing.

Cloud-based smart parking system uses a novel algorithm that increases the effectiveness of the system and also develops network architecture based on the IoT technology. Through this algorithm user can automatically find a free space at the least cost based on new performance metrics and also calculates the cost by taking into account, distance and the total number of free slots in each car park.

A survey on Smart Parking System states that Intelligent Parking Service is a part of Intelligent Transportation Systems (ITS). This study audits different Intelligent Parking Services used for parking management, parking facilities and gives an insight into the cost-effective analysis of such papers. In this survey, various systems that provide intelligent parking management services are discussed. These systems can answer the parking problems that arise due to the unavailability of a stable, modern and proficient parking system. The use of different and modern techniques such as Expert Systems, GPS based. Vehicular communication based, wireless sensor based, fuzzy based, and Vision based can reduce the parking related issues [11].

This system reduces human efforts at the parking area to a great extent such as in case of searching of free lots by the driver and providing security to the vehicle within the parking area by generating alert messages. The information of users given at the time of registration. If user tries to breach the system policies the information of users given at the time of registration is useful to trace him/her. The moment vehicle is parked IR sensor will detect the status of parking slot and updates the application [1]. A novel approach is to use Android application that will manage the number of cars that have to be parked on the parking area by automating the Parking and Un-parking of the car with the help of commands [12].With the help of application user can save his fuel time and can also book a parking slot for them from any remote site.

A new parking system called Smart Parking System (SPS) assist drivers to find vacant parking slot in parking area in stipulated time uses ultrasonic (ultrasound) sensors to detect tenancy of slot or improper parking actions.

Features of Smart Parking System include vacant parking slot detection, improper parking detection, displays availability of parking spaces, payment facilities [13].

III. PROPOSED METHODOLOGY

To overcome the problems of searching parking lot we propose smart parking management system as shown in the figure.

Cloud:

Server of IBM MQTT is hosted on cloud. Cloud acts as a database to collect all the data related to parking areas as well as end users that have access to the system. Our system maintains all the information such as time at which the vehicle was parked, time duration for parking a vehicle and it keeps a record of every user connected to the system. This system is scalable .The system is also reliable for any kind of system failure.

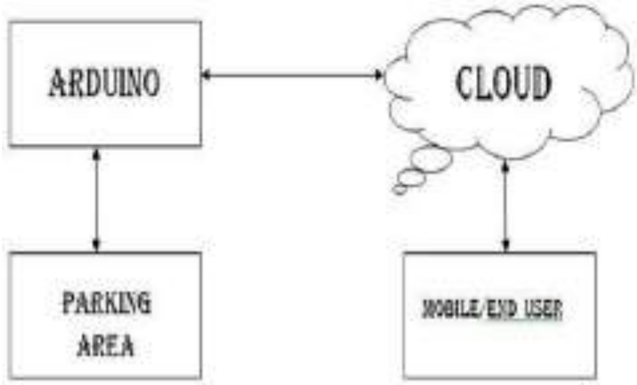


Fig 3.1: Block Diagram of Smart Parking Management System for Vehicles.

Mobile Application:

The IBM MQTT server is connected with the application through a secure channel which provides two factor authorization.The application provides information about availability of parking slots and also allows the end user to book a slot simultaneously. Apache Cordova and Angular Js framework are used to develop the application using JavaScript.

Parking Area:

Entire parking area is divided into numerous slots and each slot is equipped with sensor. These sensors are in turn connected to Arduino (micro controller) which controls the parking system and also connects to cloud server through an Internet connection to transfer data.

Arduino:

Arduino UNO ATmega328 is connected to an RFID reader. Arduino module will control the opening of the door for the vehicle and also connects to cloud server through an Internet connection to transfer data. The card reader authenticates the user information. If the information of the RFID tag or card is correct, then user is allowed to enter in parking area.

IV. IMPLEMENTATION

The user has to register with the help of application which indeed stores this information onto server hosted in the cloud so that user can be trailed in case he/she tries to breach the system policies. Application facilitates the user to view status of parking slot and allows to reserve a parking space where he can actually park his vehicle. RFID reader is present in the parking area which captures the RFID information of each vehicle. RFID tag which are present on each vehicle plays an important role in authentication, with the help of this tag, the flow diagram of the vehicle authentication by RFID reader is shown in the fig 4.

System will calculate the amount to be paid by each user for time duration the vehicle is parked in the parking lot. Arduino UNO processor controls above activities through Internet.

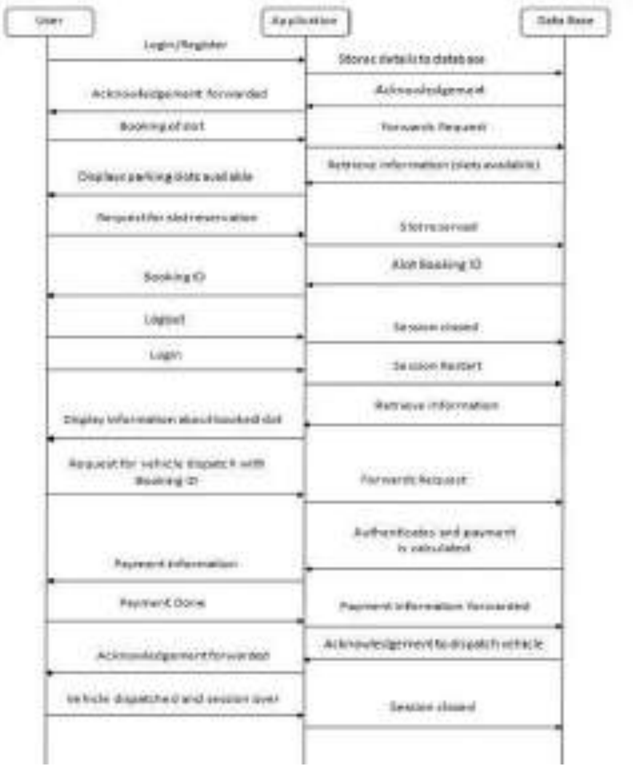

User
Application
Data Base
Login/Register
Stores details to database
Acknowledgement forwarded
Acknowledgement
Booking of slot
Forwards Request
Retrieve information (slots available)
Displays parking slots available
Request for slot reservation
Slot reserved
Alot Booking ID
Booking ID
Logout
Session closed
Login
Session Restart
Retrieve information
Display information about booked slot
Request for vehicle dispatch with Booking ID
Forwards Request
Authenticates and payment is calculated
Payment information
Payment Done
Payment information forwarded
Acknowledgement to dispatch vehicle
Acknowledgement forwarded
Vehicle dispatched and session over
Session closed

Fig 4.1: Sequence Diagram of Smart Parking Management System for Vehicles

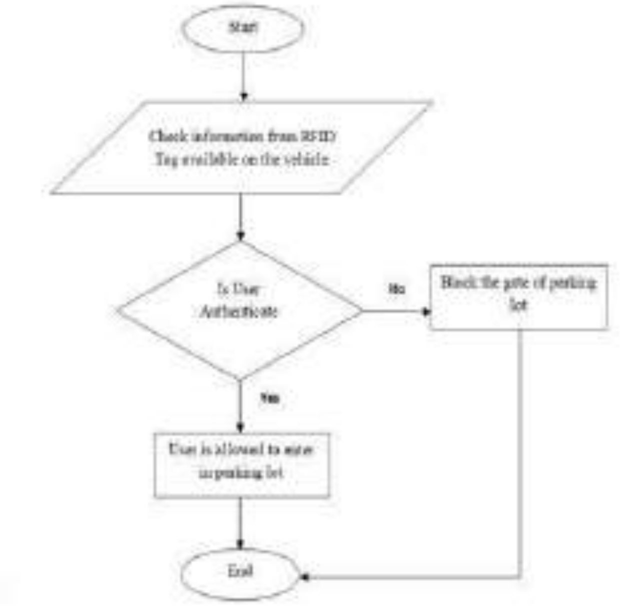

Start
Check information from RFID Tag available on the vehicle
Is User Authenticate
No
Block the gate of parking lot
Yes
User is allowed to enter in parking lot
End

Fig 4.2: Flow chart of authentication of vehicle at the entry point of parking area by RFID reader

V. RESULTS

The following are the results of Smart Parking Management system for vehicles. The Fig 5.1 shows the code of Arduino UNO interface. The application is built using Android studio.

Fig 5.1: Code of Arduino UNO interface

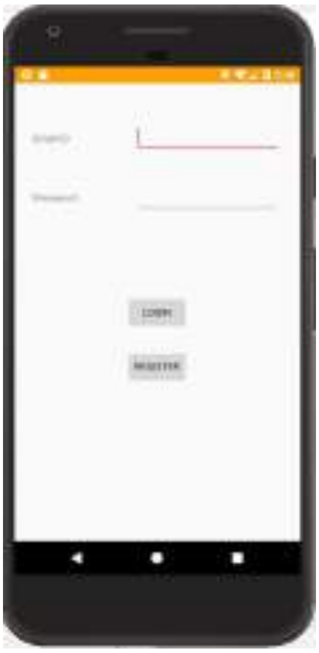


Fig 5.2: Login page

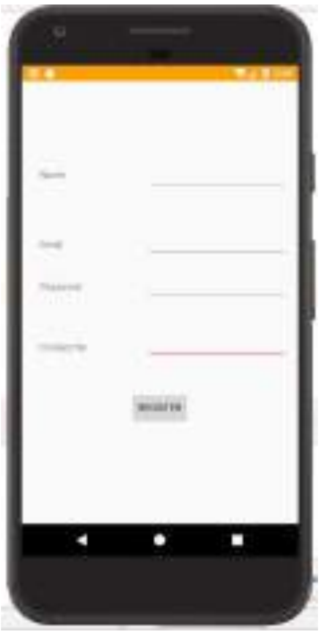


Fig 5.3: Registration page

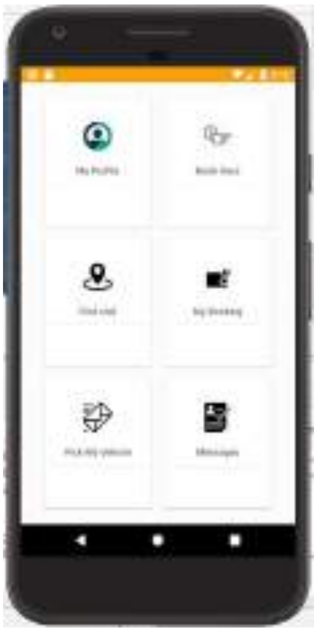


Fig 5.4: Dashboard

CONCLUSION

Our Smart Parking System for Vehicles helps in automating the parking system using android application. Also our project implements functionalities like authentication using RFID tag to attain security , guiding towards parking lot using map, e-payment options etc.

REFERENCES

[1] K.Manasa, K.Madhuvani, Iot Based Automatic Parking Sysytem Using Rfid, National Conference on Emerging Trends in Information, Management and Engineering Sciences IEEE,2018

[2] Rongxing Lu, Student Member, IEEE, Xiaodong Lin, Member, IEEE, Haojin Zhu, Member, IEEE, and Xuemin (Sherman) Shen, Fellow, IEEE, An Intelligent Secure and Privacy-Preserving Parking Scheme Through Vehicular Communications, VOL. 59, NO. 6, JULY2010

[3] Vanessa W.S. Tang, Yuan Zheng, Jiannong Cao, An Intelligent Car Park Management System based on Wireless Sensor Networks, 1st International Symposium on Pervasive Computing and Applications,2006

[4] Omar Abdulkader , Alwi M. Bamhdi , Vijey Thayananthan , Kamal Jambi, Muasaad Alrasheedi, A novel and Secure Smart Parking Management System (SPMS) based on integration of WSN, RFID, and IoT ,IEEE ,2018.

[5] Vimal A , Yashvanth R , Seetha J, a Smart Parking Bay Based on Image processing and Internet of Things, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 pp. 4423-4427,2018

[6] Venkanna U, Sanskar Sharma, Bhaumik katiyar,Yamala Prashanth, Wireless Sensor Node Based Efficient Parking Slot Availability Detection System For Smart Cities,Recent Advances on Engineering, Technology and Computational Sciences (RAETCS) IEEE, 2018

[7] Hina C. Parmar, Nisha N. Shirvi, Development of an Android Application for Smart Parking System , International Journal of Engineering Development and Research ,Volume 6, Issue 2,2018

[8] Ho Gi Jung, Dong Suk Kim, Pal Joo Yoon, and Jaihie Kim, Light Stripe Projection based Parking Space Detection for Intelligent Parking Assist System , Proceedings of the 2007 IEEE Intelligent Vehicles Symposium Istanbul, Turkey,2007

[9] Aboli Saswadkar, Chaitanya Kulkarni, Sayali Ghige, Shubham Farande, Sagar Salunke, Mobile Application for IoT based Smart Parking System , International Journal of Engineering Science and Computing, ISSN 2321 3361, April 2018

[10] Miss. Wagh Rupali, Miss. Kaklij Vaishnavi, Miss. Dagade Jayashri, Miss. Dake Pooja, , Easy And Smart Car-Parking System Using Interne-Of-Things , International Journal of Research in Engineering, Technology and Science, Vol. VIII, Issue IV, April 2018

[11] Faheem, S.A. Mahmud, G.M. Khan, M. Rahman and H. Zafar, A Survey of Intelligent Car Parking System, Vol. 11, October 2013

[12] Prof. D. J. Bonde ,Rohit S. Shende, Ketan Gaikwad, Akshay S. Kedari,Amol U. Bhokre, Automated Car Parking System Commanded by Android Application,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 3001-3004,2014

[13] Amin Kianpisheh, Norlia Mustaffa, Pakapan Limtrairut and Pantea Keikhosrokiani, Smart Parking System (SPS) Architecture Using Ultrasonic Detector, International Journal of Software Engineering and Its Applications Vol. 6, No. 3, July,2012

[14] AamirShahzad , Jae-youngChoi , NaixueXiong , Young-GabKim , Malrey Lee, Centralized Connectivity for Multi wireless Edge Computing and Cellular Platform: A Smart Vehicle Parking System ,2014

[15] Rosario Salpietro, Luca Bedogni, Marco Di Felice, Luciano Bononi, Park Here! A Smart Parking System based on Smartphones Embedded Sensors and Short Range Communication Technologies , IEEE, 2015

[16] Mr. Amit Hatekar, Harsh Bajaj, Neha Nayar, Shruti Parab, Intelligent Vehicle System , International Research Journal of Engineering and Technology (IRJET) Volume: 05 ,2018

[17] P.Mirunalini, B.Bharathi, Nirupan Ananthamurugan, Skanda Suresh, Shreyas Gopal, Multi-Level Smart Parking System , IEEE 2nd International Conference on Computer, Communication, and Signal Processing (ICCCSP),2018

Implementation of New Scheduling Algorithms in Hadoop YARN Clusters

Pavan M[1], Computer Science and Engineering, Vivekananda Institute of Technology,Bengaluru, India.
Shruthi B Gowda[2], Assistant Professor, Computer Science and Engineering, Vivekananda Institute of Technology, Bengaluru, India.
shruthi.b.gowda@gmail.com[1], pavan21manju@gmail.com[2]

***Abstract*— The data that is obtained from different media are getting stored in the form of scalable semi-structured and unstructured format, due to which the default MapReduce framework have been widely used to handle these types of data. To give better performance, Hadoop ecosystem has evolved into its second generation scheme for job scheduling that is Hadoop YARN which mainly concerns on fairness and efficiency. The current YARN does not yield effective optimal resource management, causing idle resources and ineffective scheduling. YARN doesn't support dependency between tasks as well as heterogeneous job features. By considering all the above failures in job scheduling, proposed a new YARN scheduler which can effectively reduce the makespan in Hadoop YARN clusters. For accommodating heterogeneity in MapReduce jobs, extended scheduling by further considering the job iteration information in the scheduling decisions. The implementation of new scheduling algorithm as a pluggable scheduler in YARN and evaluated it with a set of classic MapReduce benchmarks. The experimented results shows that YARN scheduler reduces the makespan and improves resource utilizations.**

Keywords: MapReduce, job scheduling, Hadoop YARN, resource management.

I. INTRODUCTION

MapReduce [1] is a framework originally designed and developed by Google to handle clusters to perform parallel computations with no database support. Hadoop is an open source implementation for MapReduce has also been adopted in both academia and industry for information analysis. The MapReduce is based on implicit parallel programming model that provides convenient way to express certain kinds of distributed computations, those that process large data sets. The Hadoop has evolved into its second generation, Hadoop YARN which consists of fine-grained resource management schemes for job scheduling. Due to the popularity of MapReduce, fairness and

efficiency become two main concerns in YARN. Current scheduling in YARN does not yield the optimal resource arrangement. MapReduce was originally used for batch data processing, it is now also being used in shared, multi-user environments in which submitted jobs may have completely different priorities. The paper aims to develop efficient scheduling schemes in YARN clusters for improving resource utilization and by reducing makespan of a given set of jobs. FIFO scheduler does not consider the optimal arrangement of cluster resources. The sequential running capability makes the FIFO scheduler for resources idleness. Fair and capacity scheduler, omit the dependency between tasks. For multiple jobs running concurrently in cluster the efficiency of resource utilization becomes crucial. Therefore, in this work a new Hadoop YARN scheduling algorithm i.e; HaSTE is presented based on task-dependency and resource-demand. The benefits of HaSTE is it efficiently utilize the resources for scheduling map/reduce tasks in Hadoop YARN and improve the makespan of MapReduce jobs. The further extension of new scheduling to dynamically determine the execution of tasks from multistage (or iterative) data processing applications. Many frameworks support multistage data processing applications such as spark, storm etc. The iterative features in the scheduling, the cluster resources cannot be efficiently utilized for executing iterative jobs, which incurs a long tail in the makespan. Therefore, a new extended version named HaSTE- A algorithm is further accommodated for heterogeneous workloads with both iterative and non- iterative jobs. HaSTE- A uses third metric (i.e. alignment) to differentiate iterative jobs from non- iterative jobs to capture the number of iterations in an application and the runtime progress of iteration jobs. HaSTE- A enforces both iterative and non- iterative jobs can be effectively reduced.

II. TECHNICAL BACKGROUND

A. MapReduce

When implementing a MapReduce program, the programmer has to implement only two functions: map(), which processes fragments of input data to produce intermediate results, and reduce(), which combines the intermediate results to produce the final output. Each map input is a key-value pair (with types defined by the programmer) that identifies a piece of work. The output of each map is an intermediate result also expressed as a key-value pair (also defined by the programmer). The reduce input is composed of all the intermediate values identified by the same key; therefore, the reduce function can combine them to form the final result. All nodes in the cluster execute the same function on different chunks of input data. The MapReduce runtime distributes and balances work across the nodes, dividing the input data into chunks, assigning a new chunk when a node becomes idle, and collecting the results. There are many runtime implementations of this model in various environments.

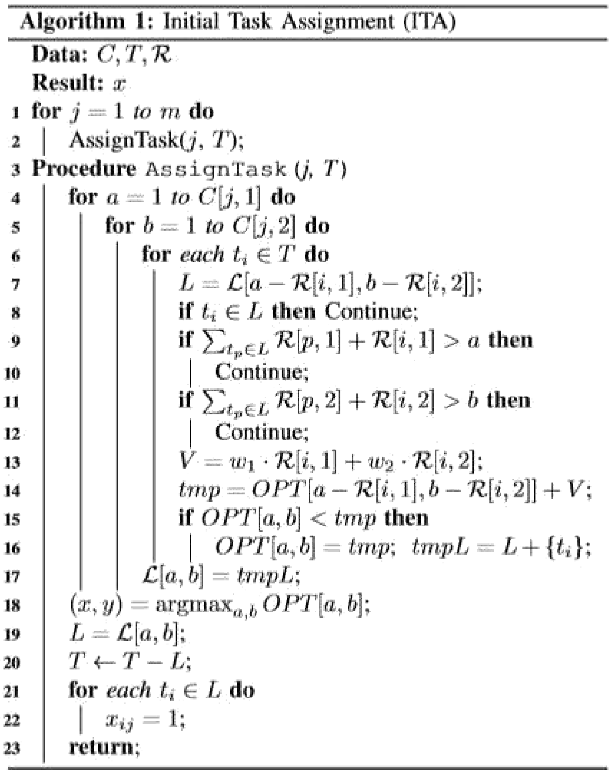
B. Hadoop

Hadoop is an open source MapReduce runtime provided by the Apache Software Foundation. It uses the Hadoop Distributed File System (HDFS) as shared storage, enabling data to be shared among distributed processes using files. The HDFS implementation has a master/slave architecture: the master process (‘NameNode’) manages the global name space and controls operations on files, while a slave process (‘DataNode’) performs operations on blocks stored locally, following instructions from the NameNode. The Hadoop runtime consists of two types of processes: ‘JobTracker’ and ‘TaskTracker’. The singleton JobTracker partitions the input data into splits using a splitting method defined by the programmer, populates a local task-queue based on the number of obtained splits, and distributes work to the TaskTrackers that in turn process the splits. If a TaskTracker becomes idle, the JobTracker picks a new task from its queue to feed it. Thus, the granularity of the splits has considerable influence on the balancing capability of the scheduler. Another consideration is the location of the data blocks, as the JobTracker tries to minimize the number of remote blocks accessed by each TaskTracker. Each TaskTracker controls the execution of tasks on a node. It receives a split descriptor from the JobTracker, divides the split data into records (through the ‘RecordReader’ component), and spawns a new worker process that actually processes all the records in the split. Such worker process will run a so-called Map task. The TaskTracker will also be in charge to run the so-called Reduce tasks as soon as they can be initiated. Notice here that a Map task will eventually result in the execution of a map() function, and that a Reduce task will behave analogously with reduce() functions. The programmer can also decide how many simultaneous map() and reduce() functions can be run concurrently on a node. When a TaskTracker finishes processing a split and is ready to receive a new one, it contacts the JobTracker. The execution of a job is divided into a Map phase and a Reduce phase. In the Map phase, the Map tasks of the job are run. Each Map task comprises the execution of the actual map() function as well as some supporting actions (for example, data sorting). The data being output by each Map task is written to a circular memory buffer. As soon as this buffer reaches a threshold, its content is sorted by key and flushed to a temporary disk file. If a Map task generates more than one such file, they are merged into a single file and then served via HTTP to nodes running Reduce tasks. During the Reduce phase, Reduce tasks are run, divided into three sub phases: data copy, key sort and finally reduce. The reduce sub-phase, which runs the actual reduce() function, is able to start after the map outputs that pertain to this particular reducer have been copied and sorted, and the final result is then written to the distributed file system. The focus of this paper is mostly put on the Map phase, which dominates the computational cost of most MapReduce applications.

C. Hadoop YARN schedulers

This section briefly introduce the scheduling process in a Hadoop YARN system and the schedulers that are currently used in YARN. It consists of multiple worker nodes and the resources that are managed by a centralized ResourceManager routine and NodeManager routines each running on worker node. The major differences between classic Hadoop system and YARN are as follows; first, unlike the jobTracker in Hadoop MapReduce, the ResourceManager no longer monitors the running status of each job. Furthermore, Hadoop YARN abandons the coarse- grained slot based resource management used in the old versions, but instead manages the system resources in a fine- grained manner. Unlike Hadoop MapReduce, YARN systems no longer explicitly distinguish map and reduce tasks such that other parallel data processing applications (such as Spark, Hive, Pig) can also be supported by YARN. The scheduling policies that are currently used in a Hadoop YARN system include FIFO, Fair, and Capacity.

- The FIFO policy sorts all waiting jobs in a non-decreasing order of their submission time. All task requests from each job will be further ordered by their priorities as well as their localities.
- Two Fair scheduling policies have been implemented in Hadoop YARN, i.e., Fair and Dominant Resource Fairness (DRF) [2]. The Fair policy only considers the memory usage of each job and attempts to assign equal share of memory to jobs, while the DRF policy aims to ensure all jobs to get on average an equal share on their dominant resource requirements (e.g., memory or cpu cores in the present YARN implementation).
- The Capacity policy works similar to the Fair policies. Under this policy, the scheduler attempts to reserve a guaranteed capacity for each job and orders these jobs by their deficit (i.e., the gap between a job’s deserved capacity and actual occupied capacity).



```
Algorithm 1: Initial Task Assignment (ITA)
   Data: $C, T, \mathcal{R}$
   Result: $x$
1  for $j = 1$ to $m$ do
2  |  AssignTask($j$, $T$);
3  Procedure AssignTask ($j$, $T$)
4  |  for $a = 1$ to $C[j,1]$ do
5  |  |  for $b = 1$ to $C[j,2]$ do
6  |  |  |  for each $t_i \in T$ do
7  |  |  |  |  $L = \mathcal{L}[a - \mathcal{R}[i,1], b - \mathcal{R}[i,2]]$;
8  |  |  |  |  if $t_i \in L$ then Continue;
9  |  |  |  |  if $\sum_{t_p \in L} \mathcal{R}[p,1] + \mathcal{R}[i,1] > a$ then
10 |  |  |  |  |  Continue;
11 |  |  |  |  if $\sum_{t_p \in L} \mathcal{R}[p,2] + \mathcal{R}[i,2] > b$ then
12 |  |  |  |  |  Continue;
13 |  |  |  |  $V = w_1 \cdot \mathcal{R}[i,1] + w_2 \cdot \mathcal{R}[i,2]$;
14 |  |  |  |  $tmp = OPT[a - \mathcal{R}[i,1], b - \mathcal{R}[i,2]] + V$;
15 |  |  |  |  if $OPT[a,b] < tmp$ then
16 |  |  |  |  |  $OPT[a,b] = tmp$;  $tmpL = L + \{t_i\}$;
17 |  |  |  $\mathcal{L}[a,b] = tmpL$;
18 |  $(x,y) = \mathrm{argmax}_{a,b} OPT[a,b]$;
19 |  $L = \mathcal{L}[a,b]$;
20 |  $T \leftarrow T - L$;
21 |  for each $t_i \in L$ do
22 |  |  $x_{ij} = 1$;
23 |  return;
```

The details are illustrated in Algorithm 1. The main algorithm is simply a loop that assigns tasks to each of the m servers (lines 1– 2). The core algorithm is implemented in the procedure AssignTask(j, T), i.e., select tasks from T to assign to server S_j . Here is a dynamic programming algorithm with two 2-dimensional matrices OPT and L, where OPT[a, b] is the maximum value of our objective function with a capacity <a, b> and L records the list of tasks that yield this optimal solution. The main loops fill all the elements in OPT and L (lines 4–17). Eventually, the algorithm finds the optimal solution (line 18) and assigns the list of tasks to S_j (lines 19–23). When filling an element in the matrixes (lines 6–17), the enumerate of all candidate tasks and based on the previously filled elements, we check: (1) if the resource capacity is sufficient to serve the task (lines 9–12); and (2) if the resulting value of the objective function is better than the current optimal value (lines 13–16). If both conditions are satisfied, we then update the matrices OPT (line 16) and L (line 17).

III. LITERATUTE SURVEY

P. Fattahi, M. S. Mehrabad, and F. Jolai, proposed Mathematical modeling and heuristic approaches to flexible job shop scheduling problems [3] in July 2007. The job shop scheduling is a branch of production scheduling, which is among the hardest combinatorial optimization problems. The job shop scheduling problem is to determine a schedule of jobs that have pre-specified operation sequences in a multi machine environment. In the classical job shop scheduling problem(JSP),n jobs are processed to completion on m un-related machines. For each job, technology constraints specify a complete, distinct routing which is fixed and known in advance. For solving the realistic case with more than two jobs, two types of approaches have been used: hierarchical approaches and integrated approaches. In hierarchical approaches assignment of operations to machines and the sequencing of operations on the resources or machines are treated separately, i.e., assignment and sequencing are considered independently, where in integrated approaches, assignment and sequencing are not differentiated. a mathematical model and heuristic approaches for flexible job shop scheduling problems (FJSP) are considered. Mathematical model is used to achieve optimal solution for small size problems. Since FJSP is NP-hard problem, two heuristics approaches involve of integrated and hierarchical approaches are developed to solve the real size problems. Six different hybrid searching structures depending on used searching approach and heuristics are presented in this paper. Numerical experiments are used to evaluate the performance of the developed algorithms. It is concluded that, the hierarchical algorithms have better performance than integrated algorithms and the algorithm which use tabu search and simulated annealing heuristics for assignment and sequencing problems consecutively is more suitable than the other algorithms. Also the numerical experiments validate the quality of the proposed algorithms.

J. Polo, D. Carrera, Y. Becerra, J. Torres, E. Ayguad´e, M. Steinder, and I. Whalley proposed Performance-driven task co-scheduling for MapReduce environments [4] in 2010. MapReduce framework with distributes the task and data across nodes. The management of MapReduce framework where the same physical resource are shared by multiple applications which consolidates the workloads in order to achieve the cost and energy savings. The proposed scheduler estimates the individual job completion for a particular given resource allocation to improve the performance goal. The main goal is to dynamically allocate resources in a cluster of machines based on the observed progress rate achieved by the jobs, and the completion time goal associated with each job. The proposed technique dynamically estimate the completion time of a job during its execution. The technique targets a highly dynamic environment [5], in which any jobs can be submitted at any time and workloads share physical resource with other workloads. Thus, the actual amount of resources available for MapReduce applications can vary over time. The dynamic scheduler uses the completion time estimate for each job given a particular resource allocation to adjust the resource allocation to all jobs.

J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S.-H. Bae, J. Qiu, and G. Fox, proposed Twister: a runtime for iterative mapreduce [7] in June 2010. MapReduce takes a more data centered approach supporting the concept of "moving computations to data". Classic parallel applications developed using message passing runtimes such as MPI[8] and PVM [9] utilize a rich set of communication and synchronization constructs offered by these runtimes to create diverse communication topologies. In contrast, MapReduce and similar high-level programming models support simple communication topologies and synchronization constructs. There are some existing implementations of MapReduce such as Hadoop and Sphere[10] most of which adopt the initial programming model and the architecture presented by Google. These architectures focus on performing single step MapReduce (computations that involve only one application of MapReduce) with better fault tolerance, and therefore store most of the data outputs to some form of file system throughout the computation. The programming model and the architecture of Twister is an enhanced MapReduce runtime that supports iterative MapReduce computations efficiently and performance comparisons of Twister with other similar runtimes such as Hadoop and DryadLINQ for large scale data parallel applications.

J. Polo, C. Castillo, D. Carrera, Y. Becerra, I. Whalley, M. Steinder, J. Torres, and E. Ayguad´e, proposed Resource-aware adaptive scheduling for mapreduce clusters [11] in 2011. The resource-aware scheduling technique for MapReduce multi-job workloads that aims at improving resource utilization across machines while observing completion time goals. Existing MapReduce schedulers define a static number of slots to represent the capacity of a cluster, creating a fixed number of execution slots per machine. The main challenge in enabling resource management in Hadoop clusters stems from the resource model adopted in MapReduce. Hadoop expresses capacity as a function of the number of tasks that can run concurrently in the system. To enable this model the concept of typed-'slot' was introduced as the schedulable unit in the system. 'Slots' are bound to a particular type of task, either reduce or map, and one task of the appropriate type is executed in each slot. The main drawback of this approach is that slots are fungible across jobs: a task (of the appropriate type) can execute in any slot, regardless of the job of which that task forms a part. The Task Scheduler is responsible for enforcing the placement decisions, and for moving the system smoothly between a placement decisions made in the last cycle to a new decision produced in the most recent cycle. The Task Scheduler schedules tasks according to the placement decision made by the Placement Controller. The Placement Controller technique leverages job profiling information to dynamically adjust the number of slots on each machine, as well as workload placement across them, to maximize the resource utilization of the cluster. In addition, our technique is guided by user-provided completion time goals for each job.

J. Wang, Y. Yao, Y. Mao, B. Sheng, and N. Mi, proposed Fresh: Fair and efficient slot configuration and scheduling for hadoop clusters [12] in 2014. The complexity that has been raised for the normal users in understanding the Hadoop system parameters and tune them appropriately is quiet difficult. Unnecessarily prolong the execution time and inefficient resource utilization while processing a batch of jobs. The proposed enhanced Hadoop system called FRESH, the best slot setting, which can dynamically configure the slots, and appropriately assign tasks to the available slots. In a classic Hadoop cluster, each job consists of multiple map and reduce tasks. The concept of "slot" is used to indicate the capacity of accommodating tasks on each node in the cluster. Specifically, the proposed new approach, "FRESH", to achieving fair and efficient slot configuration and scheduling for Hadoop clusters. The solution attempts to accomplish two major tasks: (1) decide the slot configuration, i.e., how many map/reduce slots are appropriate; and (2) assign map/reduce tasks to available slots. The targets of approach include minimizing the makespan as the major objective and meanwhile improving the fairness without degrading the makespan. FRESH includes two models, static slot configuration and dynamic slot configuration. In the first model, FRESH derives the slot configuration before launching the Hadoop cluster and uses the same setting during the execution just like the conventional Hadoop. In the second model, FRESH allows a slot to change its type after the cluster has been started. When a slot finishes its task, our solution dynamically configures the slot and assigns it the next task. The experimental results show that when serving a batch of MapReduce jobs, FRESH significantly improves the makespan as well as the fairness among jobs.

S. Tang, B.-S. Lee, and B. He, proposed Dynamic job ordering and slot configurations for mapreduce workloads [13] in 2015. MapReduce workload generally contains a set of jobs, each of which consists of multiple map tasks followed by multiple reduce tasks. Due to 1) that map tasks can only run in map slots and reduce tasks can only run in reduce slots, and 2) the general execution constraints that map tasks are executed before reduce tasks, different job execution orders and map/reduce slot configurations for a MapReduce workload have significantly different performance and system utilization. The two classes of algorithms to minimize the makespan and the total completion time for an offline MapReduce workload. Our first class of algorithms focuses on the job ordering optimization for a MapReduce workload under a given map/reduce slot configuration. In contrast, our second class of algorithms considers the scenario that the perform optimization for map/reduce slot configuration for a MapReduce workload. The target at one subset of production MapReduce workloads that consist of a set of independent jobs (e.g., each of jobs processes distinct data sets with no dependency between each other) with different approaches. For dependent jobs (i.e., MapReduce workflow), one MapReduce can only start only when its previous dependent jobs finish the computation subject to the input-output data dependency. In contrast, for independent jobs, there is an overlap computation between two jobs, i.e., when the current job completes its mapphase computation and starts its reduce-phase computation, the next job can begin to perform its map-phase computation in a pipeline processing mode by possessing the released map slots from its previous job. Having proposed job ordering algorithms that optimize the makespan and total completion time, also shows that they are stable, i.e., the optimized orders produced by job ordering algorithms do not change even if some MapReduce servers fail at execution time. Slot Configuration Optimization. Moreover, the slot configuration can have a significant impact on performance for MapReduce workloads. The performed simulations as well as experiments on Amazon EC2 and show that proposed algorithms produce results that are up to 15% „ 80% better than currently unoptimized hadoop, leading to significant reductions in running time in practice.

IV. ESISTING SYSTEM

With the growth of applications in YARN systems, more and more iterative algorithms are adopted for the MapReduce paradigm. For example, the k-means algorithm can be modeled as a set of identical MapReduce jobs such that each job's execution represents one iteration of the algorithm. Pagerank is another example of iterative algorithms, which has multiple stages in each iteration and also needs to instantiate a sequence of jobs for each iteration. The iterative feature of these algorithms determines that a single round of the map-reduce procedure is not enough for processing data. These applications often submit more than one jobs to the YARN cluster. The number of jobs for an application depends on the number of its stages as well as its input dataset. For example, the stop condition for k-means is controlled by either the pre-defined maximum number of iterations or the pre-defined convergence threshold. Observations shows that without considering the iterative feature, the current scheduling (even including HaSTE) cannot work well under the workloads with iterative applications. Two limitations can be found under those scheduling algorithms: (1) a long tail appears in the makespan due to the delayed execution of iterative algorithms, and (2) cluster resources (e.g., memory and cpu cores) cannot be fully utilized during the execution of those delayed iterative algorithms.

V. PROPOSED SYSTEM

The performance evaluation of HaSTE and HaSTE-A by conducting experiments in a Hadoop YARN cluster. Implemented HaSTE, HaSTE-A and FFD-DotProduct (abbrev. FFD-DP) schedulers in Hadoop YARN version 2.2.0 and compared them with the built-in schedulers (i.e., FIFO, Fair, Capacity, and DRF). The performance metrics considered in the evaluation include makespans of a batch of MapReduce jobs and resource usage of the Hadoop YARN cluster. For HaSTE-A, average response time is additional metric is considered. In the experiment, the consideration of different resource requirements such that a job can be either memory intensive or cpu intensive. The resource requirements of map and reduce tasks of a MapReduce job can be specified by the user when that job is submitted. The user should set the resource requirements equal to or slightly more than the actual resource demands. Observed results shows that all the conventional schedulers (i.e., FIFO, Fair, and DRF) cannot efficiently utilize the system resources, e.g., under 60% cpu core usage and under 30% memory usage. Although these conventional schedulers obtain similar resource usage, FIFO outperforms Fair by 23.8% and DRF by 29.3%. That is because under Fair and DRF when multiple jobs are running concurrently in the cluster, their reduce tasks are launched and thus occupy most of the resources, which may dramatically delay the execution of map phases. Similarly, the makespan under the FFD-DP scheduling policy is 10% larger than under FIFO, although FFD-DP achieves the highest resource usage, e.g., 86.6% cpu cores usage in average. While, the new scheduler HaSTE solves this problem by considering the impacts of both resource requirements (i.e., fitness) and dependency between tasks (i.e., urgency) and thus achieves the best makespan, which is, for example, 27% and 44.6% shorter than FIFO and Fair, respectively.

VI. CONCLUSION

The two scheduling policies, named HaSTE and HaSTE- A, which is implemented in Hadoop YARN v.2.2.0 reduced the makespan of a given set of MapReduce jobs. Based on each task's fitness and urgency, HaSTE dynamically schedules tasks for execution when resources become available. By further considering each task's alignment, our extended scheduler HaSTE-A effectively addresses the long tail issue caused by iterative jobs. The experimental results demonstrated that HaSTE and HaSTE-A improve the performance in terms of makespan under different workloads.

ACKNOWLEDGMENT

I am very thankful to the Staff members who have directly or indirectly contributed for completion of this Paper. I am very much grateful to the entire CSE department of Vivekananda Institute of Technology, Bangalore for giving me all facilities and work environment which enable me to complete my task. I am also thankful to all researchers which I have mentioned in reference list.

REFERENCES

[1] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," in OSDI '04: Sixth Symposium on Operating System Design and Implementation, San

Francisco, CA, December 2004, pp. 137–150. [Online]. Available: http://labs.google.com/papers/mapreduce.html

[2] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica, “Dominant resource fairness: fair allocation of multiple resource types,” in USENIX NSDI, 2011.

[3] P. Fattahi, M. S. Mehrabad, and F. Jolai, “Mathematical modeling and heuristic approaches to flexible job shop scheduling problems,” Journal of intelligent manufacturing, vol. 18, no. 3, pp. 331–342, 2007.

[4] J. Polo, D. Carrera, Y. Becerra, J. Torres, E. Ayguad´e, M. Steinder, and I. Whalley, “Performance-driven task co-scheduling for mapreduce environments,” in Network Operations and Management Symposium (NOMS), 2010 IEEE. IEEE, 2010, pp. 373–380.

[5] D.Carrera, M.Steinder, I.Whalley, J.Torres, and E.Ayguad´e, “Enabling resource sharing between transactional and batch workloads using dynamic application placement,” in Middleware ’08: Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, 2008, pp. 203–222.

[6] M. Yazdani, M. Amiri, and M. Zandieh, “Flexible job-shop scheduling with parallel variable neighborhood search algorithm,” Expert Systems with Applications, vol. 37, no. 1, pp. 678–687, 2010.

[7] J. Ekanayake, H. Li, B. Zhang, T. Gunarathne, S.-H. Bae, J. Qiu, and G. Fox, “Twister: a runtime for iterative mapreduce,” in Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. ACM, 2010, pp. 810–818.

[8] MPI (Message Passing Interface). Available: http://wwunix.mcs.anl.gov/mpi/

[9] PVM (Parallel Virtual Machine). Available: http://www.csm.ornl.gov/pvm

[10] Y. Gu and R. L. Grossman, "Sector and Sphere: the design and implementation of a high-performance data cloud," Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, vol. 367, pp. 2429-2445, 2009.

[11] J. Polo, C. Castillo, D. Carrera, Y. Becerra, I. Whalley, M. Steinder, J. Torres, and E. Ayguad´e, “Resource-aware adaptive scheduling for mapreduce clusters,” in Middleware 2011. Springer, 2011, pp. 187–207.

[12] J. Wang, Y. Yao, Y. Mao, B. Sheng, and N. Mi, “Fresh: Fair and efficient slot configuration and scheduling for hadoop clusters,” in Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on. IEEE, 2014, pp. 761–768.

[13] S. Tang, B.-S. Lee, and B. He, “Dynamic job ordering and slot configurations for mapreduce workloads,” 2013.

Face Spoofing Attack Detection Using LBP and LTP Methods

Ms. Anusha P S[1], Department of CSE, J N N College of Engineering, Shimoga, Karnataka, INDIA
Dr. Sanjeev R Kunte[2], Professor, Department of CSE, J N N College of Engineering, Shimoga, Karnataka, INDIA
sanjeevkunte@jnnce.ac.in[1], anushaps96@gmail.com[2]

***Abstract* – The face of a human being conveys a lot of information about identity and emotional state of the person. Face recognition is an interesting and challenging problem, and impacts important applications in many areas such as identification for law enforcement, authentication for banking and security system access, and personal identification among others. Today's, biometric systems are vulnerable to spoof attacks made by non-real faces. The problem is when a person shows in front of camera a print photo or a picture from cell phone. This proposed system is an anti-spoofing solution for distinguishing between 'real' and 'spoof' faces. In this approach overlapping block LBP and LTP operator is used to extract features in each region of the image with different radius. Finally, Chi-Square histogram distance is used to determine whether the input image corresponds to a real face or not. Experimental analysis on a publically available NUAA face anti spoofing database following the standard protocols showed good results.**

Keywords – Face Recognition, Face Spoofing Attacks, Local Binary Pattern (LBP), Local Ternary Pattern (LTP), Chi-Square histogram distance and NUAA dataset.

I.INTRODUCTION

A facial recognition system is a technology capable of identifying or verifying a person from a digital image or a video frame from a video source [1]. There are multiple methods in which facial recognition systems work, but in general, they work by comparing selected facial features from given image with faces within a database. It is also described as a Biometric Artificial Intelligence based application that can uniquely identify a person by analyzing patterns based on the person's facial textures and shape. Face recognition systems are vulnerable to spoofing attacks with printed photos or replayed videos [2]. The recent availability of richer imaging sensors is opening new possibilities for designing improved face recognition and spoofing attack detection solutions. A spoofing attack occurs when a person tries to masquerade as someone else by falsifying data and thereby gaining illegitimate access and advantages.

Face spoofing is a form of attack that is, presenting a fake sample to the acquisition sensor with facial information of a valid user including showing photographs, video, 3D facial model of a valid user etc. Face spoofing attacks may be image-based or video-based. In-order to detect the spoofing attack, there are lots of methods are available that detect whether a biometric sample is original or not. These methods include frequency-based approaches [3], texture-based approaches and motion-based approaches. During the capturing process of synthetic biometric data, there are noise information and artifacts are present such as blurring effect, printing artifacts and banding effects. These noise information and artifacts are enough to determine the spoofing attacks.

In general, there are three possible ways to generate a face spoof attack, they are:

- Generating a photograph of a valid user
- Reproducing a video of a valid user
- Presenting a 3D reproduction of the face of a valid user.

Typical countermeasure against spoofing is liveness detection that aims at detecting physiological signs of life such as eye blinking, facial expression changes, mouth movements etc [4]. Another existing countermeasure to spoofing attacks consists of combining face recognition with other biometric modalities such as gait and speech. Indeed, multimodal systems are intrinsically more difficult to spoof than uni-modal systems. As illustrated in Figure 1, face images captured from printed photos can look very similar to face images captured from real faces. From these characteristics of an image is captured and compared with the image which is stored in the database and by classifying it is recognized as a real face or a fake face.

The organization of the rest of the paper is as follows: In this paper, we reviewed other techniques related to face spoofing attacks in section II. In Section III we present the methodology used to determine whether the input image has a real or spoofed face using both LBP and LTP methods. Section IV experimental results are then introduced and discussed. Finally, Section V includes conclusion of our proposed system in this paper.

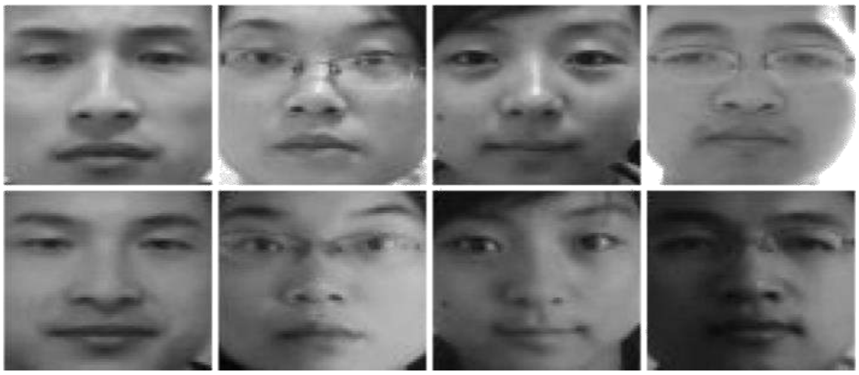


Figure 1: Example of images captured from real faces (upper row) and from Printed photos (lower row).

II. RELATED WORK

There are many approaches implemented in face spoofing detection. The existing methods for face spoofing detection can be classified into four groups: user behaviour modeling, user co-operation, methods that require additional software and hardware and methods based on data-driven characterization. Several approaches related to the face recognition system and detection of face spoofing detection methodologies are briefly explained in this section.

Tronci et al. [5] proposed a method based on the motion information and clues that are take out from the scene by combining two types of processes, referred to as static and video-based analysis. The static analysis consists of combining different visual features such as color, edge, and Gabor textures. The video-based analysis combines simple motion-related measures such as eye blink, mouth movement, and facial expression change. The static analysis is used to find the abnormalities related to the input samples at verification process. Fusion was carried out at score level by using a weighted sum. Photo detection gives a higher weight in combination. Movement measures contribute only very little weight. This performs both video and static analysis in order to employ complementary information about motion, texture and liveness and consequently to obtain a more robust classification.

TopiM aenpaa *et al.* [6] presents two novel ways of extending the local binary pattern (LBP) texture analysis operator to multiple scales. First, large-scale texture patterns are detected by combining exponentially growing circular neighborhoods with Gaussian low-pass filtering. Second, cellular automata are proposed as a way of compactly encoding arbitrarily large circular neighborhoods. The performance of the extensions is evaluated in classifying natural textures from the database. The operator works by thresholding a 3×3 neighborhood with the value of the center pixel, thus forming a local binary pattern, which is interpreted as a binary number. The joint distribution of LBP codes and cellular automaton rules proved to be too sparse to be statistically reliable, even when infrequently occurring entries were removed. Statistical reliability seems to be the key issue in using the distributions of cellular automaton rules.

Zhenhua Guo *et al.* [7] explained about Local binary pattern (LBP), fast and simple for implementation, has shown its superiority in face and palm-print recognition. To extract representative features, “uniform” LBP was proposed and its effectiveness has been validated. However, all “non-uniform” patterns are clustered into one pattern, so a lot of useful information is lost. In this study, the authors propose to build a hierarchical multi-scale LBP histogram for an image. Local binary pattern (LBP), fast and simple for implementation, has shown its superiority in face and palm-print recognition. The useful information of “non-uniform” patterns at large scale is dug out from its counterpart of small scale. The performance of single LBP operator is limited. Multi-scale or multi-resolution could represent more image feature under different settings. Traditionally, LBP features of different scale are extracted first, and then the histograms are concatenated into a long feature. Joint distribution could contain more information, but it suffers from huge feature dimension.

Yogesh Raja *et al.* [8] explained about pair wise-coupled reformulation of LBP-type classification which involves selecting single-point features for each pair of classes across multiple scales to form compact, contextually-relevant multi-scale predicates known as Multi-scale Selected Local Binary Features (MSLBF), and a novel binary feature selection procedure, known as Binary Histogram Intersection Minimization (BHIM) designed to choose features with minimal redundancy. Multi-scale Selected Local Binary Features (MSLBF) predicates are used in a pair wise-coupling approach with multiple binary classifier, one for each pair of classes, along with a scoring procedure to perform multiclass discrimination. Each classifier is a joint density generated from individual bit features selected from across scales in the training data.

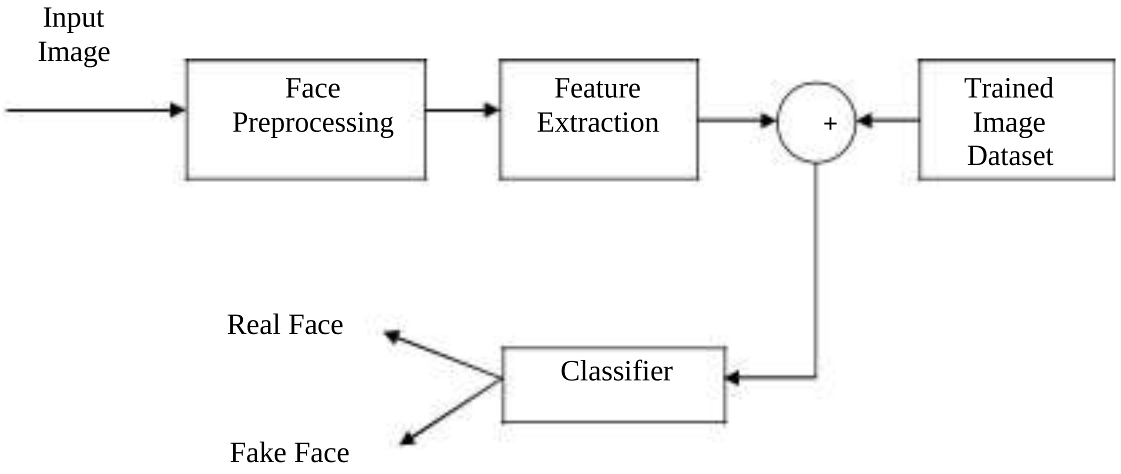


Figure 2: The Proposed Approach

III. SPOOFING DETECTION USING LBP AND LTP

This approach of anti-spoofing used to differentiate between live faces and fake ones in photographs. The process of person identification starts with input image by using face recognition can be split into four main phases and ends with the result as real face or spoofed face. They are face preprocessing, feature extraction, trained image dataset and classification. The block diagram of our anti-spoofing approach can be seen in figure 2.

A. Face Preprocessing: The preprocessing of face includes conversion of an image to gray, filtering of a gray image and detection of a face image using Viola – Jones algorithm. The Viola-Jones algorithm is one of the most popular and exploited methods in face detection history. The main components of this face detection framework are integral imaging, Adaboosting and cascading. A new representation of the image called the integral image [9].

B. Feature Extraction: There exist several methods for extracting the most useful features from (preprocessed) face images to perform face recognition. The feature extraction methods used in this proposed system are Local Binary Pattern (LBP) and Local Ternary Pattern (LTP).

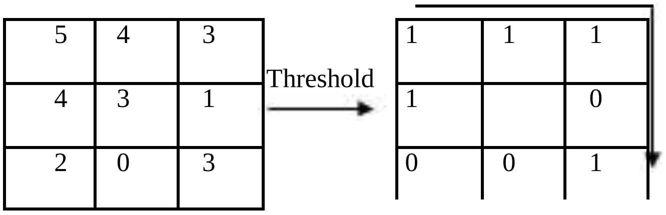
1) Local Binary Pattern (LBP): The LBP is an image operator which transforms an image into an array or image with more detail. The basic LBP introduced by Ojala et al. [10], was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. The original LBP works in a 3x3 pixel block of image. The pixels in this block are threshold by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighborhood consists of 8 pixels, a total of 28=256 different labels can be obtained depending on the relative gray values of the center and its neighborhood as shown in Figure 3.

The LBP(P,R) operator used a circular neighborhood. The notation (P, R) is generally used for pixel neighborhoods to refer to sampling points and circle of radius. So the calculation of the LBP(P,R) codes can be easily done. The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R}=\sum_{P=0}^{P-1} s(gp-gc)2^{p} \qquad (1)$$

Where gc corresponds to the gray value of the center pixel(xc, yc), gp refers to gray values of P equally spaced pixels on a circle of radius R , and s defines a thresholding function as follows:

$$s(x)=\begin{cases} 1 & if\, x \geq 0 \\ 0 & otherwise \end{cases} \qquad (2)$$



5	4	3
4	3	1
2	0	3

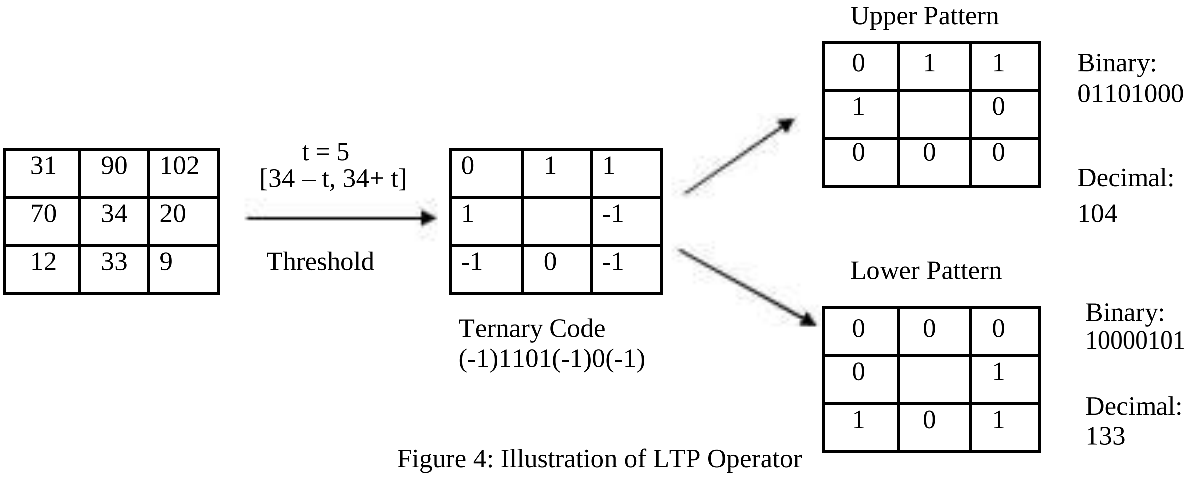
Threshold

1	1	1
1		0
0	0	1

Binary: 111011001 Decimal: 233

Figure 3: Illustration of LBP Operator

(2) Local Ternary Pattern (LTP): LTP are an extension of local binary patterns (LBP) [11]. Unlike LBP, it does not threshold the pixels into 0 and 1; rather it uses a threshold constant to threshold pixels into three values. In this way, each threshold pixel has one of the three values. Neighboring pixels are combined after thresholding into a ternary pattern. Computing a histogram of these ternary values will result in a large range, so the ternary values will result in a large range, so the ternary pattern is split into two binary patterns [12]. Histograms are concatenated to generate a descriptor double the size of LBP.



31	90	102
70	34	20
12	33	9

0	1	1
1		-1
-1	0	-1

0	1	1
1		0
0	0	0

0	0	0
0		1
1	0	1

Figure 4: Illustration of LTP Operator

Considering k as the threshold constant, c as the value of the center pixel, a neighboring pixel p, the result of threshold is:

$$s(x) = \begin{cases} 1, & \text{if } p > c + k \\ 0, & \text{if } p > c - k \text{ and } p < c + k \\ -1, & \text{if } p < c - k \end{cases} \qquad (3)$$

Here's an illustration of LTP shown in figure 4, given a 3 x 3 image patch and a threshold t. The range assign a pixel in a window to 0 is when the threshold is between c - t and c + t, where c is the centre intensity of the pixel. Therefore, because the intensity is 34 in the centre of this window, the range is between [29, 39]. Any values that are beyond 39 get assigned 1 and any values that are below 29 get assigned -1. Once determine the ternary codes, split up the codes into upper and lower patterns. Basically, any values that get assigned a -1 get assigned 0 for upper patterns and any values that get assigned a -1 get assigned 1 for lower patterns. Also, for the lower pattern, any values that are 1 from the original window get mapped to 0. The final pattern is reading the bit pattern starting from the east location with respect to the centre (row 2, column 3), then going around counter-clockwise. Therefore, probably modify function so that outputting both lower patterns and upper patterns in image.

C. Trained Image Database: The NUAA spoofing face database [13] which plays an important role in static face liveness detection and is available to the publically published in 2010, and both the images of real client and imposter attacks are included. Each individual face images is collected in three different sessions of which each is held every two weeks and the environment and lighting conditions are different for each session. The NUAA database use traditional webcams whose resolution is 680*480 to obtain 15 persons images, and each person are captured almost 500 images. Only nine out of fifteen objects present in the training set under the live human circumstance and only three out of nine objects present under the photos. Thus we can know that there is such a big difference between persos present in test and training sets. The training set contains 3099 images, the test set contains 2623 images and does not overlap with the training set to form a database.

D. Classification: System consists of database that contains predefined patterns that compares with detected object to classify in to proper category. Face spoofing attacks are most likely performed by displaying the targeted faces using prints, video displays or masks to the input sensor. The most crude attack attempts performed, e.g. using small mobile phone displays or prints with strong artifacts, can be detected by analysing the texture and the quality of the captured gray-scale face images [14][15][16]. A model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of independent normally distributed data, which is valid in many cases due to the central limit theorem. Chi-square histogram distance is one of the distance measures that can be used to find dissimilarity between two histograms. The effect is demonstrated by measuring the similarity between the local binary pattern (LBP) descriptions extracted from a genuine face with another face of the same person. The similarity is measured using the Chi-square distance:

$$d_x^2(H_x, H_y) = \sum_{i=1}^{N} \frac{(H_x(i) - H_y(i))^2}{H_x(i) + H_y(i)} \qquad (4)$$

where Hx and Hy are two LBP histograms with N bins. In addition to its simplicity, the Chi-square distance is shown to be effective to measure the similarity between two LBP or LTP histograms. Chi-square distance does not necessarily indicate that there are no intrinsic disparities in the gray-scale texture representation that could be exploited for face spoofing detection. More specifically, we computed mean LBP histograms for both real and fake face images in the training set and used these two models to compute a Chi-square distance based score value for each sample in the test set as follows:

$$d(H_x, H_r, H_f) = d\chi 2(H_x, H_r) - d\chi 2(H_x, H_f) \quad (5)$$

where H_x is the LBP or LTP histogram of the test sample, and H_r and H_f are the reference histograms for real and fake faces, respectively. Classifier is first trained using a set of positive (real faces) and negative (fake faces) samples from the dataset.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experiments are carried by considering real face images and spoofed face images of print attack from publicly available *NUAA* database. Around 1000 images from 8 different persons are considered for experimentation. The system was trained with around 600 images. The developed system is tested using another set of images which were not included in the training. Experiments are carried out by extracting *LBP* and *LTP* features and obtained results are analyzed based on comparison between trained dataset and testing dataset for both real images and spoofed images. The performance of the developed system is evaluated using recognition rate as the parameter using the following equation:

$$\textit{Percentage of Recognition rate} = \frac{\textit{Total number of images correctly}}{\textit{Total number of images tested}} \quad (6)$$

For all images total number of LBP and LTP features extracted is 5900. It has been observed from Table I has overall recognition rate of both the feature extraction methods showing LTP has better results than the LBP.

Table I: Results of images whose Block size of 16 x 16

Feature Extraction Method	Number of test images		Number of correct recognition		Overall Recognition Rate (%)
	Real	Spoof	Real	Spoof	
Local Binary Pattern (LBP)	150	100	132	83	86
Local Ternary Pattern (LTP)	150	100	147	97	97.6

V. CONCLUSION

In this work, an approach for anti-spoofing detection using LBP and LTP is presented that discriminate live faces from fake ones. Only print attacks are considered in the system. The face region in the given image is detected first with good accuracy. LBP and LTP features are extracted from the face region. A chi-square histogram distance is used for classifying the image as a real face or a spoofed face. The system is tested for both real and spoofed images. The developed system exhibits a good recognition rate of 86 % for LBP and around 97.6% for LTP. Comparatively feature extraction using LTP method gives a good result than the LBP method.

REFERENCES

[1] J. Määttä, A. Hadid, and M. Pietikäinen, “Face spoofing detection from single images using micro-texture analysis,” in Proc. IEEE Int. Joint Conf. Biometrics, Oct. 2011, pp. 1–7.

[2] A. Silva Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, “Video-based face spoofing detection through visual rhythm analysis,” in Proc. 25th SIBGRAPI Conf. Graph., Patterns Images, Aug. 2012, pp. 221–228.

[3] T.-W. Lee, G.-H. Ju, H.-S. Liu, and Y.-S. Wu, “Liveness detection using frequency entropy of image sequences,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., May 2013, pp. 2367–2370.

[4] Samarth Bharadwaj, Tejas I. Dhamecha, MayankVatsa and Richa Singh, “Computationally Efficient Face Spoofing Detection with Motion Magnification”, in Proc. IJETAE Int. vol. 3. Dec. 2013, pp. 627-633.

[5] R. Tronci et al., “Fusion of multiple clues for photo attack detection in face recognition systems,” in Proc. IEEE Int. Joint Conf. Biometrics, Oct. 2011, pp. 1–6.

[6] Topi M¨aenp¨a¨a, Matti Pietik¨ainen, “Multi-Scale Binary Patterns for Texture Analysis”, in Machine Vision Group, Infotech Oulu University of Oulu. Dec 2013.

[7] Zhenhua Guo, Lei Zhang, David Zhang, “Hierarchical Multiscale Lbp For Face And Palmprint Recognition”, in Proceedings of 2010 IEEE 17th International Conference on Image Processing. September 26-29, 2010.

[8] Yogesh Raja, Shaogang Gong, “Sparse Multiscale Local Binary Patterns”, University of London Mile End Road, London E1 4NS, U.K. December 2013.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in Proceedings of Computer Vision and Pattern Recognition, Vol. 1, pp. 1-9. 2001.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns" in proceedings of Pattern Analysis and Machine Intelligence, IEEE, vol. 24, no. 7, pp. 971–987, 2002.

[11] Renuka Patnaik, Raj Gupta, "Local Ternary Patterns and Maximum Bipartite Matching for Face Recognition", Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2011.

[12] Vasudha, Deepti Kakkar, "Facial Expression Recognition with LDPP & LTP using Deep Belief Network", 5th International Conference on Signal Processing and Integrated Networks (SPIN), 2018.

[13] Tan X, Li Y, Liu J, et al. ,"Face liveness detection from a single image with sparse low rank bilinear discriminative model", in Proceedings of European Conference on Computer Vision(ECCV), Springer-Verlag, pp. 504-517, 2010.

[14] Aruni Singh, Sanjay Kumar Singh, "Effect Of Face Tampering On Face Recognition", in Signal & Image Processing : An International Journal (SIPIJ) Vol.4, No.4, August 2013.

[15] Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid, "Face Spoofing Detection Using Colour Texture Analysis", in IEEE Transactions on Information Forensics and Security. August 2016.

[16] Hoai Phuong Nguyen, Florent Retraint, "Face spoofing attack detection based on the behavior of noises", IEEE 978-1-5090-4545-7/16, 2016.

An Efficient Solution For Establishing Stem Education And Ict Environment For Smart Schools

Ashwini S S[1], Cercilia T[2]

[1] *Assistant Professor, Information Science Engineering, Nagarjuna College of Engineering and Technology, Karnataka, Bangalore, India*

[2] *Student, Information Science Engineering, Nagarjuna College of Engineering and Technology, Karnataka, Bangalore, India*

---****---

***Abstract*-The smart school is a technology-based teaching learning institution for preparing children for the Information Age. To achieve smarts schools educational objectives, these teaching and learning concepts should be covered: curricular, pedagogy, assessment and teaching-learning materials, STEM education. Information and communication technology (ICT), as second pillar of smart school, plays many roles in a smart school, from facilitating teaching and learning activities to assisting with school management. For instance, some of technologies which can equip a smart school might be classrooms with multimedia courseware and presentation facilities, computer laboratory for teaching, multimedia development centre and server room equipped to handle applications, management databases, and web servers. Although in recent years some efforts have been done for developing smart schools, there is not a pre-defined and an efficient solution for establishing ICT environment for smart schools. The main objective of this paper is, it describes learning in digital age and gains increase attention. A smart school makes children smart and will help them grow with the technology.**

***Keywords*- Smart schools, E-Learning, STEM education, ICT.**

I. INTRODUCTION

The development of new technologies enables children to learn more effectively, efficiently, flexibly and comfortably. Children will utilize smart devices to access digital resources through wireless network and to immerse in both personalized and seamless learning

Engaging primary school students in Science, Technology, Engineering and Maths (STEM) learning is difficult, due to the often abstract notions and concepts involved[2]. One common alternative proposed to improve engagement and learning about such subjects, is to involve students in scientific inquiries, in which students are involved in formulating hypotheses and gathering and analyzing real data. Very often, this gathering of data is done outdoors, using increasingly available mobile and sensing technologies. However, the application of these approaches in authentic setting conditions faces simple but quite important constraints in terms of timing and effort.

In our view, one of the important programs for using information technology and communication (ICT) is smart school[4]. This reality was investigated through some observation in ministry of education. Although there are few efforts to implement smart school in private and public sector, it was realized these efforts uses different information technology infrastructure (platform), ideas and teaching material. In order to encourage for successful development of smart school, an ICT infrastructure framework should be developed.

With our interest in the goals for students in Vision 2020, we are beginning to look at what it takes to educate our children for the world of the future and what skills that will need to acquire to become productive citizens[3]. It is interesting to explore the wide range of skills in communication, critical thinking, and even problem solving that the world of work would ask educators to consider when planning curriculum, as well as the advanced technical skills associated with the information society we are going to become. To make this shift, the education system under the guidance of the National Philosophy of Education, must undergo a radical transformation. The schooling culture must be transformed from one that is memory-based to one that is informed, thinking, creative and caring. One way to make this happen is through the use of leading edge technology.

II. SMART SCHOOLS

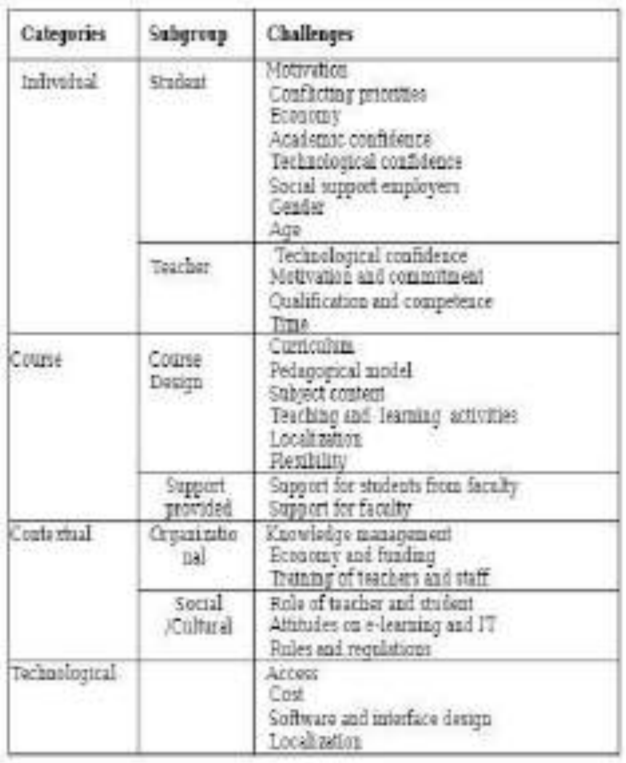
The smart school is a technology-based teaching-learning institution for preparing children with the effective functionality of smart school which requires skilled staff, and well-designed teaching, learning and supporting processes[1]. It encourages active thinking process while its' environment motivates students to use personal computers (PCs), the internet and intranets as research and communication tools. Students are able to access online libraries, use electronic mail (e-mail) or combination of desktop video-conferencing and chat rooms for doing tutorials.

The idea of the smart school is defined to revolutionize the education system through development of a holistic approach that concerns on making value based education available to anyone, anytime and anywhere. Implementing smart schools successfully will be a complex task, requiring changes teaching-learning processes; management functions; people, skills and responsibilities; and technology.

III. E-LEARNING

E-Learning can be described as the use of ICT in learning process. Various tools and technologies including e-mail, internet, video streaming and virtual classrooms can be applied for this purpose[4]. For example, one of the concerns of e-Learning in context of a learner is to connect him/her to a network in order to access course materials. This also will supported be by other tools like course management system and virtual classrooms. Andersson and Grönlund had analyzed several related papers regard to e-learning activities in different developing countries and finally they developed a conceptual framework for e-learning as shown in Table 1.

Table 1: Conceptual Framework for Challenges of E-learning
(Andersson & Grönlund)



Categories	Subgroup	Challenges
Individual	Student	Motivation Conflicting priorities Economy Academic confidence Technological confidence Social support employers Gender Age
	Teacher	Technological confidence Motivation and commitment Qualification and competence Time
Course	Course Design	Curriculum Pedagogical model Subject content Teaching and learning activities Localization Flexibility
	Support provided	Support for students from faculty Support for faculty
Contextual	Organizational	Knowledge management Economy and funding Training of teachers and staff
	Social /Cultural	Role of teacher and student Attitudes on e-learning and IT Rules and regulations
Technological		Access Cost Software and interface design Localization

Developing countries will benefit e-learning if they think for innovative ways to deliver online content on the national backbone.

IV. STEM EDUCATION

Engaging primary school students in Science, Technology, Engineering and Maths (STEM) learning is difficult, due to the often abstract notions and concepts involved[2]. One common alternative proposed to improve engagement and learning about such subjects is to involve students in scientific inquiries, in which students are involved in formulating hypotheses and gathering and analyzing real data. Very often, this gathering of data is done outdoors, using increasingly available mobile and sensing technologies. However, the application of these approaches in authentic setting conditions faces simple but quite important constraints in terms of timing and effort (e.g., logistics of such data gathering trips, matching between weather and curriculum sequence constraints, etc.).

At Tallinn University, they are starting a project that takes a different perspective on this problem of student engagement in STEM and its constraints:

instead of (or, in addition to) "taking students to the data", the Smart School project aims to "bring the data to students", while still keeping it

authentic and relevant to them. The general idea of the project is to support the next generation of scientists and engineers by having them learn in a data-rich school environment.

V. ICT

Information and communication technology (ICT), as second pillar of smart school, plays many roles in a smart school, from facilitating teaching and learning activities to assisting with school management[4].For instance, some of technologies which can equip a smart school might be classrooms with multimedia courseware and presentation facilities, computer laboratory for teaching, multimedia development centre and server room equipped to handle applications, management databases, and web servers.

Schools use a diverse set of ICT tools to communicate, create, disseminate, store, and manage information[5].In some contexts, ICT has also become integral to the teaching-learning interaction, through such approaches as replacing chalkboards with interactive digital whiteboards, using students' own smartphones or other devices for learning during class time, and the "flipped classroom" model where students watch lectures at home on the computer and use classroom time for more interactive exercises.

When teachers are digitally literate and trained to use ICT, these approaches can lead to higher order thinking skills, provide creative and individualized options for students to express their understandings, and leave students better prepared to deal with ongoing technological change in society and the workplace.

ICT issues planners must consider include: considering the total cost-benefit equation, supplying and maintaining the requisite infrastructure, and ensuring investments are matched with teacher support and other policies aimed at effective ICT use. Although in recent years some efforts have been done for developing smart schools, there is not a pre-defined and an efficient solution for establishing ICT environment for smart schools.

VI. ADVANTAGES

A. *Easy access to information*

This is perhaps the biggest benefit of a smart classroom[6]. This is because such a classroom uses many gadgets that have internet connection such as mobile, laptops, and tabs. Thanks to all these devices it is easier for students to gain access to information on the internet. Now, students have all the information that they need about any kind of topic they want to know about. It is basically at their fingertips. In fact, this is beneficial for the teachers as well. This way, they are able to learn outside the syllabus as well and this only enriches them further.

B. *Taking notes on a digital medium*

One of the biggest features of a smart school is that students can take their notes by using digital devices such as pens and tabs. This way they are able to save a significant amount of their learning time as well. No longer do they need to carry all those heavy notebooks and textbooks to class. Technology such as Google Docs has also made it a lot easier for the teachers to come up with documents and presentations. This can add major value to their students' education.

C. *A better understanding of topics through digital tools*

If students learn to topics in a better way, chances are that a smart class would be a much better bet for the same. The learning sessions at these classes normally use many visual aids. This includes PowerPoint presentations, word documents, audio sessions, and video screenings, to name a few. They play a major role in conveying the lessons in a way that is a lot more understandable. At times, a simple picture is a good enough substitute for thousands of words.

D. A great option for students who are absent

One thing with smart classes is that they are always being recorded. This means that even if you were to miss a couple of sessions you need not to feel bad. This is because there would always be the video recordings to fall back on in case you need them. You can go through the videos and also email the teachers in case you happen to have any doubt about the same. In a traditional classroom, the students always find it hard to copy down all the information that the teacher is providing them.

E. Learning in a dynamic manner

It is common knowledge that not all students in a class have the same power to grasp things. As far as the weak students are concerned they always find it hard in the traditional classrooms. But, the smart classrooms – by being as dynamic as they are – have made it easier for these students to learn at a pace that they are comfortable with. Mention needs to be made of the visual effects concepts that are being used in these classes. These effects make the entire lesson a lot clearer and this benefits these weak students in particular.

F. A teaching environment that is interactive

The digital tools are now an integral part of these smart classes. This is why the teaching environment over here can be called interactive in the truest sense of the term. It is one where both students and teachers have definite roles to play. This kind of learning makes the entire process a lot more transparent. It is a lot better than what it would otherwise have been. This improves the interaction between the students and the teachers in a significant way as well. It also improves the bond between these individuals as well. This is because they are able to communicate outside the school with the help of emails and messages.

G. Paper being replaced by digital tools

Each year students use tons of paper in order to take notes. These days, there is a lot of emphasis on being environment-friendly and this is the reason why most of the businesses are going paperless as well. In fact, by facilitating note taking on digital instruments it is now actually possible for the environment to be saved in a way. They are creating a way to replace usage and wastage of paper. In such a concept there is no place for photocopies and printouts, and this is helping reduce the carbon footprint as well.

H. They are easy to maintain

In traditional classrooms, students have to spend a lot of money each year to buy the necessary educational items such as books and pens to name a few. Even as they go up these costs increase as well. However, with smart classes, such expenses can be kept to the bare minimum. In these classes, all you need to do in this regard is to make a onetime investment in the first year and buy those electronic gadgets.

VII. CONCLUSION

With the help of the technology available in these smart classes, students actually have the chance to learn from the experts of various subjects[6]. This is one facility that is not available in traditional classrooms where it is the same teachers who take care of various subjects and this is something that does not change at all. Apart from all these, smart classes provide students more learning opportunities, makes the process of learning a fun-filled one, help them learn new technology, allow them the option to collaborate with others and learn, and their grades improve as well.

REFERENCES

[1].Zhi-Ting Zhu, Ming-Hua Yu and Peter Riezebos. A research framework of smart

education [2016] https://doi.org/10.1186/s40561-016-0026-2

[2].Luis P. Prieto1 , Mar´ıa Jes´us Rodr´ıguez-Triana, Marge Kusmin , and Mart Laanpere. Smart School Multimodal Dataset and Challenges[2017]

[3].Mohammed Sani Ibrahima *, Ahmad Zabidi Abdul Razaka , Husaina Banu Kenayathullaa. Smart principals and smart schools[2013]

[4].Siavash Omidinia, Maslin Masrom, Harihodin Selamat. An Examination of the Concept of Smart School: An Innovation to Address Sustainability [2nd International Conference on Advances in Computer Science and Engineering (CSE 2013)]

[5].https://learningportal.iiep.unesco.org/en/issue-briefs/improve-learning/curriculum-and-materials/information-and-communication-technology-ict

[6].https://fedena.com/blog/2019/02/what-is-smart-school-and-what-are-their-benefits.html

Adopting Block Chain Technology In The Distribution Of Databases On E-Voting Systems

Ms.Prabhavathi K[1],Mrs.NagaveniNimbal[2]

Assistant professor[1,2]

Computer Science and Engineering

Amruta Institute of Engg& Management Sciences

Bidadi, Ramanagara District

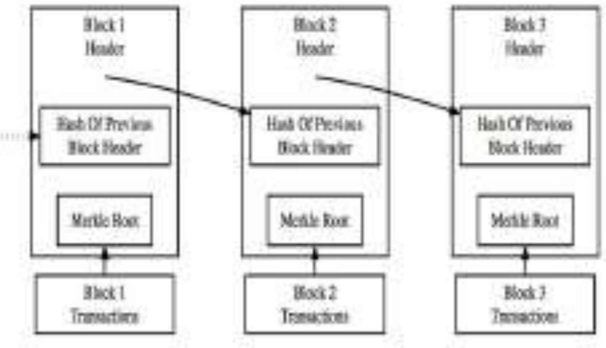
prabhavathi138@gmail[1].comnagaveninimbal29@gmail.com[2]

Abstract

E-Voting is the key public sectors that can be disrupted by blockchain technology. The idea in blockchain-enabled e-voting (BEV) is simple. To use a digital-currency analogy, BEV issues each voter a "wallet" containing a user credential. Each voter gets a single "coin" representing one opportunity to vote. Casting a vote transfers the voter's coin to a candidate's wallet. A voter can spend his or her coin only once. However, voters can change their vote before a preset deadline. In this project, we prove that blockchains might address two of the most prevalent concerns in voting today: voter access and voter fraud. The idea is as follows. Eligible voters cast a ballot anonymously using a computer or smartphone. BEV employs an encrypted key and tamperproof personal IDs. For example, the mobile e - voting platform of BostonbasedstartupVoatz employs smart biometrics and real-time ID verification. The public ledger ties each cast ballot to an individual voter and establishes a permanent, immutable record. No bad actor can engage in nefarious activities because such activities will be evident on the ledger or corrected by a peer-to-peer consensus network. To compromise the network, hackers would need to successfully hack most of the blocks (files with transaction records) before new blocks were introduced. The blockchain's audit trail ensures that no vote has been changed or removed and that no fraudulent and illegitimate votes have been added.Put simply, blockchains enable the creation of tamper-proof audit trails for voting. In this paper, we propose some BEV implementations and the approach's potential benefits and challenges.

1. Introduction

Voting, whether traditional ballet based or electronic voting (e-voting), is what modern democracies are built upon. In recent years voter apathy has been increasing, especially among the younger computer/tech savvy generation. E-voting is pushed as a potential solution to attract young voters. For a robust e-voting scheme, a number of functional and security requirements are specified including transparency, accuracy, auditability, system and data integrity, secrecy/privacy, availability, and distribution of authority. Blockchain technology is supported by a distributed network consisting of a large number of interconnected nodes. Each of these nodes have their own copy of the distributed ledger that contains the full history of all transactions the network has processed. There is no single authority that controls the network. If the majority of the nodes agree, they accept a transaction. This network allows users to remain anonymous. A basic analysis of the blockchain technology (including smart contracts) suggests that it is a suitable basis for e-voting and, moreover, it could have the potential to make e-voting more acceptable and reliable. There are number of papers that have explored this idea including now this one.Obvious advantages of e-voting using blockchains includes: i) greater transparency due to open and distributed ledgers, ii) inherent anonymity , iii) security and reliability (especially against Denial of Service Attacks) and iv) immutability (strong integrity for the voting scheme and individual votes). Existing works explore how blockchains can be used to improve the evoting schemes or provide somestrong guarantees of the above listed requirements.



2.Literature Approach

[1] "SELES: an e-voting system for medium scale online election" C. Garcia-Zamora ; F. Rodriguez-Henriquez ; D. Ortiz-Arroyo Sixth Mexican International Conference on Computer Science (ENC'05)

[2] "Building a reliable e-voting system: functional requirements and legal constraints" C. Lambrinoudakis ; D. Gritzalis ; S. Katsikas Proceedings. 13th International Workshop on Database and Expert Systems Applications.There are several types of Blockchain

PermissionlessBlockchain, like Bitcoin orEthereum, all can be a user or run a node, anyone can "write", and anyone can participate in a consensus in determining the state's validity.

Permission Blockchain inversely proportional to the previous type, operated by known entities such as consortium blockchains, where consortium members or stakeholders in a particular business context operate a Blockchain permission network. This Blockchain permission system has means to identify nodes that can control and update data together, and often has ways to control who can issue transactions.

Private blockchain is a special blockchain permitted by one entity, where there is only one domain trust.

3.Proposed work

In this paper, we incorporate Block chain technology as the solution for the problems seen in the traditional voting systems, because it embraces a decentralized system and the entire database are owned by many users. Blockchain itself has been used in the Bitcoin system known as the decentralized Bank system. By adopting blockchain in the distribution of databases on e-voting systems can reduce one of the cheating sources of database manipulation. Advantages: This technique introduces the decentralized system and thus the entire database are owned by many users. Cheating sources of the database manipulation is reduced. Simple and feasible solution. The results of the election will be available on the go and will be accurate and cheating-free.The motivation behind the proposed e-voting protocol, is to have a blockchain based scheme that meets the above stated goals. In addition to those properties the protocol must allow for a voter to change one’s mind and cancel one’s vote, replacing it with another. As a secondary goal, it has been actively pursued to provide the maximum degree of decentralisation and to create a protocol which the voters control as a network of peers. After careful consideration, however, it was decided that a certain degree of centralisation is necessary to reach the primary goal. This is because when using the blockchain, one is unable to store secret information in the public ledger without the use of external oracles that maintain such information. So if the identity of the voters is to remain secret, whilst at the same time permitting only eligible voters to participate in the elections, a Central Authority needs to be introduced that acts as a trusted third party.The main reason for using the blockchain in an e-voting protocol is to take advantage of the fact that it enables a group of people to maintain a public database, that is owned, updated, and maintained by every user, but controlled by no one. Since the protocol is based on the blockchain, it will be realised as a network of peers. Each voter will be a peer i.e. a node in a network of equals. Every voter will be responsible for making sure that fraudulent votes are rejected, hence that consensus is maintained according to the election rules. The blockchain also has the additional advantage of being increasingly well-known and well-trusted to operate as intended, as evidenced by the sheer size of the cryptocurrency market.

E-VOTING

Voting mechanisms using electronic means, or ’e-voting mechanisms’, to aid casting and counting votes have been studied in both the commercial and the academic world. In order for an e-voting protocol to be deemed secure certain formally-stated properties must hold.

• Fairness: No early results should be obtainable before the end of the voting process; this provides the assurance that the remaining voters will not be influenced in their vote.

• Eligibility: This property states that only eligible voters should be allowed to cast their vote and they should do so only once.

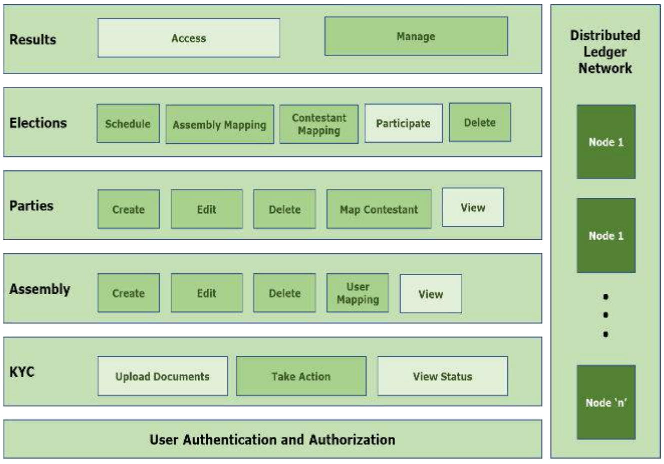
since voters need to prove their identity before being deemed eligible or not.

• Privacy: The way that an individual voter voted should not be revealed to anyone. This property in non-electronic voting schemes is ensured by physically protecting the voter from prying eyes.

• Verifiability: This property guaranties that all parties involved have the ability to check whether their votes have been counted or not. Typically two forms of verifiability are defined, individual and universal verifiability. Individual verifiability gives an individual voter the ability to verify that one’s vote has been counted. Universal verifiability requires that anyone can verify that the election outcome is the one published.

• Coercion-resistance: A coarser should not have the ability to distinguish whether a coerced voter voted the way they were instructed to.

4.System Architecture



User Authentication & Authorization: User should register or login to the application.

KYC Process: Here KYC going to check whether the given user information is right or wrong.

Assembly Component: Here an admin can perform Create, Edit and Delete the assembly. User can see the assembly information.

Parties Management: Here an admin can add the contestant with respect to party. Here both user and admin can view.

Election Process: Here the actual implementation of block chain take place.After every user cast their vote everything is gathered in form of blocks.

Result view service: Admin is goung to unveil the result after the allotted schedule is close. and both user and admin can view the result.

5. SHA-256 Algortithm

SHA-256 operates in the manner of MD4, MD5, and SHA-1: The message to be hashed is first (1) padded with its length in suchaw ay that the result is a multiple of 512 bits long, and then (2) parsed into 512-bit message blocks M(1);M(2);:::;M(N).

The message blocks are processed one at a time: Beginning with a fixed initial hash value H(0), sequentially compute H(i) = H(i1) + CM(i)(H(i1)); where C is the SHA-256 compression function and + means word-wise mod 232 addition. H(N) is the hash of M.

Description of SHA-256

The SHA-256 compression function operates on a 512-bit message block and a 256bit intermediate hash value. The initial hash value H(0) is the following sequence of 32-bit words (which are obtained by taking the fractional parts of the square roots of the first eight primes):

H(0) 1 = 6a09e667

H(0) 2 = bb67ae85

H(0) 3 = 3c6ef372

H(0) 4 = a54ff53a

H(0) 5 = 510e527f

H(0) 6 = 9b05688c

H(0) 7 = 1f83d9ab

H(0) 8 = 5be0cd19

We will use following notation:

^	bitwise AND
_	bitwise OR
:	bitwise complement
+	mod 232 addition
Rn	right shift by n bits
Sn	right rotation by n bits

Table1: Notations

Preprocessing

Computation of the hash of a message begins by preparing the message:

1. Pad the message in the usual way: Suppose the length of the message M, in bits, is `. Append the bit \1" to the end of the message, and then k zero bits, where k is the smallest non-negative solution to the equation `+1+k 448 mod 512. To this append the 64-bit block which is equal to the number ` written in binary.For example, the (8-bit ASCII) message \abc" has length 8*3 = 24 so it is padded with a one, then 448(24+1) = 423 zero bits, and then its length to become the 512-bit padded message

01100001 01100010 01100011 1 $\underbrace{00\cdots0}_{423}\ \underbrace{00\cdots011000}_{64}.$

The length of the padded message should now be a multiple of 512 bits.

2. Parse the message into N 512-bit blocks M(1);M(2);:::;M(N). The first 32 bits of message block i are denoted M(i) 0 , the next 32 bits are M(i) 1 , and so on up to M(i) 15 .We use the big-endian convention throughout, so within each 32-bit word, the left-most bit is stored in the most significant bit position.

Main loop

The hash computation proceeds as follows:

For = I to N (N = number of blocks in the padded message)

Initialize registers a, b, c, d, e, f, g, h with the (i — I) St intermediate hash value (=the initial hash value when i — 1)

$$a \leftarrow H_1^{(i-1)}$$
$$b \leftarrow H_2^{(i-1)}$$
$$\vdots$$
$$h \leftarrow H_8^{(i-1)}$$

- Apply the

SHA-256 compression function to update registers a, b, ..., h •For J — O to 63 Compute Ch(e, f, g), Maj(a, b, c), Eo(a), El(e), and wJ (see Definitions below)

$$T_1 \leftarrow h + \Sigma_1(e) + Ch(e, f$$
$$T_2 \leftarrow \Sigma_0(a) + Maj(a, b, c)$$
$$h \leftarrow g$$
$$g \leftarrow f$$
$$f \leftarrow e$$
$$e \leftarrow d + T_1$$
$$d \leftarrow c$$
$$c \leftarrow b$$
$$b \leftarrow a$$
$$a \leftarrow T_1 + T_2$$

Compute the ith intermediate hash value H(i) H(i) 1 a + H(i1) 1 H(i) 2 b + H(i1) 2 . . . H(i) 8 h + H(i1) 8H(N) =(H(N) 1 ;H(N) 2 ;:::;H(N) 8) is the hash of M.

Definition

Six logical functions are used in SHA-256. Each of these functions operates on 32-bit words and produces a 32-bit word as output. Each function is defined as follows:

Ch(x, y, z)

Maj(x, y, z)

Eo(x) = s^2 (x) s ^{13} (x) s^{22} (x)

$\Sigma_1(x)$= S^C (X) S ^{ll}(x) S^{25} (x)

01 x) = S^7 (x) S ^{IS} (x) R^3 (x)

CI(.T) = S^{IT} (x) S ^{19} (.T) R

6. THE ADVANCED ENCRYPTION STANDARD ALGORITHM (AES)

AES is a symmetric block cipher that can encrypt and decrypt information. The AES is capable of cryptographic keys of 128,192 or 256 bits. Other input, output, cipher key lengths are not permitted by this standard. For a long period of time, the Data Encryption Standard (DES) was considered a standard for the symmetric key encryption. DES has a key length of 56 bits [2]. For the time being, this key length is considered small and can easily be broken. For this reason, the National Institute of Standards and Technology (NIST) announced as a result of computation among 15 algorithms that the Rijndael cipher will replace the DES cipher and will become a new AES [16]. The Rijndael cipher has three possible block and key lengths: 128, 192, or 256 bits. Therefore, the problem of breaking the key becomes more difficult. In general, hardware implementations of encryption algorithms and their associated key schedules are physically secure, as they cannot easily be modified by an outside attacker. The basic block diagram of encryption module is shown in below fig. The decryption function is similar to that of the encryption function except that the keys have to be read in reverse order, they must be calculated prior to applying any input, therefore they are stored in a stack like buffer [9].

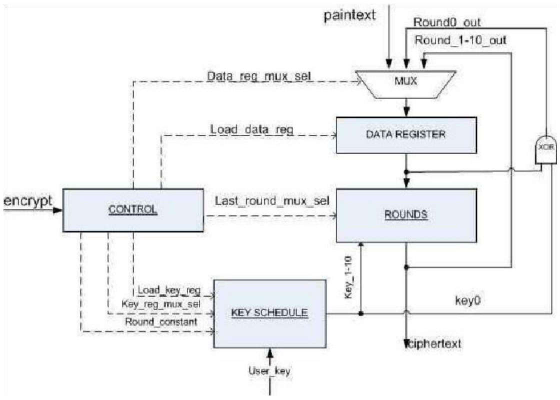


Figure.1: Basic block Diagram of Encryption Module

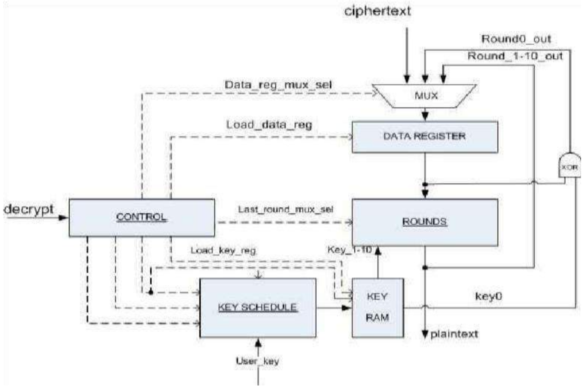


Figure.2: Block Diagram of decryption Module

TRANSFORMATIONS USED IN AES

The Advanced Encryption Standard (AES) is a block cipher. The algorithm may be used with the three different “flavors” may by referred to as “AES-128”, “AES-192”, and “AES-256”. The input to each round consists of a block of message called the state and the round key [2] . It has to be noted that the round key changes in every round. The state can be represented as a rectangular array of bytes. This array has four rows; the number of columns is denoted by Nb and is equal to the block length divided by 32. The same could be applied to the cipher key. The number of columns of the cipher key is denoted by Nk and is equal to the key length divided by 32. The cipher consists of a number of rounds - that is denoted by Nr - which depends on both block and key lengths [14].Each round of AES encryption function consists mainly of four different transformations:

ByteSub Transformation:

The ByteSub transformation is a non-linear byte substitution, operating on each of the state bytes independently. The ByteSub transformation is done using a once-pre-calculated substitution table called S-box. That Sbox table contains 256 numbers (from 0 to 255) and their corresponding resulting values.

Shift Row Transformation:

In ShiftRow transformation, the rows of the state are cyclically left shifted over different offsets. Row 0 is not shifted; row 1 is shifted over one byte; row 2 is shifted over two bytes and row 3 is shifted over three bytes.

Mix-Column Transformation:

It operates on each column individually. It takes all the columns of the state and mixes their data to produce new column.

Add Key Transformation :

The round key is applied to the state - resulted from the operation of the Mix- Column transformation - by a simple bitwise X-OR. The round key length is equal to the block length. Each Round Key consists of Nbwords from the key schedule. Those Nbwords are each added into a column of the state. The output of the above transformations is called the 'state'. The state consists of the same byte length as each block of the message. Decryption process is performed according reversed encryption scheme although mentioned earlier Trans formations have different definition to be complementary transformations. Only Add RoundKey is identical for encryption and decryption. Also the same keys are used for rounds in decryption as they were use in encryption. Such for decryption round the following transformations are in InvByteSub Transformation, InvShiftRow Transformation and InvMix Column Transformation. The below fig. shows the modified implementation of AES Encryption and Decryption Module.

IMPLEMENTATION OF AES ENCRYPTION AND DECRYPTION MODULE

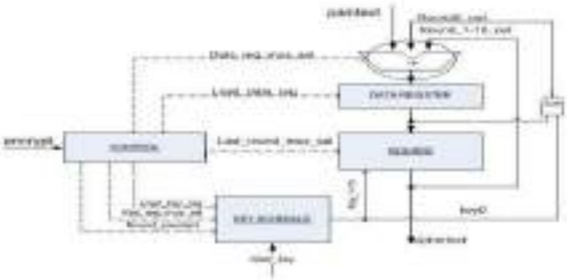


Figure.3: Modified block Diagram of Encryption Module

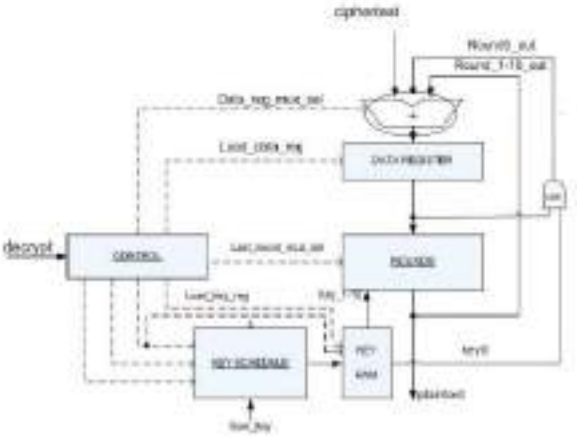


Figure.4: Modified block Diagram of Decryption Module

In the modified block diagram instead of multiplier we use shift-and-add multiplier. So it is mainly used to reduce the power consumption nearly 75% and increase the speed [10].

The Controller:

This block controls the sequencing operations of the rounds. It generates the round constant associated with each round. It also generates the control signal at the appropriate time, those control signals are: Key_reg_mux_sel:The key schedule input selector which selects the input key before the first round, and then it reads the output key as the next input.

Load_key_reg:The key schedule output enable.

load_data_reg:the input registers enable. data_reg_mux_sel:The data register selector which selects input register data from the plaintext input, last_mux_sel:it goes to the round layer to indicate the last round.

Round_constant:

it goes to the key schedule layer to indicate which round is running at the moment and it applies a constant value according to the running round. Key Generator: This block is responsible for creating the keys for each round. It contains four S boxes.

The Round Layer:

It contains the row shifter followed by two blocks running in parallel. The first one is Mix-Column block combined with an S-Box block, and the second one is S-Box block. The first block calls four look-up tables to calculate its output. The second block is the normal S-Box block. The final output is taken from S-Box combined with MixColumn block.

7. Improvements and Extensions

In this section, we discuss some possible further improvements and extensions when applying the e-voting protocol in special elections and scenarios.

Privacy of Data Transmission

In our protocol, the communication through the blockchain network may divulge voters' IP addresses, which may lead to the exposure of connections between voters and ballots via network analysis. To enhance voters' privacy, we recommend voters to use anonymity services like proxies or TOR [26], with which voters can hide their IP addresses.

Data Confidentiality and Neutrality

According to our protocol, because of the transparency property from blockchain, ballots are visible when they are cast to the blockchain network.

This exposes the progress of the election during the voting phase, and may greatly influence the outcome of the election. Here, we provide two possible solutions for this problem.

The permissioned blockchain is more flexible and there exist several promising solutions of access control (like [27]). However, providing certain extent of data confidentiality, transparency is somehow lost. To keep transparency, we may also introduce a ballot encryption mechanism here.

Dishonest Behaviors from the Organizer and Inspectors

Corruption may happen if the organizer and inspector conspire together, since both of their signatures are components of a valid ballot. To avoid this kind of dishonest behaviors, we can introduce more inspectors such that the corruption cost is greatly increased.

8. CONCLUSION

E-voting, as discussed in the paper, is a potential solution to the lack of interest in voting amongst the young tech savvy population. For e-voting to become more open, transparent, and independently auditable, a potential solution would be base it on blockchain technology. This paper explores the potential of the blockchain technology and its usefulness in the e-voting scheme. The paper proposes an e-voting scheme, which is then implemented. The implementation and related performance measurements are given in the paper along with the challenges presented by the blockchain platform to develop a complex application like e-voting. The paper highlights some shortcomings and presents two potential paths forward to improve the underlying platform (blockchain technology) to support e-voting and other similar applications. Blockchain technology has a lot of promise; however, in its current state it might not reach its full potential. There needs to be concerted effort in the core blockchain technology research to improve is features and support for complex applications that can execute within the blockchain network.

References

1. J. Demuro, "Here Are the 10 Sectors That Blockchain Will Disrupt Forever," TechRadar Pro, 16 Jan. 2018; https://www.techradar.com/news /here-are-the-10-sectors-that -blockchain-will-disrupt-forever.

2. B. Dickson, "Blockchain Tech Could Fight Voter Fraud—and These Countries Are Testing It," VentureBeat, 22 Oct. 2016; https://venturebeat .com/2016/10/22/blockchain -tech-could-fight-voter-fraud-and -these-countries-are-testing-it.

3. J. Hall, "Can Blockchain Technology Solve Voting Issues?," Bitcoin Magazine, 7 Mar. 2018; https://www .nasdaq.com/article/can-blockchain -technology-solve-voting-issues -cm931347.

4. A. Sandre, "Blockchain for Voting and Elections," Hackernoon, 14 Jan. 2018; https://hackernoon.com /blockchain-for-voting-and-elections -9888f3c8bf72.

5. G. Prico, "Sierra Leone Pilots Blockchain-Based Voting for Political Elections," 22 Mar. 2018; https: //www.nasdaq.com/article/sierra-leone -pilots-blockchain-based-voting-for -political-elections-cm938309.

6. B. Miller, "Blockchain Voting Startup Raises $2.2M," Government Technology, 8 Jan. 2018; http: //www.govtech.com/biz/Blockchain -Voting-Startup-Raises-22M.html.

7. A. Perala, "Voatz Raises $2.2 Million in Seed Funding," Mobile ID World, 9 Jan. 2018; https://mobileidworld .com/voatz-seed-funding-901093.

8. M. Hochstein, “Moscow’s Blockchain Voting Platform Adds Service for High-Rise Neighbors,” CoinDesk, 15 Mar. 2018; https://www.coindesk .com/moscows-blockchain-voting -platform-adds-service-for-high-rise -neighbors.

9. “Digital Home Blockchain Voting System, Active Citizen in Moscow

Opens,” BitccoinExchangeGuide .com; https://bitcoinexchangeguide .com/digital-home-blockchain-voting -system-active-citizen-in-moscow -opens.

10. M.D. Castillo, “Russia Is Leading the Push for Blockchain Democracy,” CoinDesk, 2018; https://www .coindesk.com/russias-capital -leading-charge-blockchain -democracy.

11. B. Kimathi, “Why You Shouldn’t Get Carried Away by Sierra Leone’s Blockchain Elections,” Crypto-Lines, 13 Mar. 2018; https://crypto-lines .com/2018/03/13/blockchain -elections.

12. S. Holder, “Can the Blockchain Tame Moscow’s Wild Politics?,” CityLab, 22 Dec. 2017; https://www.citylab .com/life/2017/12/can-the-blockchain -tame-moscows-wild-politics/547973.

13. “A South Korean Province Used Blockchain Tech for Resident Voting,” CCN.com, 8 Mar. 2017; https://www.ccn.com/south-korean -province-used-blockchain-tech -resident-voting.

14. “South Korea Uses Blockchain Technology for Elections,” KryptoMoney, 2 May 2017; https://kryptomoney .com/south-korea-uses-blockchain -technology-for-elections.

15. S. Waterman, “Nasdaq Says Estonia E-Voting Pilot Successful,” CyberScoop, 25 Jan. 2017; https://wcyberscoop.com/nasdaq-estonia -evoting-pilot.

Identification of Anomaly Activities on Social Network

Mrs. Mangala.C.N[1],Prabhudev B K[2],Mayur Jain[3],Prashanth A[4],Sagar Reddy N J[5]

Assoc. Professor[1],Student[2]

Dept. of Computer Science & Engineering

East West Institute of Technology,Bengaluru, India

mangalacn@ewit.edu.in[1], prabhudevkumar007@gmail.com[2], mayurjainkm@gmail.com[3], prashanthasharma@gmail.com[4], sagarreddynj@gmail.com[5]

Abstract: Social media is arguably the richest source of human generated text input. Opinions, feedbacks and critiques provided by internet users reflect attitudes and sentiments towards certain topics, products, or services. Every day, millions of messages are created, commented, and shared by people on social media websites, such as Twitter and Facebook. This provides valuable data for researchers and practitioners in many application domains, such as marketing, to inform decision-making. Distilling valuable social signals from the huge crowd's messages, however, is challenging, due to the heterogeneous and dynamic crowd behaviors. These are the anomalies caused by a user because of his/her variable behavior towards different sources. Due to such risk parameters, it is always a best practice to have a mechanism to assign a risk score to each online social network user. This paper therefore put forth risk analysis of Facebook. This work hence acts as a risk indicator to the administrator of the Face book services so that they can formulate strategies to overcome the same.

Keywords: *Anomaly Detection, Social network, SVM, Data Analytics*

I INTRODUCTION

Over the recent years, the surge of social media, such as Twitter and Facebook, has significantly advanced the way that people publish, acquire, and share news and information. All day long, millions of messages are created, commented on, and disseminated by over one billion active social media users [10]. Such publicly available texts as well as their propagation patterns among people provide great potential for researchers and practitioners in a variety of fields, such as political science and marketing, to make data-informed decisions. While there is abundant information on social media, not every posting is equally valuable, important or informative. The first challenging question is: which particular message streams are worth looking into? To be efficient, analysts aim to identify anomalous

Anomalies are the unexpected behavior of the user which results in irregular and suspicious activity causing threats to the information and the regular network users. With the advent of increased online social interaction sites, user tracking and anomaly detection in social networks are two of the major areas of research. The primary goal of detecting anomalies is to identify the accustomed trends of incredulous activities in the network [17]. A lot of research has been carried out to build a generalized method for anomaly detection. A number of well-developed methods are available for detecting them under specific conditions on different domains.

Over the last few years, detection of the anomalies has been taken as a serious research which required efficient approaches for improved identification. However, the approaches proposed so far are valid for networks under certain pre-defined parameters which mostly involves the level of information exchange between the source and the users.

One of the most popular Online Social Network is Facebook which provides platform where user can keep in touch with family, friends and share the information among them. Facebook is also used for commercial purposes. Despite of drastic increase in OSN usage – Facebook, for instance, has now 1 billion daily users, 1.3 billion mobile users, 1.55 billion monthly active users; which has led to lot of security and privacy concerns.

Anomaly detection is based on the idea that the characteristics of normal behavior can be distinguished from abnormal behavior [1].

In this paper, we describe a new method for anomaly detection in social media posts. The basis for this method is a series of patents filed in [4] [5] and [6].The rest of this paper is organized as follows. Section II discusses the existing sensing technologies. Section III presents the proposed methodology of fine-grained sentiment analysis. Section IV provides the system architecture for proposed method using real world anomaly buzzwords. Lastly, in Section V, we conclude this study along with all the references.

II RELATED WORK

The anomaly detection in online social networks can be carried out in a number of ways. Over the years, multiple variants of anomalies have been identified and targeted with strategic solutions. These solutions focus on the categorization of the anomaly and then provide solutions which can resolve the problem of user identification.

Compromised Account-based Anomaly Detection

Another aspect of the anomalies in the online social networksis the compromised accounts which have been examined by Egele et al. [6] [3]. The authors developed an approach under the name of COMPA, which can identify the compromised accounts in most of the social networking sites. The authors analyzed and tested their approach on a large data set comprising approximately 1.4 billion Twitter messages which are publicly available. These systems can be classified as Intrusion Detection Systems (IDS) particularly focusing on the anomaly detection in the online social networks as stated by Sommer and Paxson [17]. The authors presented the utility of machine learning approaches to the formation of an IDS which can efficiently track the network anomalies.

Interaction-based Anomaly Detection

Point of interaction can be another solution for identifyinganomalies. Such approach utilizes the concept of anomaly scores by analyzing the sources with which a user interacts. Takahashi [12] proposed change-point detection technique which uses the Sequentially Discounting Normalized Maximum Likelihood (SDNML). The authors utilized the anomaly scores obtained from these experiments to identify the link anomalies. In the other approach by Yu et al. [4], the authors proposed a Group Latent Anomaly Detection (GLAD) approach which uses the pair-wise as well as pointwise data to inference at the final decision of anomalies. Their approach is efficient but lacks applicability to the horizontal anomalies because of its dependency on group features for each individual, whereas horizontal anomalies arise due to an individual's activity irrespective of the group to which it belongs.

Previous work presented in this section clearly shows that most of the existing approaches have been generic in the detection of the anomalies and have not considered for live anomaly detection. Thus, efficient approaches are required which can not only identify the threat level caused by those anomalies but also resolves these efficiently by taking further actions.

The first major limitation is that the training data needs tobe large enough to allow sufficient representation of full target domains. In the real-world social media context, it is hard to determine the effective size for a training dataset. This is due to the diversity of the social discussion being unknown a priori. The second major limitation is that the training data must be of good quality, requiring domain experts to clean up the training data. This makes learning-based approaches too costly and impractical to be applied to new domains. Their over-reliance on training databases means the machine learning- based methods are not directly applicable since training datasets are not always available.

To tackle the limitations as well as the challengesmentioned above, this study proposes an new anomaly detection method which enhances the current methods for anomaly detection, through vectors acting live on on social media. The applicability of the proposed method is demonstrated using a live model of social network.

III EXISTING TECHNOLOGIES

Sentiment analysis methods can be broadly categorized into two types: learning-based and lexical-based [8] [9]. Learning based method uses known properties derived from labelled training data to make predictions about unlabeled new data. In text data, it derives the relationship between the features of the text segment. Some examples of learning-based methods are the Naive Bayes (NB) classifier [10] [13], Maximum Entropy (MaxEnt) classifier [11], support vector machine (SVM) [14][18] and Extreme Learning Machine (ELM) [19] [2].To be effective, models using such learning-based methods typically require a sufficiently large labelled training dataset[2] [20] to achieve an acceptable classification accuracy [20][7]. However, in most social media contexts, it is difficult to determine what size of labelled dataset qualifies as being sufficient because the diversity of the social discussion is not known a priori [5] [14]. In addition, the labelling task would be costly or even prohibitive [5] [8] [14], not to mention wasteful because the training results could not be really applied to other datasets.

IV PROPOSED METHOD

Our proposed demonstrates how anomaly can be detected in live on social media such as Facebook. The proposed system creates a social network hosted on a cloud service. The social network acts as a replica to actual Facebook site. The users can register on the social network and can login with login credentials anytime .The users are greeted with home page and are provided with an option to share his/her opinions through posts. As soon as the user posts, the text is stored and compared with predefined feature vectors created using anomaly buzz words, the text are dimensionally reduced and concluded as either anomaly affected or not. In case of mixed posts the Gradient Based SVM is applied on the text with the help of feature vectors and anomaly is detected, as shown in Figure 1.

Depending upon result of this anomaly detection module a report is generated which contains information about users and number of the time they have voided the rules. This information is provided graphically to the admin using data analytics. Depending upon report generated appropriate action are taken (sending warning mail/blocking the user)

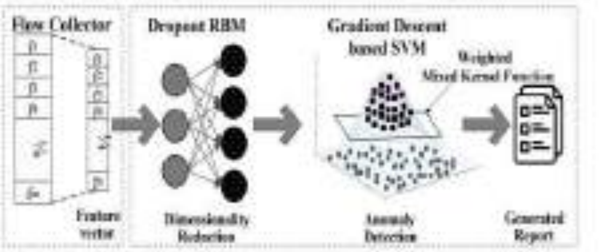


Figure 1: Anomaly detection module

VSYSTEM ARCHITECTURE

User first logins to the social network cloud with fresh registration or login credentials. The user then enters a text messages, and these text messages are stored in database. From the database, all the messages are extracted and the classification of the messages is done, that is messages are classified into Anomaly posts and normal posts. The classification of the posts is done using SVM. For the input messages anomaly identification rules are applied to determine the category of the words in the messages. Neurons are trained using the training dataset. Finally result of the messages is predicted that is whether the entered posts are anomaly affected or not.

The system keeps a count of number of times a user tweets a anomaly text. After 3 anomaly tweets the system warns the user through email and after 5 anomaly tweets the user will be blocked and notified the same through email,The admin of the social network has full control to remove anomaly tweets any time. The admin is also provided with data analytics about all the users and their number of anomaly actions.The system also allows users to categorize available tweets and view them through notepad. The users are provided with options of categorizing tweets and also an option to remove their anomaly tweets personally. Figure 2 shows the system architecture of proposed model.

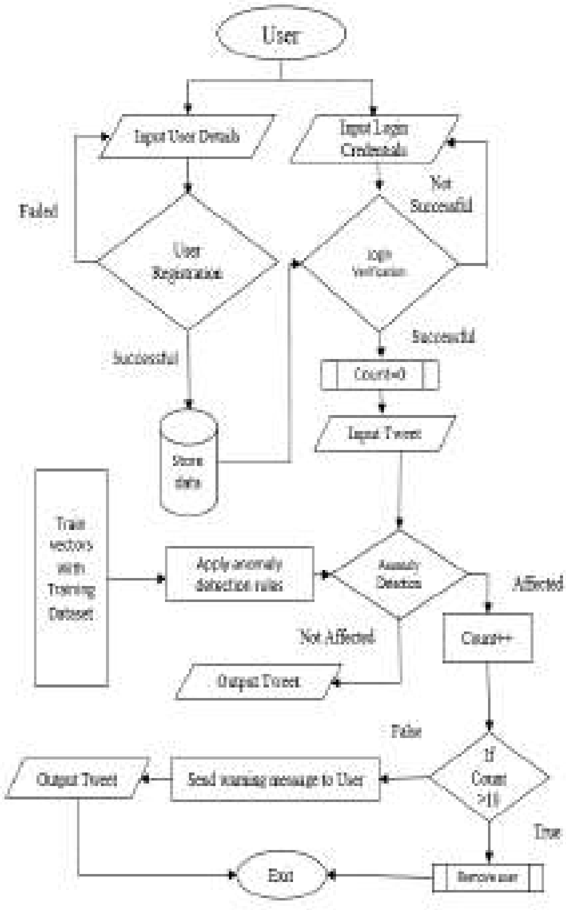


Figure 2: An Overview Architecture for AnomalyDetection.

VIALGORITHM

The Anomaly check on the text can be determined using support vector machines. The anomaly in the text is distinguished in two ways as direct anomaly and mixed anomaly. When a person enters a comment or post in social media it is compared with vectors created using training data.

The proposed algorithm creates vectors (Hidden and visible) with the training data provided. The target class contains testing data in the form of posts. The training objective is created for the vectors created. The contents of the target class are made to pass through this objective function. As a result the vectors trained will check for the buzzwords in the testing data set to classify the testing posts as anomaly posts or not.In case of mixed posts the anomaly is identified using +, - properties of the trained vectors. In case of direct posts the anomaly is found by passing trained vectors and the test data into a objective function created for anomaly detection. The algorithm with input and output is mentioned below as stepwise.

Algorithm: Anomaly text classification using support vector machine

```
Input: Training dataset with anomaly buzz words
Output: classification of testing text to be affected by anomaly or not
1: Load training dataset
2: Sample training vector from training dataset
3: Initialize weights W and bias a and b
4: Set m visible units (v)
5: Set n hidden units (h)
6: Compute conditional probability P for all v
7: Compute conditional probability P for all h using dropout
8: Initialize target class c = {c1, c2, · · ·,ct}
9: Set training objective
10: To deal with the computational problem, compute gradient of log P(ct, vt), i.e., ∂ log P(ct, vt)/∂θ
11: Repeat the procedure G times to classify all target class members
12: Return classified texts
```

VIIRESULTS AND DISCUSSION

This algorithm helps in differentiating the text from the normal and anomaly text, which in turn will help in keeping the social network free from anomaly tweets. As said,this algorithm plays a major role in categorizing the text.The test text goes into the classification algorithm, according to the algorithm the result of the process is defined. The action on anomaly users are takes based on the value of the count variable.The result can be either of the three followings:no action (if count<5), send warning mail to the user (if count=5) and remove user account (if count>10)

VIICONCLUSION AND FUTURE WORK

In this paper, a vector based anomaly detection model is proposed that accounts for efficient detection of anomalies in online social networks. The proposed model uses support vector machine and detect anomaly action as soon as performed (Live).Study results indicated that SVM based anomaly detection algorithm is efficient in identifying anomaly actions and the visualization is useful for analysts to discover insights and comprehend the model.The proposed approach uses limited formulations even for identifying complex anomalies and the number of iterations required for arriving at a decision is less than the theoretical values.

In the future, we will further investigate anomaly detection models for Twitter conversational threads and improve the current algorithm to allow a faster analysis. In addition to anomaly detection, it is interesting to integrate other content features (e.g., topics and semanticinformation) to the current system.Results suggest that the proposed model proves to be efficient in terms of significant gains attained in comparison with the existing approaches over various parameters namely, anomaly filtering rate, accuracy in anomaly detection, convergence value, approach failures.

ACKNOWLEDGEMENT

First and foremost, we express our sincere thanks to Dr. ArunBiradar,Head, Department of computer science and engineering for his moral support. We express our sincere gratitude to our principal Dr.KChennakeshavalu for his constant support and encouragement, we also thank all the faculties of East West Institute of Technology for their co-operation and support.

REFERENCES

[1] M Swarnasudha ,KArunPriya,"Data mining approach for anomaly detection in social network analysis",2018,ICICCT conference 2018

[2] Z. Wang and Y. Parth, "Extreme Learning Machine for Multi-class Sentiment Classification of Tweets," Proc. ELM-2015, Springer Int. Publ. 2016, vol. 1, pp. 1–11, 2016.

[3] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Towards detecting compromised accounts on social networks," IEEE Transactions on Dependable andSecureComputing,vol.14,no.4,pp.447 – 460,2015.

[4] R. Yu, X. He, and Y. Liu, "Glad: group anomaly detection in social media analysis," ACM Transactions on Knowledge Discovery from Data, vol. 10, no. 2, pp. 18–22, 2015.

[5] Z. Wang, J. C. Tong, and D. Chan, "Issues of social data analytics with a new method for sentiment analysis of social media data," in 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, 2014, pp. 899–904.

[6]M.Egele ,G. Stringhini," Compa: Detecting compromised accounts on social networks.," in NDSS, 2013,

[7] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Computer Science, vol. 17, pp. 26-32, Jan.2013.

[8] P. Gonçalves and M. Araújo, "Comparing and combining sentiment analysis methods," Proc. first ACM Conf. Online Soc. networks. ACM., pp. 27–38, 2013.

[9] B. Yuan, Y. Liu, and H. Li, "Sentiment classification in Chinese microblogs:Lexicon-based and learning-based approaches," Int. Proc. Econ. Dev. Res., vol. 68, pp. 1–6, 2013

[10] J. Ortigosa-Hernández, J. D. Rodríguez, L. Alzate, M. Lucania, I. Inza, "Approaching sentiment analysis by using semisupervisedlearning of multi-dimensional classifiers," Neurocomputing, vol. 92, pp. 98-115, Sep. 2012.

[11] H. Ji, H. Deng, and J. Han, "Uncertainty reduction for knowledge discovery and information extraction on the World Wide Web," Proceedings of the IEEE, vol. 100, no. 9, pp. 2658-2674, Sep. 2012.

[12] T. Takahashi, R. Tomioka, and K. Yamanishi, "Discovering emerging topics in social streams via link anomaly detection,"2011 IEEE 11th International Conference on Data Mining, pp. 1230–1235,2011.

A Survey on Bloom Filter and its Application

Pooja N[1], Information Science and Engineering, Global Academy of Technology, Bengaluru, India
Lakshmi R[2], Information Science and Engineering, Global Academy of Technology, Bengaluru, India,
Pratiksha P Kulkarni[3], Information Science and Engineering, Global Academy of Technology, Bengaluru, India
Nischitha G[4], Information Science and Engineering, Global Academy of Technology, Bengaluru, India.
poojanarasimhamurthy10@gmail.com[1], lakshmi505@gmail.com[2], pratikshakulkarni814@gmail.com[3]
nischithagowda2000@gmail.com[4]

***Abstract* - Bloom filters are used for membership queries over sets with allowable errors also called false positive. It is extensively used in databases, networks and distributed systems and it has great potential for distributed applications where systems need to share information about available data. This paper presents the various applications of blooms filter. Bloom filter plays a vital role in searching technique. It is a probabilistic data structure that is used to test whether an element is a member of a given set. Number of applications are using this technology for accessing and processing the data. They are also used widely in many network applications. Recently, data de-duplication, that has got tremendous scope for research, has received a broad attention from both academia and industry. Some researches concentrate on the approach by which to diminish more redundant data. And the others investigate how to do de-duplication at higher speeds. Bloom Filter plays a vital role in data de-duplication and query optimization in Big Data.**

Keywords—Bloom Filter, De-duplication, Big Data, Query Optimization

I. INTRODUCTION

Bloom filter was invented by Bloom (1970) and is used widely today for diverse purposes including data de-duplication. The theory behind the Bloom filter is described in this section. At first, the Bloom filter is described and then its enhancement to meet the requirement of string detection is explained. Bloom filter is a probabilistic data structure which is used to test a membership of an element in a given set. A Bloom filter suggests a striking choice for string matching. Using this technique, a group of strings is compressed at first by calculating multiple hash functions over each string. Then, compressed set of strings is stored in memory. This set can be queried to find out if a given string belongs to it. The two important properties of a Bloom filter that make it a viable solution for string matching are the following:

Scalability: Bloom filter uses a constant amount of memory to compress each string irrespective of the length of the original string. Thus, large strings can be stored with smaller memory space. This makes it highly scalable in terms of memory usage.

Speed: The amount of computation involved in detecting a string using Bloom filter is constant. This computation is a calculation of hash functions and the corresponding memory lookups. Efficient hash functions can be implemented in hardware easily with little resource consumption. Hence, a hardware implementation of Bloom filter can do string matching at high speeds. Bloom filters use less memory space to store the compressed strings. The amount of memory depends on the number of strings being compressed and typically is few megabits. For instance, to store 10,000 strings, around 200k bits are required.

The working of Bloom filter is explained in this section. Given a string x, the Bloom filter computes k hash functions on it producing hash values ranging from 1 to m. It then sets k bits in an m bit long vector at the addresses corresponding to the k hash values. The same procedure is repeated for all the members of the set. This process is called *programming* of the filter. The *query* process is similar to programming, where a string whose membership to be verified is given as input to the filter. The Bloom filter generates k hash values using the same hash functions which are used to program the filter. The bits in the m bit long vector at the locations corresponding to the k hash values are looked up.

If at least one of these k bits is found not set then the string is declared to be a non-member of the set. If all the bits are found to be set then the string is said to belong to the set with a certain probability. This uncertainty in the membership comes from the fact that those k bits in the m bit vector can be set by any of the n members. Thus, finding a bit set does not necessarily imply that it was set by the particular string being queried.

II. RELATED WORK

Hash table data structure and algorithm which outperforms the conventional hash table algorithms by providing better bounds on hash collisions and the memory access per lookup. Hash table algorithm extends the multi-hashing technique, Bloom filter, to support exact match. However, unlike the conventional multi-hashing schemes, it requires only one external memory for lookup. By using a small amount of multi-port on-chip memory, how the accesses to the off-chip memory, either due to collision or due to unsuccessful searches, can be reduced significantly is examined. Through theoretical analysis and simulations authors show that hash table is significantly faster than the conventional hash table. Thus, Fast Hash Table can be used as a module to aid several networking applications. Among the conventional avenues to improve the hash table performance, using sophisticated cryptographic hash functions such as MD5 does not help since they are too computationally intensive to be computed in a minimum packet-time budget; devising a perfect hash function by pre-processing keys does not work for dynamic data sets and real-time processing; and multiple-hashing techniques to reduce collisions demand multiple parallel memory banks (requiring more pins and more power). Hence, en engineering a resource efficient and high-performance hash table is indeed a challenging task [1].

Network Intrusion Detection and Prevention Systems (IDPS) use string matching to scan Internet packets for malicious content. Bloom filters offer a mechanism to search for a large number of strings efficiently and concurrently when implemented with Field Programmable Gate Array (FPGA) technology. A string matching circuit has been implemented within the FPX platform using Bloom filters. Using 155 block RAMs on a single Xilinx VirtexE 2000 FPGA, the circuit scans for 35,475 unique signatures. By using Bloom filters, an IDPS can be implemented that scans for tens of thousands of strings at Gigabit per second rates, all within a single FPGA. If a Bloom engine detects a match, a hash table is queried to determine if an exact match occurred. If the queried signature is an exact match, the malicious content can be blocked and an alert message is generated within an User Datagram Protocol (UDP) packet, informing a network administrator, an end user, or an automated process that a matching signature has been detected [2].

There are numerous examples where one would like to use a list in a network. Especially when space is an issue, a Bloom Filter may be an excellent alternative to keeping an explicit list. The drawback of using a Bloom Filter is that it introduces false positives. The effect of a false positive must be carefully considered for each specific application to determine whether the impact of false positives is acceptable [3].

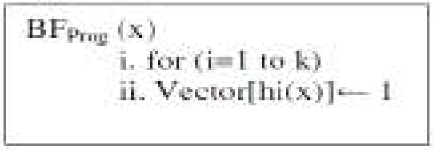
[4] Analyzed the practical performance of hashing functions used in hardware applications such as page tables for address translation. The practical performances of bit extraction, exclusive ORing, and H3 class of hashing are addressed. The results showed that by choosing functions randomly from this class of hashing functions, which can be rapidly implemented in hardware, an analytically predicted performance of hashing schemes with real life data can be achieved.

III. Bloom Filter working

A. Programming a Bloom Filter

A Bloom filter is essentially a bit vector of length m which is used to efficiently represent a set of bit-strings. Given a set of strings S, with n members, a Bloom filter is “programmed” as follows. For each bit string x, in S, k hash functions, $h_1()...h_k()$, are computed on x producing k values each ranging from 1 to m. Each of these values addresses a single bit in the m bit vector; hence each bit-string x causes k bits in the m-bit vector to be set to 1. It is to be noted that if one of the k hash values addresses a bit that is already set to 1, then that bit is not changed. Figure 3.1 illustrate Bloom filter programming. Two bit-strings, x and y are programmed in the Bloom filter with k = 3 hash functions and m = 16 bits in the array. It is to be noted that different strings can have overlapping bit patterns. The following pseudo-code describes adding a bit-string, x, to a Bloom filter. Pseudo-code for programming the Bloom filter is given in Table 1.

Table 1: Pseudo-code for programming the Bloom Filter



```
BF_Prog (x)
    i. for (i=1 to k)
    ii. Vector[hi(x)]← 1
```

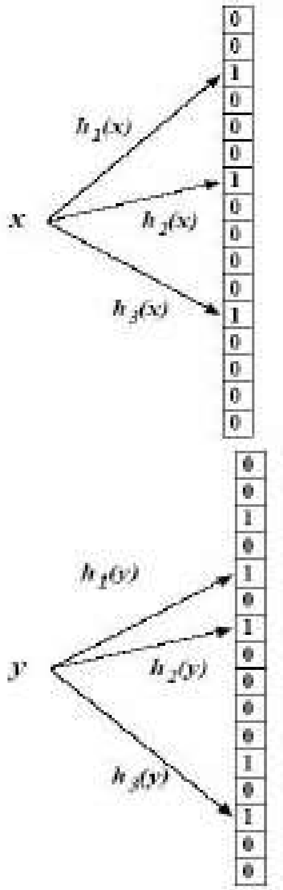
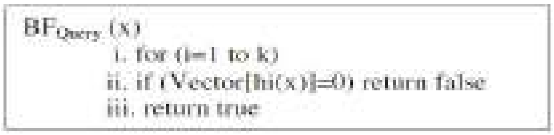


Figure 1: Programming Strings ‘x’ and ‘y’ in Bloom filter

B. Querying a Bloom Filter

Querying the Bloom filter for set membership of a given bit-string x, is similar to the programming process. Given bit-string x, k hash values are generated using the same hash functions used to program the filter. The bits in the m-bit vector at the locations corresponding to the k hash values are checked. If at least one of the k bits is 0, then the bit-string is declared to be a non-member of the set, as discussed in Figure 3. If all the bits are found to be 1, then the bit-string is said to belong to the set with a certain probability, as shown in Figure 2. If all the k bits are found to be set and x is not a member of S, then it is said to be a false positive. The following pseudo-code describes the query process. Pseudo-code for querying the Bloom filter is given in Table 2.

Table 2: Pseudo-code for querying the Bloom Filter



```
BF_Query (x)
        i. for (i=1 to k)
        ii. if (Vector[hi(x)]=0) return false
        iii. return true
```

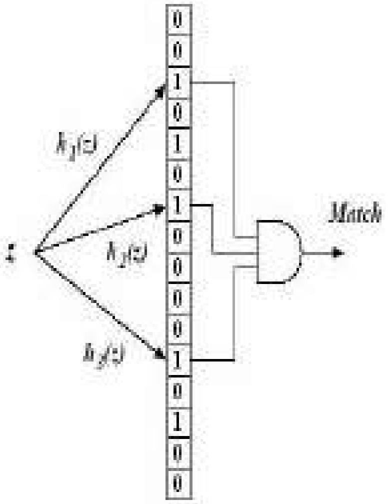


Figure 2: Querying z where all k bits are 1

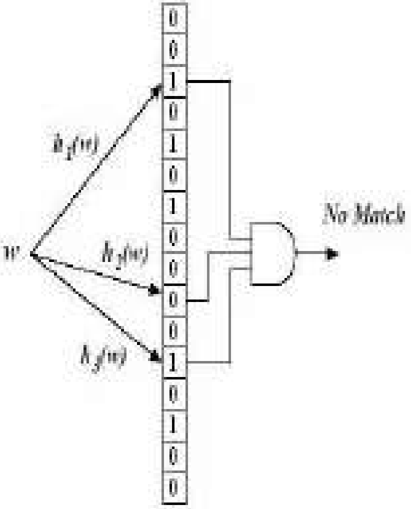


Figure 3: Querying w where one of the k bits is 0

IV. Applications of Bloom Filter

A. Bloom In Flash for data de-duplication

Usually bloom filters are implemented using Random Access Memory (RAM) which is limited and expensive. This leads to higher false positive probability which decreases the de-duplication efficiency. A Flash Based Segmented Bloom Filter for De-duplication is proposed in [5] which implements its bloom filter (BF) on solid state drive where only part of the whole bloom filter will be kept in RAM while the full bloom filter is on solid state drive. This improves duplicate lookup in three ways. First the size of the bloom filter can be sufficiently large. Second, more number of hash functions can be used. And last, there will be more RAM space for fingerprint cache.

B. Locality Sensitive Bloom Filter in Big Data

Implementing Bloom's Filter over big data will result into efficient query accessing in big data. An approach to implement Locality Sensitive Bloom Filter (LSBF) technique in big data is proposed in [6]. To remove the drawbacks of simple hashing technique, the LSBF must be implemented to store data in the bloom filter which will help to search the most approximate result by using the Locality Sensitive Hashing approach.

Locality Sensitive Hashing (LSH) associates similar data elements in the same hash buckets with a high probability to serve main memory algorithms for searching the similar items. LSH was introduced in [7]. For similarity searching we need to hash the query q into buckets in multiple hash tables, and then by combining items from the chosen hash bucket by ranking to their distance from the point *q.* we can thus choose the closest element to be the most exact to the queried one. The elements present close to each other will have a higher colliding probability than the elements that are far apart in LSH

C. Other Applications

Transactional Memory has recently applied Bloom filters to detect memory access conflicts among threads.

Any unique identification system has to generate a unique number for newly registered users.If the number of user registrations increase dramatically checking with the database is too expensive. In the case a bloom filter can tell if a number has already been generated or not.if yes simply generate a new random number and check with filter again.This is done until the blooms filter returns false.

URL shortness are usually done by making a call to sever which generates fresh URL and sends it back. But it is very difficult to check the uniqueness .In that case Blooms filter is used to tell if the URL has already been generated early and it keeps generating the new one until it returns false.So this makes the work much easier and a considerable price.

Caches Optimisation When a website page is revisited we come across caches which usually have LRU policy ,which means this one hit wonder will take up valuable memory in the server cache .To solve this blooms is used to store page entries only in cache If the blooms filter returns true.

Oracle uses blooms filter to perform bloom pruning of partitions for certain queries.For example when joining a date dimension table with a large fact table partitioned by data,a bloom filter can be built over the date dimension and used to prune partitions from the fact table.

Compact Representation of a Differential File:A differential file contains a batch of database records to be updated .For performance reasons the database is updated only periodically or when the file grows above certain threshold .however to preserve integrity each reference to the database has to access the differential file to see if a particular record is scheduled to be updated.To speed up the process with little memory and computation overhead the file is represented as blooms filter

LSBF(locality sensitive bloom filter):It is a proposed structure which can be efficiently used in many real world application due to its properties such as input mistake tolerance ,fast query response, assistance to similar query and system performance improvement.

Estan and Varghese present an excellent application of Bloom filters to traffic measurement problems inside of a router, reminiscent of the techniques used in the Stochastic Fair Blue algorithm . The goal is to easily determine heavy flows in a router. Each packet entering is hashed k times into a Bloom filter. Associated with each location in the Bloom filter is a counter that records the number of packet bytes that have passed through the router associated with that location. The counter is incremented by the number of bytes in the packet. If the minimum counter associated with a packet is above a certain threshold, the corresponding flow is placed on a list of heavy flows. Heavy flows can thereby be detected with a small amount of space and a small number of operations per packet. A false positive in this situation corresponds to a light flow that happens to hash into k locations that are also hashed into by heavy flows, or to a light flow that happens to hash into locations hit by several other light flows .The idea of a conservative update, an interesting variation that reduces the false positive rate significantly for real data. When updating a counter upon a packet arrival, it is clear that the number of previous bytes associated with the flow of that packet is at most the minimum over its k counters. Call this Mk. If the new packet has B bytes, the number of bytes associated with this flow is at most Mk +B. So the updated value for each of the k counters should be the maximum of its current value and Mk + B. Instead of adding B to each counter, conservative update only changes the values of the counters to reflect the most possible bytes associated with the flow.

V. CONCLUSION

A Bloom filter is a simple space-efficient representation of a set or a list that handles membership queries. As we have seen in this survey, there are numerous networking problems where such a data structure is required. Especially when space is an issue, a Bloom filter may be an excellent alternative to keeping an explicit list. The drawback of using a Bloom filter is that it allows false positives. Their effect must be carefully considered for each specific application to determine whether the impact of false positives is acceptable.

This leads us back to: The Bloom filter principle: Wherever a list or set is used, and space is at a premium, consider using a Bloom filter if the effect of false positives can be mitigated. There seems to be plenty of room to develop variants or extensions of Bloom filters for specific applications. For example, we have seen that the counting Bloom filter allows for approximate representations of multi-sets or dynamic sets that change over time through both insertions and deletions. Bloom filters 506 Internet Mathematics are now starting to receive significant attention from the algorithmic community, and while there have been a number of recent results, there may well be further improvements to be found. We expect that the recent burst of applications is really just the beginning. Because of their simplicity and power, we believe that Bloom filters will continue to be used in modern network systems in new and interesting ways.

REFERENCES

[1] S. Dharmapurikar, H. Song, J. Turner, and J. W. Lockwood, “Fast Hash Table Lookup Using Extended bloom Filter: An Aid to Network Processing”, *ACM/SIGCOMM*, pp. 181-192. Philadelphia, 2005.

[2] M. Attig, S. Dharmapurikar, and J.L. Lockwood. “Implementation Results of Bloom Filters for String Matching”, *Proc. of IEEE Symp. on Field- Programmable Custom Computing Machines*, pp. 322-323, 2004.

[3] A. Broder and M. Mitzenmacher, “Network Applications of Bloom Filters: A Survey”, *Internet Mathematics*, vol. 1, no. 4, pp. 485-509, July 2003.

[4] M. Ramakrishna, E. Fu, and E. Bahcekapili, “Efficient Hardware Hashing Functions for High Performance Computers”, *IEEE Trans. on Computers*, vol. 48, no. 12, pp. 1378-1381, 1997.

[5] G. Dagnaw, A. Teferi, and E. Berhan, A. Abraham, P. Krömer, & V. Snášel (eds.), “Flash Assisted Segmented Bloom Filter for Deduplication”, *Afro-European Conf. for Ind. Advancement*, Advances in Intelligent Systems and Computing 334, DOI: 10.1007/978-3-319-13572-4_7 © Springer International Publishing Switzerland 2015.

[6] Mayank Bhushan, Monica Singh , Sumit K Yadav *“*Big Data query optimization by using Locality Sensitive Bloom Filter” *2nd International Conference on Computing for Sustainable Global Development (INDIACom), 2015*

[7] Indyk and Motwani, “Approximate Nearest neighbors: Towards Removing the Curse of dimensionality,” Proc. 13th Ann. ACM Symp. Theory of Computing, pp. 604-613, 1998.

Design and Implementation of Efficient Email Spam Detection Enabling Privacy Preservation

Vaishnavi S Hegde[1],Ambika L G[2],Yashaswini G[3],Kavya M[4], Sahana Y B[5]
Dept. of Computer Science and Engineering,
Rajarajeswari College of Engineering, Bangalore, India
vais.hegde@gmail.com[1],ambikalg1991@gmail.com[2],yashaswinig710@gmail.com[3],kavyam451@gmail.com[4],sahanabalkrishna@gmail.com[5]

Abstract

Spam emails are a chronic issue in computer security. It is very costly economically and extremely dangerous for computers and networks. Despite of social networks and other Internet based information exchange venues, dependence on email communication has been increased over the years. Although there are many spam filters created to detect these spam emails by entering a user's inbox. Currently, Naive Bayes is one of efficient method of spam classification because of its simplicity and efficiency and it is very accurate. But it is unable to correctly classify all the emails when they consists leetspeak or diacritics. Thus, in this proposes, we implemented a novel algorithm to enhance the accuracy of the Naive Bayes Spam Filter so that it can detect and correctly classify the email as a spam or ham. Keyword based and machine learning algorithms are used to increase the accuracy of Naive Bayes and Hidden Markov model. Additionally, we have discovered a relationship between the Naïve Bayes Classifier and Hidden Markov Model to get more accurate spam emails.

Keywords - Email,Spam Filter, Naive Bayes Classifier, Text Classification, Hidden markov model.

I. Introduction

The growth of the emails has also led to unprecedented increase of the number of illegitimate mail, or spam - 49.7% of emails sent is spam because current spam detection methods have lack in accurate spam classifier. Spam is a problematic but not only because it is also the carrier of malware, and also spam emails hoard network bandwidth, storage space, and computational power. Additionally, the commercial world has a significant interests in the spam detection because of spam it causes loss in work productivity and financial loss.

This project investigates a comparison between two different approaches for classifying emails. Naive Bayes and Hidden Markov Model (HMM), are the two different machine learning algorithms, both have been used for detecting whether an email is important or spam.

Naive Bayes Classifier is based on conditional probabilities. By taking small datasets as an input it processes fast. It considers independent words as a feature. HMM is a generative model that provides with the distribution over the sequences of observations. HMMs can handle inputs of variable length and help programs come to the most likely decision, based on both previous decisions and current data. In this paper, we will discuss related methods, definitions, our new method, results.

II. Existing System

Enrico Blanzieri, Anton Bryl AI review29(1), 2008 In this paper we give an overview of the state of the art of machine learning applications for spam filtering, and different ways of evaluation and comparison of different filtering method. It also provides a brief description of anti-spam protection and discuss the use of various machine learning algorithms and approaches anti-spam software solution.

Andronicus A Akinyelu, Aderemi O Adewumi, 2014 This paper investigates and reports the use of random forest machine learning algorithm in classification of phishing attacks, additionally it as a major objective that is improved phishing email classifier with better accuracy and with less features.

Aery, M., & Chakravarthy, S. (2005). eMailSift: eMail classification based on structure and content. Data Mining, Fifth IEEE Int., 2005. IEEE.This paper gave an approach which is based on the premise that patterns can be extracted from a pre-classified email folder and the same can be used effectively for classifying incoming emails. As emails consists a format in the form of headers and body of the email, the correlation between different terms can be showed in the form graph. Hence there is a graph mining technique for pattern extraction and classification.

Jing Jiang, 2008 Domain adaption of statistical classifier is the problem that arises when the data distribution in our test domain is different from that in our training domain. For example, spam filters can be trained on some public collection of spam and ham emails. But when applied to an individual person's inbox, we may want to "personalize" the spam filter to adapt the spam filter to fit the person's own distribution of emails in order to achieve better performance.

D Karthika Renuka, T Hamsapriya, 2011 The spams or damages in the email are oocured by the increasing amount of the electronic business and financial transactions directly increases in the email threats. Email spam is one of the major problem's, bringing financial damage to companies and annoying individual users.

Ammar Almomani, BB Gupta, 2013 In this paper, we present a survey of the state of the art research on such attacks. This paper discusses the methods of the protection against attacks of phishing email. It also shows the various techniques which is used to detect phishing email.

Islam, R., & Zhou, W. (2007). Email Categorization Using Multi-stage Classification Technique. Parallel and Distributed Computing, Applications and Technologies, 2007. PDCAT '07. Eighth International Conference on, 2007. IEEE.This paper showed a way which proposed a multi-stage classification technique using different popular learning algorithms such as SVM, Naive Bayes and boosting with an analyzer which reduces the False Precision substantially and increases classification accuracy compared to similar existing techniques

Teli, S. P., &Biradar, S. (2014). Effective Email Classification for Spam and Non-Spam. The research paper written showed us a three phased system that they engineered for their way of spam detection.The first phase is a users interacts with html interface and creates the rule for classification. Rules are nothing, but the keywords/phrases that occur in mails for respective legitimate or spam mails. The second phase is a training phase by giving training set. Here the classifier will be trained using a spam and legitimate emails manually by the user. Then with the help of algorithm the keywords are extracted from classified emails.

Although, the work has been done for the spam filter improvement over the years, many of the spam filters today have a limited success because of its dynamic nature of spam. Spammers have been constantly developing new techniques to bypass filters, and also include word obfuscation and statistical poisoning. Although these two text classification issues were recognized, research today it has largely neglected to provide a successful method to improve spam detection and recognition by counteracting word obstruction and Bayesian poisoning, and also many other common spam filters have been unable to detect them.

The Naive Bayes spam detection method is also a supervised machine learning probabilistic model for spam classification technique based on the Bayes Theorem. Supervised machine learning is a method of teaching computers without direct programming, or machine learning, which uses a known data set which are a training set to train the technique to classify whether email is spam or ham, which contains input and response values, to make predictions on another dataset, the testing set. We have preferred Naive Bayes for its speed, multi-class prediction ability, and small training set and also for its accuracy and efficiency. Since Naive

Bayes is the baseline for most of the spam filters, by improving the Naive Bayes it will inevitably improve most spam filters overall.

Although the Naive Bayes classifier has high accuracy in testing, but also it has several major flaws in that surface in real life spam detection as exemplified by increasing the spam. The only way for which spammers can bypass spam filters easily by using tokenization attacks. A tokenization attack disrupts the feature selection process by inserting forms of word obfustication and also the concept of using a non-decrypt able piece of data to represent by reference, sensitive or secret data. The word obfustication includes some different forms like embedding special characters, using HTML comments, character-entity encoding, or ASCII codes.

One of the most important aspects is that although humans are able to discern the actual words, the computer is incompetent. By using leetspeak and diacritics common spam senders are able to bypass the spam detectors. Leetspeak is an informal language used, where the standard letters are replaced by numerals,special characters or symbols.

Diacritics is a sign, such as an accent, when it is written above or below a letter that makes a difference in pronunciation from the same letter when it is unmarked. Leetspeak allows the spam senders to change the letters or words into some numerals or special characters. For example, “WELCOME” can be written as “\/\/3|C0M3”. When the words which are presented in the email are modified using leetspeak and diacritics, then the spam detectors fail to identify or predict the email as spam, which results in a false positive. Bayesian poisoning is a technique used by e-mail spammers to attempt to degrade the effectiveness of spam filters that rely on Bayesian spam filtering.

Additionally, there is a speculation about the existence of Bayesian Poisoning because it always requires the knowledge of which words has to be considered as ham. However, we have seen that in both a passive attack and an active attack, when the spammers come to know which words are ham words, then the performance of Naive Bayes decreases gradually. Bayesian poisoning results in a spam emails incorrectly classified as ham and shows high false negative rate. Naive Bayes makes the assumption as that all the feature vectors are independent to one another, and also Naive Bayes will fail to detect if certain words or phrases are related.

A keyword based statistical method like Naive Bayes mostly depends on the strict lexical correctness of the words. The Naive Bayes algorithm is also unable to detect all forms of word obfuscation and also fails to detect the emails which contain leetspeak and diacritics that are spam. Minimal research has been done on the forms of word obfuscation. One previous research paper written showed us a three phased system that they engineered for their way of spam detection. In the starting phase, the user firstly creates the rule and checks for the classification. Rules are nothing, but the keywords/phrases that occur in mails for respective legitimate or spam mails.

The second phase is a training phase by giving training set to it. Here the classifier will be trained using a spam and legitimate emails manually by the user. Then with the help of algorithm the keywords are extracted from classified emails. According to this paper, in sender based detection, the email sender information such as the writing style and the email sender user name is used as the major features.

However, although the efficiency of the Spam-Assassin and Naïve bayes was approximately 20%, so the study failed to detect for the word which contains leetspeak and diacritics and also the word boundary or optimization. The research in proposed is by using both the Naïve Bayes and Hidden Markov Model (HMM) method for email classification.

III. IMPLEMENTATION

PROPOSED SYSTEM

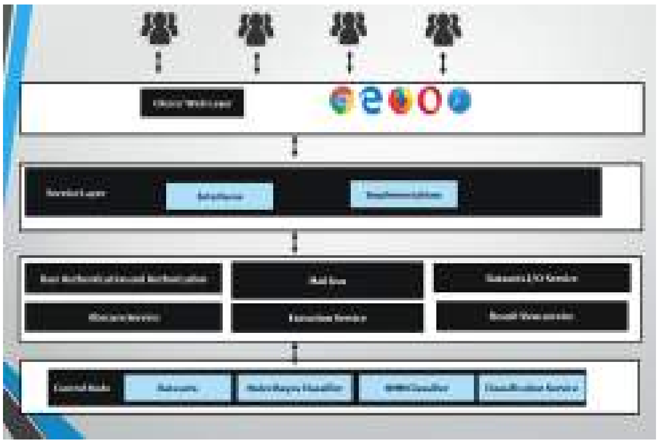


Fig 1.0: System archcitecture for proposed system

List of modules

Module 1: User Authentication and Authorization

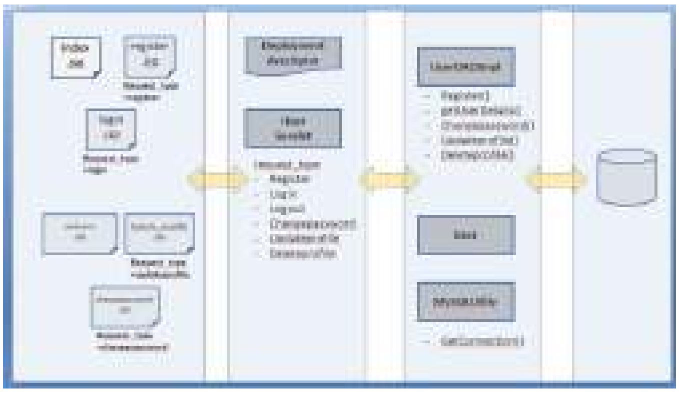


Fig 1.1: User's account operations

Here, the end users of the project will be provided with an interface where they can request access to our project. They will be provided with an HTML interface using which they can create a new account by providing the required information (like email ID, phone number, first name, last name, etc). The user's registration is subject to approval from the project administrators. Once the administrators approve a user registration, the user can proceed with logging in to our portal, perform various account operations (edit profile, change password, etc). After this, the users will be granted access to rest of the modules of the project.

Module 2: Data I/O Operations

Here, the end users having access to the project will be provided with an interface using which he/she can perform various file input-output operations (like upload, download, and delete) on the Hadoop Distributed File System location which is configured in module 2. The solution presented here will be a thread-based solution to enable transfer of multiple files concurrently to improve the speed of the operation.

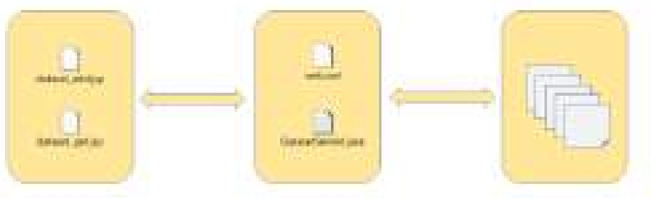


Fig 1.2: Data set operations

Module 3: Control node configuration

Here, the users can specify which control node to be used for performing the core operations (email classification) of the project. That said, the solution proposed will be capable of getting installed at multiple nodes individually and can operate independently from each other. This gives an added advantage to our project that, when there are multiple instances of the servers processing millions of client's request, the load from these clients can be evenly distributed across multiple control nodes. This configuration will be made in the module.

Module 4: Obscure process

This component will perform the hash operation on the input dataset using MD-5 algorithm. The hashed datasets are called obscured datasets. The obscured datasets are then given to run component to execute the algorithm. Hence, the email classification component, either HMM or NB, will not have an access to the user's plain text. This way, we preserve the privacy of the sender.

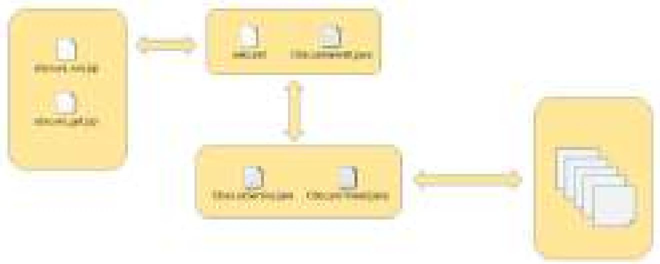


Fig 1.3: Obscure operations

Module5: Classification Service Here, we will be developing an intelligent solution for classifying the emails into "spam" and "not spam" using couple of algorithms – Naive Bayes approach and Hidden Markov Model. The solution will then decide which of these approaches an ideal candidate for the given datasets is. The output of the more accurate of the two solutions will be presented to the users.

Module 6: Result View service
This module presents a user with the output of the email classification algorithm which is developed in the previous module. The user will be provided with an improved visual representation which is easy even for a layman to perform analytics kind of operations.

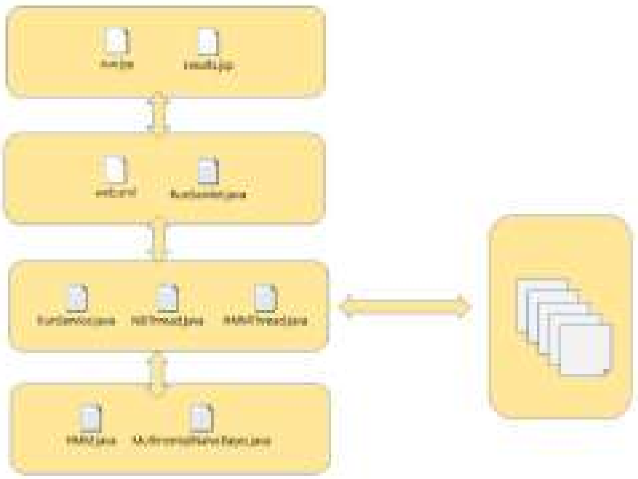


Fig 1.4: Run and Results

Advantages of proposed system

- Filtering of email into Important and Spam efficiently.
- Improves the existing Naïve's Bayes algorithm and blocks the hackers.
- Enabling privacy preservation.

IV. METHODOLOGY

. The above diagram indicates the use case of the design and implementation of efficient email spam detection enabling privacy preservation. Firstly a user has to have an account. In this not only one user many users can have the account. Firstly user has to create and register the account by entering his details like name, mobile number, email id, password, gender and address.

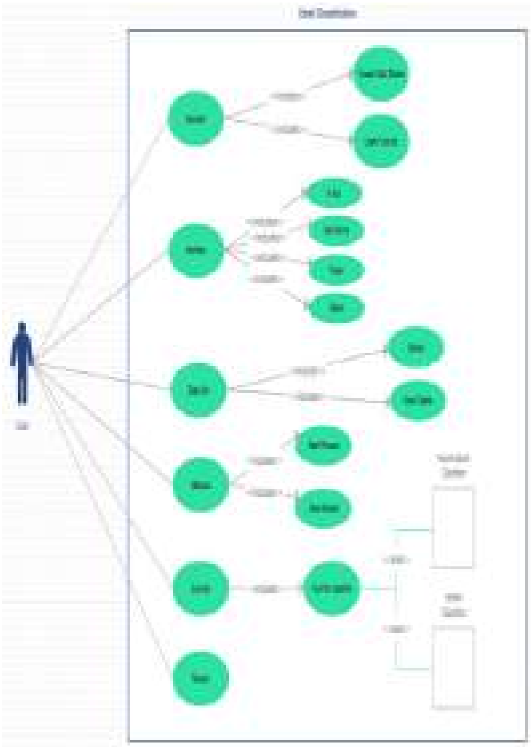


Fig 1.5:Use Case diagram for proposed system

Mailbox which includes inbox, sent items, trash, spam etc. In the dataset option the end-user uploads the email (datasets) and check whether it is spam or ham. Here we can view, delete and download the datasets.
So after uploading the email we have to check whether it is spam or ham. So, before that here we have an obscure option for enabling the privacy preservation. The function of obscure is for enabling the privacy preservation here the email

which is in the form of plain text is converted to cryptographic cipher text. So, the hackers cannot see or hack the contents of the sender as it is in the cryptographic cipher text form. So, in this proposed system it as an advantage of enabling the privacy preservation. Here in this obscure it includes start process and view results. In this obscure option it takes the input as the emails (dataset) and the start the process it converts plain text to the cryptographic cipher text. Finally, look output or result of the obscure process. User should able to see the obscure file to get the confirmation that the classifier not using the plain text.

After this it as an execution process. Here execution invokes the classifier and runs the classifier by using both Naïve Bayes and Hidden Markov Model (HMM). Naïve bayes is based on conditional probabilities. It runs fast and works with small dataset and gives accurate result. It considers independent words as a feature. HMM is a generative, probabilistic model. The state is not directly visible but output (in the form of data or token) dependent of state is visible. The sequence of tokens generated by HMM gives information about sequence of states. In HMM it uses various combinations of NLP (Natural Language Processing) techniques that is stop words removing, stemming, lemmatizing. Stopwords removing are (commonly used words like the,a,an,in) that search engine has programmed to ignore. Stemming is the process of reducing word to its word stem that affixes to suffixes and prefixes or to root of words. Lemmatizing is normally aiming to remove inflectional endings only and return base or dictionary form of a word. It is being tried on both the algorithms to inspect the differences in accuracy as well as to find the best method among them.

V. CONCLUSION

In this paper we have proposed both the Naïve Bayes and Hidden Markov Model algorithms for enhancing the accuracy and also enabling the privacy preservation. Both the algorithm are implemented and tested in real-time environment over the internet. We also carried out two main types of spam encryptions: leetspeak and diacritics to overcome it. By using both the algorithms Naïve Bayes and Hidden Markov Model it improves the accuracy of the spam server by coding new addition to the current server. By using these algorithms not only increases the accuracy of email sorting but also enabling the privacy preservation.

VI. FUTURE ENHANCEMENTS

The concept can be used in “Design and implementation of efficient email spam detection enabling privacy preservation”, Naïve bayes and Hidden Markov Model. The future enhancement is done using other algorithms like clustering and ANN, it can increase in speed, accuracy and efficiency of the system.

REFERENCES:

[1] Enrico Blanzieri, Anton Bryl, “A survey of learning based technique of email spam filtering”, Artificial Intelligence review 29(1), 63-92, in 2008.

[2] Andronicus A Akinyelu, Aderemi O Adewumi, “The use of random forest machine learning algorithm in classification of phishing attacks”, journal of applied machine learning, 2014.

[3] Aery, M., & Chakravarthy, S. (2005). eMailSift: eMail classification based on structure and content. Data Mining, Fifth IEEE Int., 2005. IEEE.

[4] Jing Jiang, “A literature survey on domain adaptation of statistical classifiers”, URL: http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey 3, 1-12, 2008.

[5] D Karthika Renuka, T Hamsapriya, M Raja Chakkaavarthi, P Lakshmi Surya, “Spam Classification based on supervised learning using machine learning techniques”, 2011
International Conference on Process Automation, Control and Computing, 1-7, 2011.

[6] Ammar Almomani, BB Gupta, Samer Atawneh, A Meulenberg, Eman Almomani, “A survey of phishing email filtering techniques”, IEEE communications surveys and tutorials 15(4), 2070-2090, 2013.

[7] Islam, R., & Zhou, W. (2007). Email Categorization Using Multi-stage Classification Technique. Parallel and Distributed Computing,

Applications and Technologies, 2007. PDCAT '07. Eighth International Conference on, 2007. IEEE.

[8] Teli, S. P., &Biradar, S. (2014). Effective Email Classification for Spam and Non-Spam. International Conference on Process Automation, and Software Engineering.

[9] Klimt, B., & Yang, Y., Boulicaut JF., Pedreschi D. (eds) Machine Learning: ECML 2004, ”A new dataset for Email Classification”. Springer, Berlin, Heidelberg. Lecture Notes in Computer Science.

[10] Bhat, V. H., Malkani, V. R., Shenoy, P. D., Venugopal, K. R., & Patnaik, L. M. (2011). Classification of email using Behavior and keyword stemming. TENCON 2011 - 2011 IEEE Region 10 Conference, Bali, 2011. IEEE.

[11] Wang, X., & Cloete, I.A.N. (2005). Learning to classify email: a survey. International conference on machine learning and cybernetics, 2005.

[12] Yitagesu1, M. E., &Tijare, M. (2016). Email Classification using Classification Method. International Journal of Engineering Trends and Technology (IJETT).

Survey on Eye Tracking For Password Authentication

Mala B M [1] Pavithra A [2] Ranjeetha J [3] Yamuna R [4] & Smt. Manikantha. K[5]
[1,2,3,4]Student BNMIT, Department of Computer Science & Engineering, VTU
[5]Assistant Professor BNMIT, Department of Computer Science & Engineering, VTU

Abstract: The personal identification numbers (PINs) is a common user authentication method for many applications, such as money transactions in automatic teller machines (ATMs), unlocking personal devices and opening doors. Authentication remains a challenge even when user enters a PIN in open or public places, makes PIN entry vulnerable to password attacks, such as shoulder surfing as well as thermal tracking. The use of PINs is especially true for banking applications where the combination of a token (e.g. bank card) and the user's secret PIN is commonly used to authenticate transactions. In financial applications PINs are typically four-digit numbers, resulting in 10000 possible numbers. The security of the system relies on the fact that an attacker is unlikely to guess the correct PIN number and that the systems (e.g., Automated Teller Machines) limit the user to few attempts (e.g., 3) for entering the correct PIN. To overcome shoulder surfing attacks, and enable users to enter their PIN without fear of being observed by developing a system that employs an eye tracking device. With safety PIN, users select PIN numbers with their eyes by simply focusing on the digits displayed on a screen. Gazed-based authentication refers to finding the eye location across sequential image frames. Haar Cascade algorithm is a machine learning approach which can be used for detecting the eye pupil location.

Keywords: Personal identification numbers (PINs), Eye centre, Pupil location, Gaze-based PIN entry, Haar Cascade.

1. Introduction.

The use of PINs (personal identification numbers) as passwords for authentication is ubiquitous nowadays. This is especially true for banking applications where the combination of a token (e.g. bank card) and the user's secret PIN is commonly used to authenticate transactions. In financial applications PINs are typically four-digit numbers, resulting in 10000 possible numbers. The security of the system relies on the fact that an attacker is unlikely to guess the correct PIN number and that the systems (e.g., Automated Teller Machines) limit the user to few attempts (e.g., 3) for entering the correct

PIN. As most applications that use PINs for authentication operate in a public setting a common attack is to try to observe and record a user's PIN entry (shoulder-surfing).

These security problems have been recognized for a long time and researchers have proposed a number of different schemes to minimize the risk of PIN entry observation. One such proposed alternate PIN entry method requires the user to input some information, which is derived from a combination of the actual PIN and some additional information displayed by the system, instead of the PIN itself. Another approach proposes the use of an elaborate hardware to make PIN entry resilient to the observation attacks. However, these methods have not been introduced into practical applications because the users would have to be retrained to use a completely different approach to PIN entry and the significant additional costs involved in the hardware setup.

Interaction with computers is not limited to keyboards and printers anymore. Different kinds of pointing devices, touch-sensitive surfaces, high-resolution displays, microphones, and speakers are normal devices for computer interaction nowadays. There are new modalities for computer interaction like speech interaction, input by gestures or by tangible objects with sensors. A further input modality is eye gaze which nowadays finds its application in accessibility systems. Such systems typically use eye gaze as the sole input, but outside the field of accessibility eye gaze can be combined with any other input modality. Therefore, eye gaze could serve as an interaction method beyond the field of accessibility. One of the security requirements for general terminal authentication systems is to be easy, fast and secure as people face authentication mechanisms every day and must authenticate themselves using conventional knowledge-based approaches like passwords. But these techniques are not safe because they are viewed by malicious observers who use surveillance techniques such as shoulder-surfing (observation user while typing the password through the keyboard) to capture user authentication data. Also there are security problems due to poor interactions between systems and users. As a result, the researchers proposed eye tracking systems, where users can enter the password by looking at the suitable symbols in the appropriate order and thus the user is invulnerable to shoulder surfing. Eye tracking is a natural interaction method and security systems based on eye movement tracking provide a promising solution to the system security and usability.

The most of eye-tracking systems work on video-based pupil detection and a reflection of an infrared LED. Video cameras became cheap over the last few years and the price for a LED is negligible. Many computer devices come already with built-in camera as, such as mobile phones, laptops and displays. Processor power still increases steadily and standard processors are powerful enough to process the display on video streaming which gives freedom for the user to move in front of their displays, are also video based. Such systems can be implemented unobtrusively with a second camera. If produced for the mass market, a future standard eye tracker should not cost much more than an optical mouse or webcam today.

Some people interact with the computer all day long, for their work and in their leisure time. As most interaction is done with keyboard and mouse, both using hands, some people suffer from overstressing particulars parts of their hands, typically causing a carpal tunnel syndrome. With a vision of ubiquitous computing, the amount of interaction with computers will increases the needs of interaction techniques which do not cause physical problems. The eyes are a good candidate because they move anyway when interacting with computers. Using the information lying in the eye movements could save some interaction, in particular hand-based interaction.

The paper describes the following sections:

Section II gives the general methods for eye tracking. Section III gives the conclusion.

2. Methods Influencing Authentication Mechanism of Eye Tracking.

The study of existing security systems that are based on eye movement tracking developed by different researchers according to their area of expert. In the following paragraphs are given several of the published researches related to the goals of this work.

A. Pin-entry against Human Shoulder-Surfing:

In computer security, shoulder surfing refers to using direct inspection techniques, such as peeping over someone’s shoulder, to acquire information. Shoulder surfing is frequently used to acquire passwords, PIN security codes and related data. To stop shoulder surfing, which is between the customer and the system, cryptographic prevention approach is hardly relevant because users are restricted in their capacity to process information. Among them, the PIN entry technique introduced was effective because of its clarity and instinctive in every round, a structured numeric keypad is colored at odd half of the keys are in black and another half in white, which is called as BW method. A customer who knows the accurate PIN digit can enter the color by pressing the distinct color key below. The primary BW method is targeted to withstand a human shoulder surfing attack. [1]

B. Gaze-Touch Pass Scheme:

With mobile devices enabling ubiquitous access to sensitive information, there is a need to protect access to such devices. Meanwhile, authentication schemes are prone to shoulder surfing attacks, where a bystander observes a user while authenticating. The attacker then gets hold of the device and tries to authenticate and access sensitive data. To overcome this attacks Gaze Touch Pass, a multimodal authentication scheme in which user define four symbols, each can be entered either via touch (a digits between 0 and 9) or via gaze (gazing to the left and to the right). Consecutive gaze inputs to the same direction would then need to be separated by a gaze to the front and switches between input modalities are used within a single password. [2]

C. Eye Gaze Classification for iTyping :

Human emotions and cognitive states are essential in developing a natural human-computer interaction system (HCI). Systems which can identify the affective and cognitive states of humans can make the interaction more natural. The knowledge of mental processes can help computer systems to interact intelligently with humans. Estimation eye gaze direction is useful in various human-computer interaction tasks. Knowledge of gaze direction can give valuable information regarding user’s point of attention. A real time framework which can detect eye gaze direction using low-cost cameras in desktops and other smart devices. Estimation of gaze location from webcam often requires cumbersome calibration procedure. Gaze direction classification as a multi-class classification problem, avoiding the need for calibration. The eye directions obtained can be used to find the EAC and thereby infer the user’s cognitive process. The information obtained can be useful in the analysis of interrogation videos, human-computer interaction, information retrieval, etc.

D. To Enhance iTyping Privacy:

Mobile devices offers the most convenient user experience ever, e.g., at anytime and anywhere, but users unavoidably face a new potential threat at the same time. The interaction between users and mobile devices may be exposed to public directly, which may leak very sensitive information of the user, e.g., passwords, private data, account information, etc. If the input of such information is not properly protected, the user's privacy can be easily emanated and compromised in public. iTyping is for entering the private information using the eye gaze. In iType, the keyboard consists of multiple buttons and each button represents unique character(s) (number or letter). For the ease of presentation, it refers password to various kinds of private information for short. To type a password, the user looks at the corresponding buttons sequentially, and iType essentially solves a decoding puzzle it reads the user's gaze, infers the buttons being looked at, and assembles the password. The iTyping is secure primarily due to the fact that the eye gaze is difficult to eavesdrop. Even an adversary in front of the user could decode the eye gaze, the gaze itself conveys no meaningful information, unless it matches with the keyboard layout, which however can be user-defined and changed. [4]

E. Accuracy and Precision of Eye Tracking:

To establish eye gaze as part of everyday interaction with computers, it needs to understand the characteristics and limitations of eye tracking in practice and derive standards for the design of gaze-enabled applications that take into account that accuracy and precision can vary widely across different tracking conditions. It collects calibration style eye tracking data from 80 participants, using two different trackers in two lighting conditions. In contrast to many eye tracking studies, It does not exclude any participant due to insufficient tracking quality and only calibrated once at the beginning. Finding a several contributions for the design and development of gaze-enabled applications.

1. Checking the accuracy and precision ranges overall and for different parts of the screen that characterize the large variation across different tracking conditions.

2. Provides a formal way to derive appropriate target sizes from measures of accuracy and precision. Based on data user gives a recommendation for the minimum size of gaze-aware regions to allow robust interaction.

3. An approach to find optimal parameters for any filter that minimizes target size and signal delay. [5]

To overcome all the above issues by implementing a real time hands-off gaze-based PIN entry technique is used, which leaves no footprints. Gaze-based authentication refers to finding the eye location across sequential image frames and tracking eye centre. Haar Cascade algorithm is a machine learning approach used for detecting the eye pupil location by Image Processing. In this technique, several stages are used to find out the movement of eye, such as Face detection and Eye detection, Edge detection. The distance between the centre point and eye pupil centre are measured using coordinates system logic. According to the eye pupil movements, the measured distance will vary. A minimum distance indicates the eye pupil is moved towards the left, and maximum values indicates the eye moved on right and if there is no movements of the eye, then it concludes that eye is in the middle position. After tracking the eye pupil position, the data is taken by the system. The entered data is compared with the trained dataset. If the entered data is not matched with the trained data then the system will throw an error as “unauthorized access” else the system confirms and allows for further transaction.

3. Conclusion

In order to protect the users from shoulder-surfing in ATMs while entering the PIN, new method of entering the PIN are being evaluated. With the eye tracking technology becoming easier, eye interaction for PIN entry is emerging as a practical solution. It has been discussed about Safety PIN, which proposes retrofitting the ATMs with an eye tracking device, so that users can enter their PIN without using keypad for pin entry. In addition to look and shoot the gaze activation method called blink activation. Initial user evaluations have yielded encouraging results, prompting further work.

4. References

[1]. R. Revathy and R. Bama, “Advanced Safe PIN-Entry Against Human Shoulder Surfing,” IOSR Journal of Computer Engineering, vol 17, issue 4, ver.II, pp. 9-15,July,Aug.2015.(http://www.iosrjournals.org/iosrjce/papers/Vol17issue4/Version2/B017420915.pdf).

[2]. Mohamed, Florian, Mariam, Emanuel, Regina and Andreas “Gaze-Touch Pass Scheme”, March 2016.

[3]. Anjith George, Aurobinda Routray "Real time Eye Gaze Direction Classification Using Convolutional Neural Network", June2016.

[4]. Zhenjiang Li1, Mo L, Prasant Mohapatra, Jinsong Han, Shuaiyu Chen

"iType Using Eye Gaze to Enhance Typing Privacy", 2017.

[5]. Puja sorate, Prof. Mrs. G. J. Chhajed

"Survey Paper on Eye Gaze Tracking Methods and Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN.

Sentiment Analysis of Twitter Corpus Using Artificial Intelligence Assistants

Rakshitha B[1],Nandini G[2],Pratiksha M Javali [3],Priyanka N Murthy[4] , Ramya N[5]
Assistant Professor[2],Student[1,3,4,5]
Department of Computer Science, Rajarajeswari College of Engineering[2]

Abstract— **The necessity of improving the quality of life and an enhancing experience is the significant research in recent days. It is required to evaluate usability and emotion to improve a user's experience. Sentiment analysis which is used to understand user's tendency is based on their opinions. User's opinion were collected from Twitter and classified into positive, negative, neutral opinions by lexicon named Valence Aware Dictionary and sEntiment Reasoner (VADER).**

Keywords— Twitter, sentiment Analysis, VADER, sentiment scores.

I. INTRODUCTION

The main aim of a product or a service is to provide a better experience to the user. This experience can be enhanced by the research on understanding and observing the user and obtaining the feedback from the users [1]. The experience can be evaluated by focusing on the user's usability and emotion. User's opinions are generally applied to evaluate emotion [2] [3].

Sentiment analysis is a kind of big data mining. Researchers use the data from social media such as Instagram, Facebook, and Twitter, which has appeared as user's subject opinions. Hence, sentiment analysis is known as opinion mining. This allows the researcher to understand the user more accurately. User opinion depends on user's values, conditions or interests. Each user has their own different sentiment. By investigating the user tendency the best experience can be provided.

In this paper, the data from twitter, a text based social media service is collected. Tweets are classified into positive, negative and neutral opinions using Valence Aware Dictionary and sEntiment Reasoner (VADER). VADER converts the opinions into sentiment scores. Each opinion which is quantified to document matrix demonstrates statistical significance between sentiment groups.

The paper is composed of six sections. Section 2 consists of related works from earlier studies to present day about sentiment analysis. Section 3 describes the design and defining of the architecture, modules and components of the system. Section 4 shows consists of the procedure to collect the data from Twitter Corpus and classifying them using lexicon. Section 5 consists of the snapshots and the experimental trials of the proposed system. The final section is about the results and summary and about the further research.

II. RELATED WORKS

According to Chae Won Park and Dae Ryong Seo, providing a significant experience is one of the current issues in the user's research in Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants. A process should improve the user's experience by evaluating their emotions and usability [1].

Action rules are data mining method for gaining information as well as knowledge from the datasets. The sub-actions are called Meta actions. The action rules for Sentiment Analysis on Twitter Data process new optimized, speed and efficient system for generating meta- actions. Meta actions are generated by implementing Specific Action Rule discovery based on Grabbing strategy (SARGS)
algorithm [2].

The social networking sites generate a huge amount of data. An approach was proposed by V. Sahayak, V. Shete, S.A.Bahrainian, A. Dengel which classifies the sentiment of tweets automatically taken from the twitter datasets. This is useful for company as well as the customers to know about the brand, its quality and feedback from the other customers [3][18].

A novel unsupervised approach is used in the Sentiment Analysis of Twitter Data using sentiment Influencers for the analysis for the twitter data based on a rule based scoring engine. Here, parts of speech of the sentence are ranked according to the influence of the sentiment of the sentence [4].

The datasets are pre- processed after the extraction from the twitter in Sentiment Analysis of Twitter Data using Machine Learning Approach and Semantic analysis. Later the adjectives from the dataset are extracted that have some meaning called feature vector. For the selected feature vector list machine learning algorithms such as Naïve Bayes, Maximum entropy and SVM algorithms are applied [5].

VADER model is used for general sentiment analysis and also it is compared to eleven benchmarks like LIWC, ANEW, SentiWordNet, and machine learning techniques in Sentiment Analysis of Social Media text by E. Gilbert, CJ Hutto, Federico Neri, Carlo Aliprandi, [6][17].

Sentiment Analysis and Opinion Mining is based on predicting sentiments to extract opinions from the internet and predict online customer’s preferences, which would help in economic and marketing of the company [7].

Supervised Learning and Volume based measures are used by B.O’Connor, R BalaSubramanyan, A.Bermingham and Smeaton.A [8][9] for the purpose of Sentiment Analysis. They evaluated against the conventional election polls and final election result to find the Social analytics.

For the task of Sentiment Analysis A.Pak,P.Paroubek, B.Pang and Lee.L[10] [11] have focused on the most popular microblogging platform,Twitter. For the purpose of Sentiment Analysis and Opinion Mining the corpus are automatically collected. Using the Corpus a Sentiment classifier is built which can determine the sentiments for a document.

III. DESIGN OF SYSTEM

The process of defining and designing the architecture, components, modules, interfaces, and data for a system to satisfy requirements is System requirements. Systems design is the process of defining and deploying and designing systems to satisfy specified requirements of the user.

The entire architecture has been implemented in eight modules. The below figure shows a general block diagram describing the activities performed by this project.

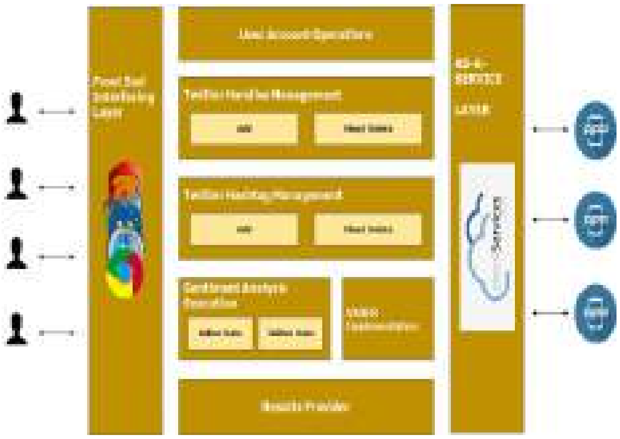


Fig.1 System Design

A. *Data Access Layer*

Data access layer is the one which exposes all the possible operations on the data base to the outside world. The internal components consists of Dao classes, interfaces, POJO’s and Utils.All the other modules will be communicating with the DAO layer for their data access needs.

B. *Account Operations*

Account operations module provides the following operations to the users of the project.

1. Register a new seller/ buyer account
2. Login to an existing account
3. Logout from the session
4. Edit the existing Profile
5. Change Password for security issues
6. Forgot Password and receive the current password over an email
7. Delete an existing Account

The Account operations use DAO layer to provide back the above functionalities.

C. *Twitter Handle and Twitter Keyword*

Here, the end users will be provided with an HTML interface where they can add as many twitter handles or twitter hashtags they want for which the sentiment analysis operations need to be carried out. The inputted twitter handle and the hashtags will be stored in the database which will be retrieved later while performing the sentiment analysis operation

D. *VADER Algorithm*

VADER is used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. Lexical approaches map words to sentiment by building ‘dictionary of sentiment.’ This dictionary can be used to access the sentiment phrases and sentences. Sentiment can be categorical – such as {negative, neutral, positive} – or it can be numerical – like a range of intensities or scores.

E. *Adhoc Run*

This component is developed using the HTML interfacing components to display the results of the VADER algorithm on the specified input text (temporary text) in a beautiful styling in the browser of the client who has requested for the same. This component will be developed in a responsive fashion so that it works consistently across various devices and browsers.

F. Twitter Data Analysis

This application is also implemented using java. This application accepts the twitter handler as an input and makes use of Twitter 4J libraries for communicating with twitter and pulls the data (posts) from that twitter account and executes the core Emotion Recognition Algorithm against each of the posts. It later invokes a utility function to decide the category of this emotion. Possible emotion categories are positive, negative, and neutral.

G. As a Service Implementation

This component is useful particularly when an external application has to make use of the solution proposed by us in this project. As part of this component, we will be exposing a REST (Representational State Transfer) API to the outside world to consume our solution. The external application will have to provide the input text which needs to be analyzed for the sentiment and this API returns the result back to the application with the calculated results.

H. Service Layer

Service Layer is a J2EE implementation of web Tier which processes the requests from various clients. This service layer will contain servlets, the configuration files, and the helper classes. The servlets is a server side programming construct in Java used to receive the requests from the HTML clients, forward the request to the appropriate business logic components and get the results back to the HTML clients. To find out the appropriate business logic component for each request, the servlet will have to communicate with the configuration file which is the deployment descriptor (web.xml) of the J2EE application. The deployment descriptor will have a mapping from the incoming request (URL path) to the business logic component. The helper classes are the ones which acts as an interfacing agents between the service layer components and the VADER algorithm components.

IV. METHODOLOGY

The experimental procedure consists by collecting the Twitter corpus and classify the data by using the lexicon method. Then after classification process, the percentage of positive, negative and neutral opinions are extracted from the lexicon method. This can be applied for the purpose of user's research.

A. Corpus

Microblogging today has become a famous communication tool among Intertxt users across the world. Many users share their opinions on different aspects on daily basis. So, microblogging sites are rich sources of content for opinion mining and sentiment analysis. One of the most popular microblogging services across the world is Twitter.

Twitter is a platform for the extraction of public opinion on specific issues in society. The number of daily active users are 126 million as of February 2019. Twitter offers the organizations a fast and effective way to analyze the customer's opinion on the products available in the market.

Tweets of the users are collected by Twitter API for the analysis process. The twitter tweets length are limited to 140 characters. So, the users have limited space to express their thoughts. The sentiments in the twitter tweets are usually incisive, uncomplicated and consistent in polarity.

The tweets have a length limitation, so the users use emotions to make their tweets more expressive.

B. Lexicon

Lexical approach aims to map the words to corresponding sentiment by building a lexicon or a 'dictionary of sentiment'. We use this to map the sentiment of phrases and sentiment, without looking anything else. Sentiment can be categorical i.e., negative, neutral, positive or it could also be numerical –like are scores or the intensity values.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and it is a rule- based sentiment analysis tool used to express sentiments of social media data. It is a open –source used the MIT License. IT is introduced in 2014, Vader text Sentiment analysis uses a human –centric approach, by combining qualitative analysis and empirical validation by using human raters.

VADER classifies the tweets related to artificial intelligence assistants into three categories, which are positive, neutral, negative words. If in a tweet there were no positive or negative word then it is considered to be neutral. However, the proportion of words that consisted of positive, negative or neutral could be expressed in a format of percentage.

V. RESULTS

First the user has to create an account by providing all the required credentials. Later he/ she can login in to the account with the required credentials. A message pops up when the user has successfully logged in. The user can modify their account by providing the required credentials. The user can change the password of the account, or logout of the account or delete the account.

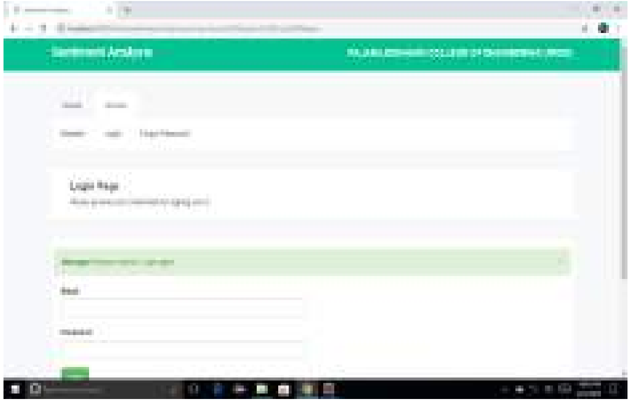


Fig.2 Login page

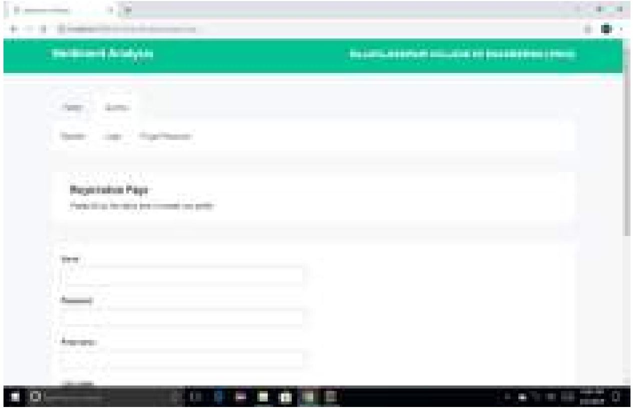


Fig.3 Registration page

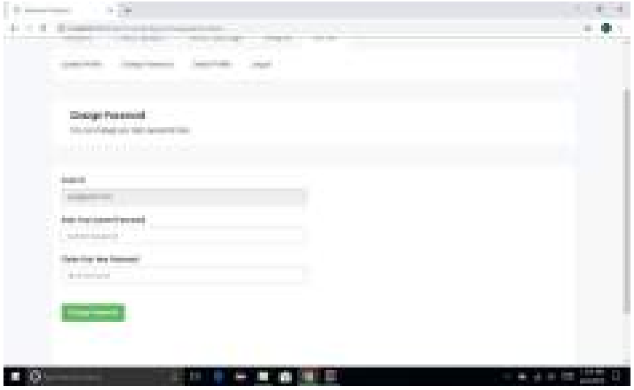


Fig. 4 Change password

After the user has logged into the account they can add the Twitter handles and the Twitter hashtags. The user can register a new Twitter handle for which the sentiment analysis have to be run. Twitter handles can be added, viewed along with its date and time, and can also be deleted.

Fig.5 Twitter Handle

Similarly, Twitter Hashtags can also be added, viewed and also deleted.

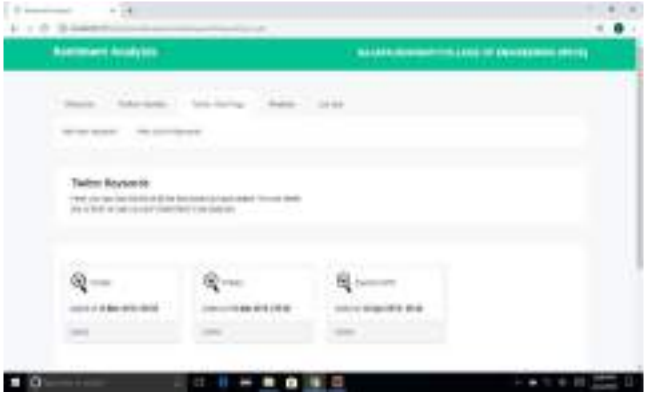
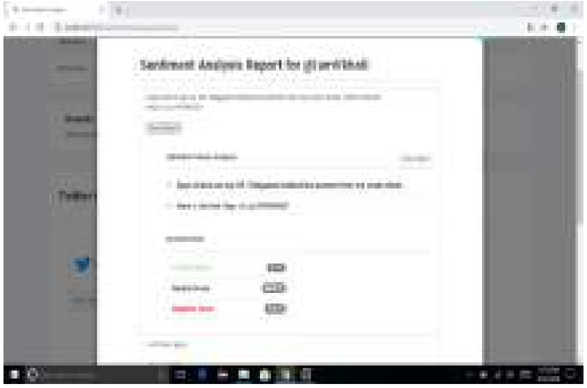


Fig.6 Twitter Hashtag

Sentiment analysis can be executed in two ways. One is Ad-Hoc method, where the user can run the VADER algorithm for a set of textual data which can be obtained from the other sources as well. The sentimental scores can be obtained for each and individual sentences or for the whole set of provided data.

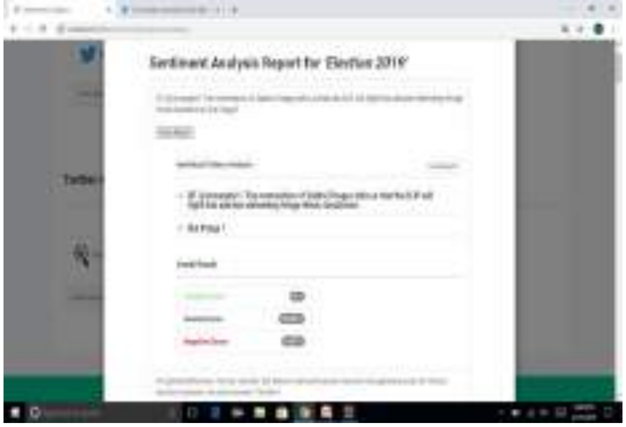
Fig.7 Individual token analysis

Other method of sentiment analysis is the run the sentiment analysis algorithm. Here, the VADER algorithm is implemented for the Twitter Handles and Hashtags. It provides the sentiment score for the recent 100 tweets of that particular handles/ hashtags. The individual scores for each tweet can also be obtained along with the overall score.



Sentiment Analysis Report for @[illegible]

Fig.8 Analysis for Twitter Handle



Sentiment Analysis Report for 'Election 2019'

Fig .9 Analysis for Twitter Hashtag

VI. CONCLUSION

In this project, we have analyzed the tweets, the tweets were collected using Streaming API and divided into positive, negative, and neutral opinions by VADER, the sentiment dictionary. A change of sentiment score was described as positive, negative, and neutral percentage.

In future, we would be working towards applying our algorithm to non-textual data as well like image, videos, and other multimedia.

REFERENCES

[1] Chae Won Park, Dae Ryong Seo ,"Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants", presented at the 2018 5th International Conference on Industrial Engineering and Applications.

[2] Jaishree Ranganathan ,Allen's , Irudayaraj and Angelina A. Tzacheva, "Action Rules for Sentiment Analysis on Twitter Data", IEEE International Conference on Data Mining Workshop,2017.

[3] Varsha Sahayak , Vijaya and Apashabi Pathan, "Sentimental Analysis on twitter Data", International journal of Innovative Research in Advanced Engineering, 2015.

[4] Munazza Ishtiaq ,"Sentimental Analysis of Twitter Data using Sentimental Influences", Journal of Intelligent Computing Volume 6 Number 1,2015.

[5] Geetika Gautam and Divakar Yadav,"Sentimental Analysis of Twitter Data using Machine learning Approach and Semantic Analysis ", International conference on Artificial Intelligence ,2014

[6] C.J Hutto and Eric Gilbert on, "VADER: A Parsimonious Rule based Model for Sentimental Analysis of Social Media Text", 2014.

[7] G. Vinodhini and R.M. Chandrashekaran, "Sentiment Analysis and opinion mining: A Suvey", International Journal, Vol 2, 2012.

[8] B. O'Connor, R.Balasubramanyan, B.R. Routledge, and N.A Smith, "From tweets to polls: linking test sentiment to public opinion time series", Proc.The Fourth International AAAI Conference on weblogs and Social Media, ICWSM, 2010.

[9] A. Bermingham and Smeaton.A, "On Using Twitter to monitor political sentiment and predict election results", Workshop on Sentiment Analysis where AI meets Psychology, 2011.

[10] B. Pang and Lee.L, "Opinion mining and sentiment analysis,"Foundations and Trends in Information Retrieval, 2008.

[11] A.Pak and P.Paroubek, "Twitter as a Corpus for Sentimental Analysis and Opinion Mining", 4th International Conference on Industrial Engineering, 2010

[12] M.Kuniavsky, "Observing the user experience: A Practitioners guide to user research", M.Kaufmann, 2003.

[13] E. Boiy, P.Hens, K.Deschact and M.F.Moens, "Automatic Sentiment Analysis in On-line Test", ELPUB, Jun 2007,pp.349- 360.

[14] S.M.Mohammad and S.Kiritchenko, "Using HashTags to capture fine emotion categories from tweets", Computational Intelligence, Vol.31,No.2,pp.301-326,2015

[15] Z.Niu, Z.Yin, X.Kong, " Sentiment Classification for Micro blog by Machine Learning", in Computational and Information Science (ICCIS), 2012 4th International Conference, pp. 286-289, IEEE,2012

[16] Y.Wu and F.Ren, “Learning Sentimental influence in Twitter”, 2011.

[17] Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cundroe, “Sentiment Analysis on Social Media”, proceedings of 2012 International Conference on Advance in Social Networks Analysis and Mining, 2012

[18] S.A.Bahrainian, A. Dengel, “Sentiment Analysis and Summarization of Twitter Data”,,2013.

Association Rules to Predict the Consumers Behaviour in Large Shopping Malls

K K Kavitha
Assistant Professor,
Dept. of ISE, New Horizon College of Engineering, Bengaluru,
kkkavitha@newhorizonindia.edu

Abstract— **Predicting consumer's location and motion is very important in a large shopping mall to provide them better service. When a consumer passes regions of a shopping mall, his/her moving trace can be recorded for prediction. Existing approaches cannot be directly used to fulfill such task because handling the ordered region sequences is quite challenging. System Proposes an improved Apriori algorithm called AprioriOS (Apriori for Ordered Sequences) to solve this problem. Using this method, association rules are mined out from ordered region sequences and then used to predict future locations of consumers. System can predict more than one region that a consumer may pass in future and System also proposes a motion prediction method based on Wi-Fi Positioning System to predict motions of consumers. Based on the proposed Idea System develops a location and motion prediction system for shopping malls.**

Keywords—Association Rule Mining, Apriori, Location & Motion Prediction.

I. INTRODUCTION

Every industries main purpose is to provide a best service to the customers, and one of the industries which are growing in urban areas is shopping mall so it is important to predict consumer's location and motion in a large shopping mall to provide them better service. When a consumer passes regions of a shopping mall, his/her moving trace can be recorded for prediction [6]. To provide better services for consumers and promote the sales of shops, it is important to predict two things. One is the regions that a consumer will walk in. The other is the motions of consumer will perform (e.g. slowly walking, quickly walking, standing up, sitting down). If we can predict locations and motions of a consumer, we can know which things catch his/her eyes and then recommend products or services to them.

II. RELATED WORK

Existing approaches related to above mentioned problem [2] cannot mine relations from ordered sequences of regions that consumers have gone, and hence cannot effectively predict locations of consumers. It is highly demanded that we can obtain the location prediction result in the form of subsequence of consecutive regions like "Region 3 → Region 5 → Region 1 → Region 2".

Moreover, the prediction of motion is not combined with the prediction of locations. Indoor positioning methods [3]are used to trace the move of consumers and generate an ordered sequence of regions that consumers passed. The outliers of positioning data are removed to avoid data distortion. Then we mine ordered region sequence set using AprioriOS to get association rules. The Algorithms proposed for association rule mining were based on sequential pattern mining techniques on the data bases. In today's world, when we are dealing with Big Data the existing algorithms [4] lag behind because of the time and computational complexity.

Data Mining has various techniques [5] such as clustering, classification, Naïve Bayes, Association rule mining. Association Rule mining is one of the most well-known technique for discovering relationships between the items in data sets. Association rules are being used widely in various areas such as telecommunication networks, risk and market management, inventory control, medical diagnosis/drug testing etc. The Association rule mining focuses on the frequent item-set. It's also important to understand the rare association algorithms are also equally important to capture a rare scenario. The document focus on the rare and interesting association rule mining.

Defense Mechanism	**Functionality**	**Limitations**
Markov-based crowdsourced mobility prediction	This approach characterizes the close relationship between the human mobility patterns	difficulty in differentiating human motilities among the relative small areas due to its coarse localization granularity
APFT(Combination of apriori and FP-tree structure	This approach is used to get Frequent item set mining	when the support threshold is high this combination is not so productive
Novel approach for indoor localization	this method is used in WiFi-based indoor localization and navigation	when the space is too large like a large shopping mall this method cannot be applied
Density-based Clustering Algorithm	This method is used to obtain wi-fi finger prints which will help in motion prediction	This method is less sensitive to outliers[11]
Similarity-based location predictions and adaptive	This mechanism is used to distinguish	This algorithm predicts the location in

background modeling approach,	betting moving humans and relocated objects also it is used to predict future location of customer	macroscopic level than predict users next location on a microcosmic level.
Hidden Markov Model	This approach is used in Human motion prediction for human-robot collaboration	Only a small part of distributions over the space of possible sequence can be represented
BFSPMiner	This approach is used to get sequential pattern mining	This method is only accurate in batch stream which is not continuous not in the bulk non batched data
Closed weighted sequential mining	This approach is used to find the sequential pattern on time series	Weighted sequential mining is only applicable in Sequential mining

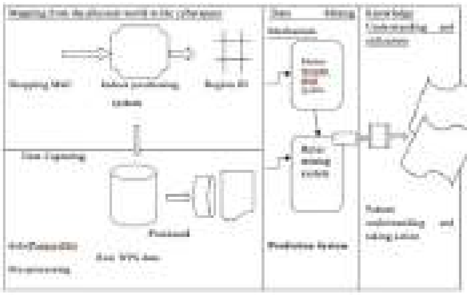
III. System Architecture

The indoor positioning system [10] provides the set of ordered region sequences and the location data of a consumer to association rule mining system and motion recognizing and recording system, respectively. Motion recognizing and recording system (motion system for short) is a central server which collects motion data from phone sensors of consumers. Then it analyses motion data to generate quater-naries (time window length, region id, consumer id, motion tag). There are two databases, association rule database and Consumer's motion database. Association rules and quaternaries are stored in them, respectively. The final quaternary (time length, region id, consumer id, motion tag) will be generated using quaternaries in database and stored back into the database. The prediction system uses the data of two databases to do location and motion prediction.

The system architecture is broadly categorized as follows:

a. *Mapping from the physical world to the cyberspace*
b. *Data Capturing & Pre-processing*
c. *Data Mining Mechanism*
d. *Knowledge Understanding & Utilization*

Figure: System Architecture



Proposed Algorithm

Function: getAllFrequentOrderedRegionSequences *(H_C)*

```
LS_1 = {All 1 frequent sequences of Consumer C};
for (k = 2;LS_{k−1}=      ; k++) do
    C_K= candidate-gen(LS_{k−1});
    foreach LH_C do
        foreach cd_kC_k do
            if Lcontains subsequencecd_k then
                cd_k.count++;
                end
        endfch
    endfch
    LS_k= {cd_kC_k | cd_k.count ≥ minsup};
end
return    _kLS_k;
```

The above algorithm shows the process of mining out all frequent ordered region sequences and shows the pseudo code of function candidate-gen(). The main difference between our AprioriOS and original Apriori is in candidate-gen(). The original Apriori needs two prune steps to delete non-frequent K itemsets from candidate frequent K itemsets. It first examines whether any K-1 sub-item set of a candidate frequent K itemset is frequent or not, then calculates the support of this candidate and deletes it if its support is less than *minsup*. But our variant only needs one step.

Mathematical model

To explain about the motion function [22] we have taken a mathematical model which will take the region id as the input and also take the Centre of the area

$$hq = \text{EXP}\left|\frac{\sum_{i=1}^{1}(xi-uqi)2}{2\sigma_{q2}}\right|$$

Where: hq : the output of the qth motion function.
$\boldsymbol{x}$: vector that represents the network inputs ($\boldsymbol{x}$ = [$x1$, $x2$, … , xI]).
$\boldsymbol{uq}$: vector that represents the center of the qth basis function ($\boldsymbol{uq}$ = [$uq1$ $uq2$ $uq3$ … . uqI]).
σq: spread or Gaussian width of the qth motion function.

Result and Analysis

The indoor locating system [13] offers the set of ordered region sequences and the position data of a customer to association rule mining system and motion recognizing and recording system, respectively. Motion recognizing and recording system (motion system for short) is a central server which collects motion data from phone sensors of consumers. Then it studies motion data to generate quaternaries (time window length, region id, consumer id, motion tag). There are two databases, association rule database and customer motion database. Association rules and quaternaries are stored in them, respectively. The final quaternary (time length, region id, customer id, motion tag) will be produced using quaternaries in record and kept back into the database. The forecast system uses the data of two databases to do location and motion prediction.

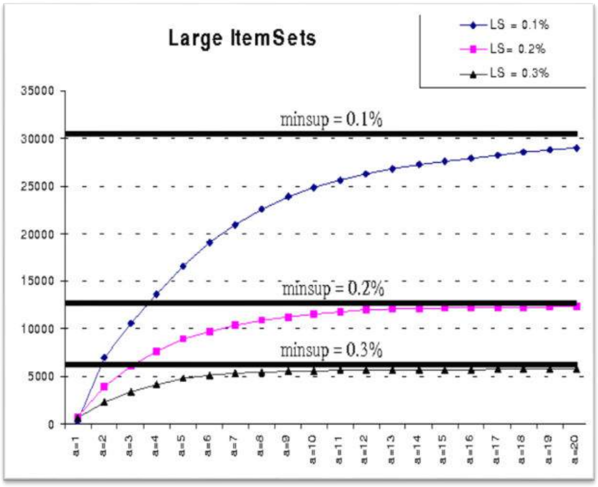


Fig 1:Different Minimum Support

To prove the validity of prediction system [13], we design a shopping simulation of a consumer who has a specific shopping habit. The shopping simulation is showed in Fig. 1. In Fig. 1, a circle with a number is a region of the shopping mall. Region 1 and region 2 are two entrances of this shopping mall. And region 3 and region 4 are two exits of this shopping mall. The red path and the dark blue path are two main paths the consumer often passes through. Paths with other colours are those that the consumer not often passes through but actually does. The number of paths is totally 100 for simulating shopping activities that basically happen about two times a week in one year. For the convenience of the presentation of experiment results, we only use two motion types in this simulation. One motion type is not moving, and the other motion type is walking. The meeting-point region, region 8, region 9 and region 15 are attractive to the consumer, so we add the motion of not moving into the synthetic simulation data of them. Therefore, in experiment results, the motion tag of not moving is much more likely to appear in those three regions. The final results of simulation test are showed in Fig. 1.

The simulation results show that the location prediction is very accurate because the actual results of location prediction are all the same with the expected results except the second experiment that only one future region can be predicted because of the insufficient confidence.

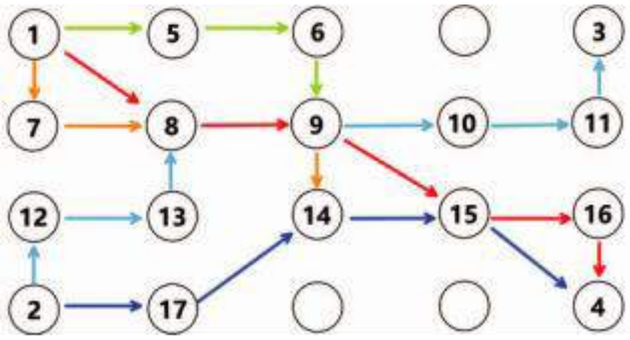


Fig 2: Habitual shopping paths of a consumer in one large shopping mall

The simulation results show that the location prediction is very accurate because the actual results of location prediction are all the same with the expected results except the second experiment that only one future region can be predicted because of the insufficient confidence. For 8→9, when the number of regions for forward prediction (forward prediction region number) is 1, the system considers region 15 as the future region with confidence 0.610. Because red path 8 →9 is a main path, the consumer is more likely to reach region 15 along the red path 9 →15

When the forward prediction region number is up to 2 for path 8 →9, due to the low confidence that is not greater than minsup, it can only predict one region not two regions in future. So the prediction result is still region 15.

For sequence of recently passed regions 8 →9 →10, because the consumer has passed through the light blue path 9 →10, the system then considers the light blue path 10 →11 →3 as the result with confidence 0.987 when the the forward prediction region number is 2.

To test let me take a customer passed regions 13 →8 →9, the system does not give out a wrong result of location prediction, region 15, through the red main path, but successfully give out a correct result, region 10, because the consumer has passed through a light blue path from region 13 to region 8 and he/she is more likely to pass through a light blue path again from region 9 to region 10.

IV. Conclusion & Future Scope

System used the indoor positioning[1] method to generate ordered region sequences, and use AprioriOS to mine out association rules [2] from them to do location prediction of consumers in large shopping malls. A special tree structure for storing association rules is specifically designed to meet the need of the location prediction method. The WPS or the RFID method can be used to recognize motions of a consumer and then the system stores them into the database as historical data for later motion prediction. The whole location and motion prediction

System can effectively predict the future path of a consumer in a large shopping mall, and predict the motions that the consumer may have in corresponding regions. The AprioriOS is aimed to mine ordered and consecutive sequences of regions and also can be used to mine other ordered and consecutive sequences in other work.

Indoor/Outdoor IPS. Outdoor positioning systems will merge with IPS in a seamless way to locate a person with a

Smartphone anywhere. This means that while current IPS systems involve specialized equipment and applications, future IPS systems will be part of the Smartphone operating system and leverage its sensors so any location-sensitive Smartphone application will use indoor or outdoor location services as they are available.

Consideration of Privacy and Security Issues in the Development of IPS. From the analysis of IPS, we noticed that the privacy and security issues regarding the user's location are only addressed in very few projects [7, 8]. Nevertheless, some authors provide evidence that these factors may influence the adoption and use of the IPS [12, 10] or argue that the system must give the users the possibility of deciding whether they want to share their locations with others [9]. Though privacy has been a concern since the very beginning of the development of IPS systems, in the future, this will become one of the main considerations for the adoption or choice of specific IPS systems.

REFERNCES

[1] Evolution of Indoor Positioning Technologies: A Survey Ramon F. Brena,1 Juan Pablo García-Vázquez,2 Carlos E. Galván-Tejada,3 David Muñoz-Rodriguez,1 Cesar Vargas-Rosales,1 and James Fangmeyer Jr.1 pp 56-59

[2] Fang Miao, Lu Bai, "Study on the association and aggregation of fusion media content for TV services", Computer Communication and the Internet (ICCCI) 2016 IEEE International Conference on, pp. 220-223, 2016.

[3] HadisKakanejadiFard, Yuanzhu Chen, Kyung Kook Son, "Indoor positioning of mobile devices with agile iBeacon deployment", Electrical and Computer Engineering (CCECE) 2015 IEEE 28th Canadian Conference on, 2015, ISSN 0840-7789, pp. 275-279.

[4] ChitreshVerma, Rajiv Pandey, "Big Data representation for grade analysis through Hadoop framework", Cloud System and Big Data Engineering (Confluence) 2016 6th International Conference, pp. 312-315, 2016.

[5] ShivamAgarwal"Data Mining: Data Mining Concepts and Techniques",Published in:Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference on

[6] Zeyu Zhang,Weiping Zhu:Location and Motion Prediction of Consumers in a Large Shopping Mall published in 2017 Fifth International Conference on Advanced Cloud and Big Data,pp.250-254,2017

[7] Jian-Qiang Li, F. Richard Yu, Genqiang Deng, Chengwen Luo, Zhong Ming, Qiao Yan, "Industrial Internet: A Survey on the Enabling Technologies Applications and Challenges", Communications Surveys & Tutorials IEEE, vol. 19, no. 3, pp. 1504-1526, 2017.

[8] X. Zhao, Z. Xiao, A. Markham, N. Trigoni, and Y. Ren, "Does BTLE measure up against WiFi? A comparison of indoor location performance," in Proc. Eur. Wireless Conf., May 2014, pp. 1–6.

[9] D. Lymberopoulos et al., "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned," in Proc.IEEE/ACM IPSN, Apr. 2015, pp.178–189

[10] C. Praneeth Kumar, Ravi Poovaiah, Ajanta Sen, Priya Ganadas, "Single access point based indoor localization technique for augmented reality gaming for children", Students' Technology Symposium (TechSym) 2014 IEEE, pp. 229-232, 2014.

[11] Amik Singh, Mohit Gupta, Manoj Misra. Parallel Progressive Sequential Pattern (PPSP) Mining.

[12] D. Lymberopoulos et al., "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned," in Proc.IEEE/ACM IPSN, Apr. 2015, pp.178–189

[13] M. Paciga and H. Lutfiyya, "Herecast: an open infrastructure for location-based services using Wi-Fi," in Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '05), vol. 4, pp. 21–28, August 2005

Structure Extraction from Texture via Relative Total Variation

Sindhu R[1], Sneha S[1], Thejaswini C[1], Prakruthi M P[1] , Rajath A N[2]
[1]Student, Dept. of CSE, GSSSIETW, Mysuru, Karnataka, India
[2]Assistant Professor,Dept. of CSE, GSSSIETW, Mysuru, Karnataka, India

***Abstract*-It is universal that important structures are shaped by or show up over finished surfaces. Removing them under the confusion of surface examples, which could be customary, close normal, or unpredictable, is testing, yet of incredible pragmatic significance. We propose new intrinsic variety and relative absolute variety measures, which catch the fundamental contrast of these two kinds of visual structures, furthermore, build up an effective enhancement framework to separate fundamental structures. The new variety measures are approved on a large number of test patches. Our methodology finds various new applications to control, render, and reuse the huge number of "structure with surface" pictures and illustrations that were customarily troublesome to be altered appropriately.**

Keywords: texture, structure, inherent ariations, smoothing, relative total variation.

I. INTRODUCTION

Numerous common scenes and human-made craftsmanship pieces contain surface. For example, graffiti and illustrations can be generally observed on block dividers, railroad freight cars, and metros; floor coverings, sweaters, and other fine creates contain different geometric examples. In mankind's history, mosaic has for quite some time been be an artistic expression to speak to nitty-gritty scenes of individuals and creatures, and copy artworks utilizing stone, glass, clay, and different materials. While looking in Google Images, a large number of such pictures and illustrations can be found rapidly. A couple of precedents from various sources are appeared in Figure 1.

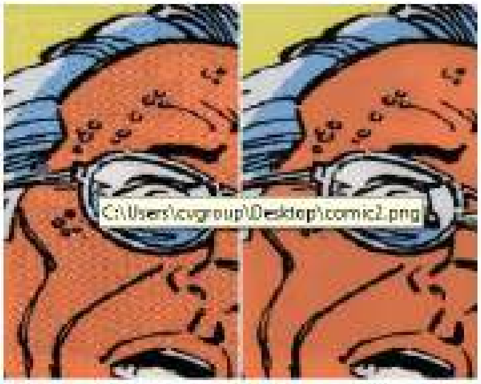


Figure 1: Significant structure extraction from texture.

They share the closeness that semantically important structures are mixed with or shaped by surface components. We call them "structure+texture" pictures. It is especially fascinating that human visual framework is completely able to comprehend these photos without expecting to expel surfaces. In brain science , it is likewise discovered that "the generally basic highlights are the essential information of human discernment, not the individual subtleties".

We present a straightforward but then successful technique dependent on novel nearby variety measures to achieve surface evacuation. We found that with respect to our new relative all out variety, which will be expounded later in this paper, surface and principle structure display totally unique properties, making them shockingly decomposable. With this finding, we present an advancement structure, in which significant substance and textural edges are punished in an unexpected way. A hearty numerical solver is additionally proposed to decay the first very non-curved advancement issue into a few direct frameworks, for which quick and powerful arrangement exists. Note that we don't accept specific normality or symmetry of the surface examples, and rather take into consideration an abnormal state of irregularity. Non-uniform and anisotropic surface, in this manner, can be dealt with in a unified structure.

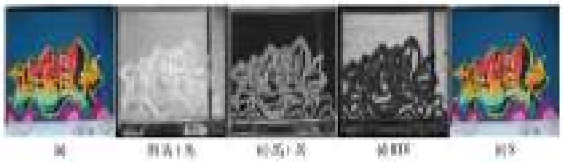


Figure 2: Effect of our variation measures. (a) Input. (b) Windowed total variation map. (c) Windowed inherent variation map. (d) Relative total variation (RTV) map, where meaningful structures are penalized much less than textures. (e) Our finally extracted structure image.

II. PROPOSED SYSTEM

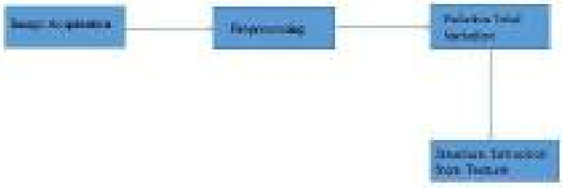


Figure 3: Block Diagram of Proposed System

Surface as a rule alludes to surface examples that are comparable in appearance and nearby measurements. Surface combination can deliver a huge consistent surface guide from little precedents. For close ordinary surfaces, spatial relationship is utilized to identify and investigations normality, empowering picture surface detachment in de-fencing. These strategies rely on the symmetry and consistency of surface and require earlier example learning. Agent structure-surface deterioration techniques that don't require broad surface data are those implementing the all out variety (TV) regularizer to protect substantial scale edges. To confirm the adequacy of the Relative Total Variation (RTV) measure, we construct a dataset, which contains a large number of patches alongside physically made names. In the first place, we gather 200 "structure with surface" pictures and request that five understudy assistants attract strokes snapping to critical structure edges. The rest of the pixels are treated as not containing significant changes. So each picture has a comparing stroke map. Ongoing edge-protecting picture altering devices don't plan to take care of a similar issue, and, accordingly, are not ideal arrangements. We present a basic but then successful strategy dependent on novel nearby variety measures to achieve surface expulsion. We found that concerning our new relative all out variety, surface and fundamental structure display totally unique properties, making them shockingly decomposable. With this finding, we present a streamlining system, in which significant substance and textural edges are punished in an unexpected way. A strong numerical solver is additionally proposed to break down the first exceedingly non-raised advancement issue into a few straight frameworks, for which quick and hearty arrangement exists. Our strategy makes a colossal number of existing "structure with surface" pictures reusable in altering and rendering. We present a few applications, including basic edge location, vectorization, consistent cloning, and structure just picture arrangement, to give some examples. Our framework additionally benefits general crease cutting, making the outcomes less blunder inclined to universal surfaces.

Our final objective isn't surface/structure classification, yet rather another difficult assignment, i.e., surface expulsion from various "structure with surface" pictures. The variety metric should be easy to frame a pragmatic answer for finely separate surface and structure from one another for every pixel.

III. APPROACH

We don't expect or physically decide the kind of surfaces, as the examples could change a ton in various models. Our strategy contains a general pixel-wise windowed absolute variety measure, composed as.

$$\mathscr{D}_x(p) = \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_x S)_q|,$$
$$\mathscr{D}_y(p) = \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_y S)_q|, \qquad (1)$$

Where q has a place with R(p), the rectangular area focused at pixel p. Dx(p) and Dy(p) are windowed all out varieties in the x and y headings for pixel p, which check indisputably the spatial distinction inside the window R(p).gp, q is a weighting capacity defined as indicated by spatial affinity, communicated as

$$g_{p,q} \propto \exp\left(-\frac{(x_p - x_q)^2 + (y_p - y_q)^2}{2\sigma^2}\right), \qquad (2)$$

Where σ controls the spatial size of the window. In a picture with notable surfaces (Figure 2(a)), both the detail and structure pixels yield vast D (Figure2 (b)), which shows that the windowed absolute variety is receptive to visual saliency.

To help recognize conspicuous structures from the surface components, other than D, our strategy likewise contains a novel windowed characteristic variety, communicated as

$$\mathscr{L}_x(p) = \Big|\sum_{q \in R(p)} g_{p,q} \cdot (\partial_x S)_q\Big|,$$
$$\mathscr{L}_y(p) = \Big|\sum_{q \in R(p)} g_{p,q} \cdot (\partial_y S)_q\Big|. \qquad (3)$$

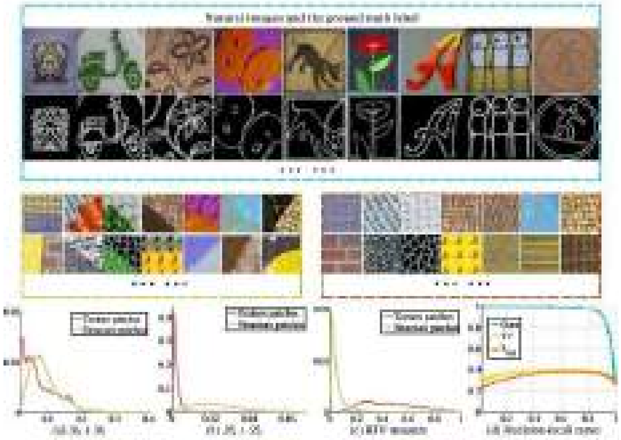


Figure 3: *Statistical validation of the inherent and relative variations. A few examples along with manually labeled main structures are shown in the top frame. Sample patches are randomly extracted from these examples, as illustrated in the middle row. Structure and texture patches are included in the red and green rectangles respectively. (a)-(c) show the distributions of D, L, and RTV. The red and green curves are for the classes of texture and structure patches respectively. (d) plots precision-recall of our normalized measure and other variations.*

L catches the by and large spatial variety. Not the same as the articulation in Eq. (1), it doesn't fuse the modulus. So the entirety of ∂ S relies upon whether the slopes in a window are incidental or not, as far as their headings, in light of the fact that ∂ S for one pixel could be either positive or negative.

IV. NUMERICAL SOLUTION

We propose an efficient solver dependent on the information that a target work with the punishment of a quadratic measure can be upgraded directly. Our methodology deteriorates the RTV measure into a non-direct term and a quadratic term. The favorable position is that the issue with the nonlinear part, intriguingly, can be changed to settling a progression of direct condition frameworks, in a path like iterative re-weighted least squares.

We first talk about the x-course measure. The y-heading term can be managed comparably. We grow the punishment as

$$\sum_p \frac{\mathscr{D}_x(p)}{\mathscr{L}_x(p)+\varepsilon} = \sum_p \frac{\sum_{q\in R(p)} g_{p,q}\cdot|(\partial_x S)_q|}{|\sum_{q\in R(p)} g_{p,q}\cdot(\partial_x S)_q|+\varepsilon}. \quad (4)$$

By re-organizing the terms and grouping elements that contain |(∂ x S)q|, we obtain

$$\begin{aligned}\sum_p \frac{\mathscr{D}_x(p)}{\mathscr{L}_x(p)+\varepsilon} &= \sum_q \sum_{p\in R(q)} \frac{g_{p,q}}{|\sum_{q\in R(p)} g_{p,q}\cdot(\partial_x S)_q|+\varepsilon}|(\partial_x S)_q| \\ &\approx \sum_q \sum_{p\in R(q)} \frac{g_{p,q}}{\mathscr{L}_x(p)+\varepsilon}\frac{1}{|(\partial_x S)_q|+\varepsilon_s}(\partial_x S)_q^2 \\ &= \sum_q u_{x\,q} w_{x\,q} (\partial_x S)_q^2\end{aligned} \quad (5)$$

Algorithm 1 RTV for Structure Extraction from Texture

```
1: **input:** image I, scale parameter σ , strength parameter λ
2: **initialization:** t =0, S0 ←I
3: **for** t=0:2 **do**
4: compute weights w and u in Eqs. (6), (7), (9), and (10)
5: solve the linear system in Eq. (14)
6: **end for**
7: **output:** structure image S
```

The second line in (5) is an estimate because of the presentation of ε s for numerical strength. The re-game plan of the terms decays the measure into a quadratic term (∂ x S)2 question and answer non-straight part ux q, wx q. They are individually

$$u_{x\,q} = \sum_{p\in R(q)} \frac{g_{p,q}}{\mathscr{L}_x(p)+\varepsilon} = \left(G_\sigma * \frac{1}{|G_\sigma * \partial_x S|+\varepsilon}\right)_q \quad (6)$$

$$w_{x\,q} = \frac{1}{|(\partial_x S)_q|+\varepsilon_s}, \quad (7)$$

Articulation (6) demonstrates that ux for every pixel really joins neighboring slope data in an isotropic spatial filter way. G σ is a Gaussian filter with standard deviation σ. The division in (6) is component savvy and ∗ is the convolution administrator. wx is just identified with the pixel-wise inclination.

Similarly, we can express the y-directional penalty as

$$\sum_p \frac{\mathscr{D}_y(p)}{\mathscr{L}_y(p)+\varepsilon} = \sum_q u_{y\,q} w_{y\,q} \left(\partial_y S\right)_q^2, \quad (8)$$

Where (∂y S) 2 q is the quadratic y-component partial derivative and uy q, wy q is similarly the non-linear part. They are respectively

$$u_{y\,q} = \left(G_\sigma * \frac{1}{|G_\sigma * \partial_y S|+\varepsilon}\right)_q \quad (9)$$

$$w_{y\,q} = \frac{1}{|(\partial_y S)_q|+\varepsilon_s} \quad (10)$$

And we can write the matrix form

$$(v_S - v_I)^T(v_S - v_I) + \lambda\left(v_S^T C_x^T U_x W_x C_x v_S + v_S^T C_y^T U_y W_y C_y v_S\right). \quad (11)$$

where vS and vI are the vector portrayal of S and I separately. Cx are Cy are the Toeplitz grids from the discrete slope administrators with forward contrast. Ux, Uy, Wx, and Wy are corner to corner grids. Their corner to corner esteems are separately Ux[i,i]=uxi, Uy[i,i]=uyi, Wx[i,i]=wxi , Wy[i,i]=wyi.

The structure in (11) empowers a unique iterative streamlining system. Because of the deterioration of the non-straight and quadratic parts, a numerically steady guess is normally gotten, which was discovered exceptionally successful in our trials to rapidly gauge the structure and surface pictures. Our advancement procedure is as per the following.

Stage 1: From the evaluated structure picture S in the past emphasis, it is direct to figure the estimations of u and w dependent on Eqs. (6), (7), (9), and (10), which structure the networks of U and W in Eq. (11).

Stage 2: Using the estimations of Ux, Uy, Wx, and Wy, minimization comes down to comprehending a straight framework in every emphasis as

$$(1+\lambda L^t)\cdot v_S^{t+1} = v_I \quad (12)$$

where 1 is a personality network and Lt =CT x Ut x Wt x Cx +CT y Ut y Wt y Cy is the weight framework registered dependent on the auxiliary vector vt S. (1+ λ Lt) is the symmetric positive definite Laplacian grid. We utilize the forward distinction to rough discrete slopes, which results in a scanty five-point Laplacian grid. Efficient solvers are accessible for it. Both the isotropic and anisotropic medications of the all out variety can be connected. The entire streamlining process is abridged in Algorithm 1.

V. CONCLUSION

We have introduced another framework for important structure extraction from surface. Our primary commitment is twofold. To start with, we proposed novel variety measures to catch the idea of structure and surface. We have broadly assessed these measures and infer that they are without a doubt amazing to make these two kinds of visual data distinguishable much of the time. Second, we formed another streamlining plan to change the first non-direct issue to a lot of sub issues that are a lot simpler to fathom rapidly. A few applications utilizing these pictures and illustrations were proposed.

Our strategy does not require earlier surface data. It could, along these lines, botch some portion of structures as surface on the off chance that they are outwardly comparative in scales.

VI. REFERENCES

[1]. Zhigang Shang, Mengmeng Li, “Combined Feature Extraction and Selection in Texture Analysis”, 2016 9th International Symposium on Computational Intelligence and Design.

[2]. Shurong Liu, Kun Han, Zhibin Song, Misheng Li, “Texture Characteristic Extraction of Medical Images Based on Pyramid Structure Wavelet Transform”, 201O International Conference On Computer Design And Appliations.

[3]. Yujing Sun, Scott Schaefer, Wenping Wang, “Image Structure Retrieval via L0 Minimization”, 2016 IEEE.

[4]. Ying Wu, “A method of Pattern Feature Extraction for Clothing Texture”, 2017 International conference on robots and intelligent system.

[5]. Rahmaniansyah, Dwi Putri, Harsa wara prabawa, Yaya Wihardi, “Color and Texture Feature Extraction on Content-based Image Retrieva”, 2017 3rd International conference on Science in Information Technology.

Survey on IDS in Cloud Computing Environment

Bhvya Jyothi A
M.Tech (CNE)
Dept. of CSE, MSRIT
Bengaluru-54
Bhvyajyothi11@gmail.com

Jayalakshi D S
Associate Professor
Dept. of CSE, MSRIT
Bengaluru-54
jayalakshmids@msrit.edu

Abstract— **Cloud computing is delegating data to storage systems that are handled by third party vendors. It can also be considered as a disk or network storage which can be accessible over the network. Microsoft's OneDrive and Apple's very own iCloud are few known examples. Since the client's information is dealt by service provider, utilizing cloud raises privacy and confidentiality concerns because service provider will have access to all the data. And sometimes deliberately or accidentally unveil it or sometimes even use it or disclose it for unauthorized purposes. This work, is an examination on clients concerns and basic investigation about the diverse security models and devices proposed in different papers.**

Keywords – Cloud computing, Ontology, Security issues, IDS.

I. INTRODUCTION

Cloud computing are a set of resources and services which are offered through the Internet. Also, all through the world. its services are delivered from data centres located Cloud computing facilitates its consumers by providing virtual resources via internet. General example of cloud services is Google apps, provided by Google and Microsoft SharePoint. Main obstacle in adopting to cloud is absence of security. And it is bounded by numerous security concerns like authenticating data, and inspecting the utilization of cloud by the vendor. Main reason for all these security challenges for the clients and vendors is because of its boom and wide adaption. This work, is an examination on clients concerns and basic investigation about the diverse security models and devices proposed in different papers.

Based on server virtualization technology, Amazon released its very own cloud computing service. Amazon released Xen-based Elastic Compute Cloud™ (EC2), object storage service (S3) and structure data storage service (Simple DB) [12] in 2006 – 2007, under the name Amazon Web Service™ (AWS) [9]. Because of its OnDemand and user friendliness of AWS, it became the pioneer of Infrastructure as a Service (IaaS) provider. Google's style is based on technique-specific sandbox. A few research papers that were distributed by Google, which concentrated on Platform as a Service (PaaS) [1-5]. Google released to public, its own one of a kind stages as a platform as a service called Google App Engine™ (GAE). Microsoft Azure™ [10] is released in Oct. 2008, which uses Windows Azure Hypervisor (WAH) as the underlying cloud infrastructure and .NET as the application container

II. CHARACTERISICS

The five distinct characteristics of cloud are [1] :
1) On-demand service: Provides services such as email, network, application and computer capabilities as and when the user needs.
2) Rapid elasticity or expansion: The resource pooling and self-service make it possible. The provider can automatically distribute resources from available pool which helps to provide elasticity in cloud.

3) Broad Network access: Utilizing standard mechanisms, processing capabilities can be accessed over the network. Which would then be able to be utilized by diverse thin or thick client platform.

4) Resource pooling: The provider can automatically distribute resources from available pool

5) Measured service Utilizing a metering ability, cloud automatically controls and manage resources.

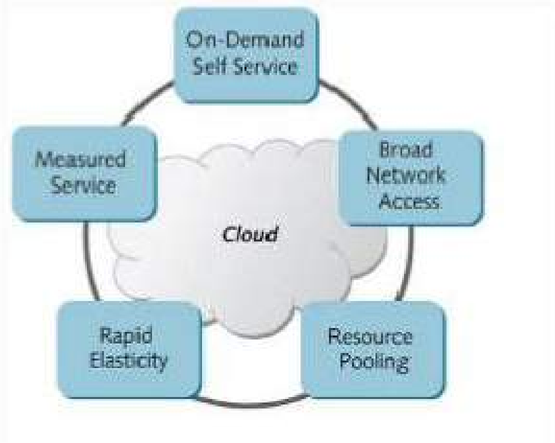


Fig 1. Characteristics of cloud [1]

III. SERVICE MODELS

In view of service models, that are utilized to give ability to the cloud customers, the following is categorized:
1)The Cloud provisions such as the use of basic IT infrastructure such as storage, network and computing capacity are provided via infrastructure as a service (IaaS).

2)The cloud provisions higher than infrastructure level services for the consumer are provided under platform as a service (PaaS). The cloud empowers its clients to design separated applications for their
own individual needs in both a runtime domain (RTE) and an amalgamated development environment (IDE)

3)The cloud gives an integrated service network dependent on the cloud foundation through Software as a Service (SaaS). Along these lines' expenses of software / equipment licenses and upkeep of the IT foundation additionally can be spared by the cloud customers.

IV. DEPLOYMENT MODELS

Cloud computing services are divided as follows into the two pure forms Public Cloud and Private Cloud and a composite form, the Hybrid Cloud, based on their deployment model.
— IT assets or software are generally given by a cloud seller by means of the Web, in Public cloud. A few clients share a similar cloud in this type of cloud. Inside the same physical infrastructure, the client's information and application dwell however they are allocated differently.
— the IT assets or software are tweaked to the necessities of a solitary client and are given only to that client by a cloud vendor, in private cloud. In contrast to public cloud, here complete access is given to singular client for their individual cloud framework.

---- There exists a wide range of blended mixed cloud computing types known as hybrid cloud, between the two, public and private cloud. These consolidate the benefits of both, public and private cloud

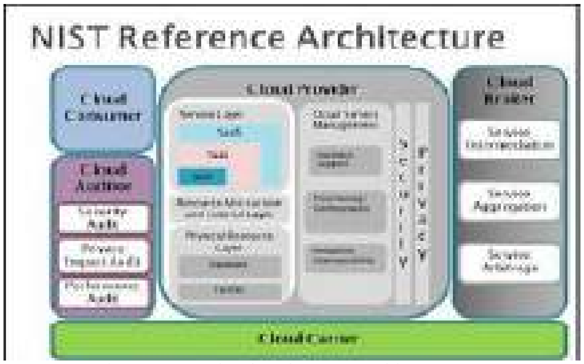


Fig 2. NIST Cloud Reference Architecture [1]

V. TAXONOMY OF THREATS

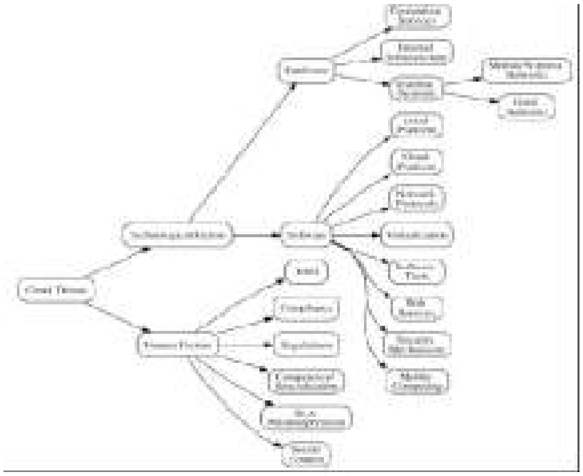


Fig 3. Taxonomy of Threats [4]

Cloud computing threats can be classed as technical (or hard threats) influences and human (or soft threats) influences. Both addressed in the regions underneath.

1) Human influences:

These influences relate to that emerge from human-driven activities that compromise a Cloud framework. Few of these threats might be related with government guidelines for some random district or nation or region, the absence of information security and consistency in some location uncommitted areas, social engineering, shortfall of technical, dimension of certainty between other Cloud partners.

These two human factors are primarily for Cloud vendors and clients,

♦ Compliance

♦ And Competence, are a dynamic security challenge

Off standard methodology might be permitted, and scare tactics to security (inside or remotely) may happen if there is no compliance. All the clients, engineers, and Cloud vendors demonstrate good practice in preventing any unwanted events due to competence of IT professionals.

The utilization of Cloud inside the network is called as social context, which consequently is affiliated with different factors, for example, trust and explicit social engineering human associated security issues, such as, absence of mindfulness or preparing, or absence of watchfulness or alert when utilizing Cloud administrations.

The dull craftsmanship that can represents a serious hazard to confidentiality, integrity, and authenticity of data are social context and related social engineering. And can be demarcated as to gain entree to confidential data by influencing people. The objectives of the assailant can be accomplished, if there is absence of competence in the cloud environment.

SLA provisioning needs careful assessment because of the complexity of Cloud Computing architectures, Also, failures related to Cloud computing can emerge because of misinterpretations of the SLA, which thus influences the Cloud security. Along these lines, commitment must be made to adhere to the SLA provided by the cloud vendors. Contrarily, if commitment is not in place it could lead to ack of monitoring of the SLA. It is important for both the clients and vendors of cloud to know what the SLA states and what it provides.

2) Technological influences:

These refer to threats caused from anything other than human or social influences.

Two categories listed beneath fall into this risk under this category

Numerous software-based security emerges because of Virtualization, for example, Denial of Service, numerous escapades to hypervisor like cross-VM side channel, hypervisor escape

Primarily these days difficulties to security are introduced because of web services, it very well may be as substantial as a cross webpage scripting too little as HTTP vulnerabilities. Indeed, even in cases like loss of AJAX security, SQL infusions can be considered as a noteworthy risk.

The most common means for general and distributed computing nowadays is mobile computing. It has opened doors to Bring Your Own Device. Presently multi day difficulties to security in mobile world is turned out to be a major concern. Application and network-based threats are introduced due to mobile computing.

3) Hardware:

In today's cloud era. like how external factors can offer ascent to threats, similarly there are many inside factors too. Overseeing discrete hardware components is getting progressively mind boggling and is liable to vulnerabilities and exploits. The fundamental purpose behind this can be increase in the unpredictability to deal with these hardware components. Potential security ruptures can emerge because of breaking down of these components, trust, SLA agreements by cloud vendors, certain network protocols, malfunctioning or failing of network components. Other reason can resemble unapproved access to specific assets by support experts, to get to remote cloud assets. Due to redistributing of the cloud administrations to outsiders, unauthorized individual can complete couple of insider assaults

VI. INTRUSION DETECTION SYSTEM

As seen in the previous section, even with Even with the firewalls in place intruder still can attack the network by breaking through it, that is where the intrusion detection system comes in to place by acting as a barrier for unauthorized access control over the network. Intrusion detection systems or IDS functions as the term suggests: they detect possible intrusions It monitors computers or networks primarily for any unauthorized entry activity, or any modification file. Although it can likewise be utilized to screen network traffic, in that way distinguishing if a system is being focused by a system assault, for example, when attempting to place an IDS in a network, an IDS may be viewed as a home - installed burglar alarm system. Although function differently, both sense when an intruder/assailant/thief is available, and in this manner, issue caution signal(s) or alert(s). The benefits of utilizing IDS are to monitor, detect, and react to any unauthorized action.

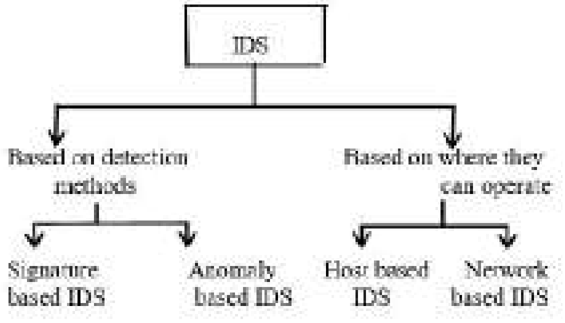


Fig 4. IDS Classification

VII. LITERATURE SURVEY

In [2], to precisely detect potential attacks they developed a system utilizing different strategies like decision free, Random forest and KNN. To conquer the impediment of the past that could not detect IPV6 attacks, this paper proposed new technique. The proposed system alongside recognizing ipv4 based assaults likewise indicated efficient and impressive outcomes. Various parameters such as accuracy of detection, precision, etc. were measured.

In [3] to detect novel anomaly called NEC, clustering and KDD can be proficiently used. To deliver high detection rate and less false passive rate, an unsupervised anomaly is utilized. To find the anomaly which does not need a labelled data set, this is an appropriate way. The NSL - KDD 2009 dataset was used to check the system. Transformation of all attributes into the real number and normalised dataset done by pre-processing model, and the predicated results with the accurate results are compared at the end of the evaluation.

In [4], a survey was conducted on Intrusion detection in data mining and machine learning, to guarantee cybersecurity. To reach networks and kernel level data Packet-header and net flow packet header are used for the instruction detection system. The algorithms and various machine learning complexities are discussed, with respect machine learning/data mining methods Intrusion Detection System. This paper gives a lot of correlation criteria which help to distinguish unapproved utilized, duplication, modification and demolition of the data framework.

In [5], for improved detection and to expand security, propelled strategy was used, for recognizing and discovering traces of the attacker. This new strategy included, machine learning, deep learning, ranking and Voronoi clustering, clustering, various Botnet techniques. It likewise guarantees decreasing the measure of data set and high detection accuracy. "ISOT" new data set has been utilized remembering the handling delay in the largescale network. The paper likewise identifies the botnet intrusions, by utilizing the system stream characteristics, despite parcel attributes of the system

In [6], Utilizing ADS-B strategies, an automatic dependent surveillance-broadcast IDS technique called ADS-IDS. To improve the air traffic controls performance, HMAC data set has been utilized. Among the Current radio, radar benchmarks in flying machine flagging, ADS-B as a superior as it gives superior location accuracy, which are the given by GPS utilizing the cyber-physical environment the attack detection is affirmed. To trade the keys utilized for the HMAC calculation safely, a system is proposed. ATC Centre starts confident handshakes with ATC's that control another zone in the flightpath to exchange the private key over open key framework (PK1) plans.

In [7] Specified an approach using techniques of data mining called hybrid IDS. This automated technique is used to create hybrid Disesteem is a combination of IDS based on both signature and specification. The advantage of this approach is that it detects up to 73 percent, but is not well equipped to capture big data logs in big data, as it can get dubious now and again. There's is no compelling reason to play out any manual analysis as this method is mechanized.

In [8], said one of cyberspace's challenging issues is cybersecurity. Along these lines, in this paper, using deep learning, a new approach to network IDs that uses a neuromorphic cognitive computing was demonstrated. Discrete Vector Factorization is utilized in this strategy. To build exactness and grouping up to 90.12% and 81.31% separately, NSL-KDD dataset is utilized. This paper attempts to accomplish human dimension performance, as it utilizes deep learning, it joins highlights of extraction arrangement.

In, [9], this paper utilizes ANN (Artificial Neural Network), to recognize any anomalies in android based mobiles on android operating system. This strategy depends on flow anomaly IDS., Considering different parameters like CPU, memory and battery power, this system offers up to 85% exactness and recognition rate. Utilizing this work lightweight, adaptable IDs can be distinguished. Utilizing machine learning calculations few examinations is performed

In [10], A Hidden Markow model-based IDS is developed for software-defined networking (SDN), to help in monitoring the security of the system, a Hidden Markow model-based IDS is created. It utilizes ANN IDS. Papers merits are there is a noteworthy increment of the security application. To get to risk in networking environment, machine learning techniques have been utilized. Dynamic control over the network can be accomplished utilizing this procedure.

According to [11], for Classification (GPFIS-Class), this paper uses a new GFS model called Genetic Programming Fuzzy Inference System for Classification (GPFIS-Class). It is based on Multi-Gene Genetic Programming (MGGP). It is not just used in the intrusion detection area yet can likewise be used to do feature selection method, to eliminate data redundancy and irrelevant features to analyse

In [12] Hybrid Cryptography, this reduces the network and routing overhead. This methodology is more powerful and secure than MANET in parameters like End-to-end delay, Battery life, Connectivity, Network capacity.

In [13] Uses real time discrete event system for PS-Poll DOS attack in 802.11 networks. To detect DOS attack, this approach uses RTDES on real time discrete event system. Even though loss of frames is a major setback, its high accuracy and detection rate tries to balance that. Encryption change in protocol is required to detect the PS-DOS attack.

In [14] discussed a Cluster-Based Intrusion Detection Framework for Monitoring the Traffic of Cloud Environments. This framework monitors network traffic. Few advantages are Telecommunication traffic, Monitoring, Intrusion detection, Virtual machining,

According to [11], for Classification (GPFIS-Class), this paper uses a new GFS model called Genetic Programming Fuzzy Inference System for Classification (GPFIS-Class). It is based on Multi-Gene Genetic Programming (MGGP). It is not just used in the intrusion detection area yet can likewise be used to do feature selection method, to eliminate data redundancy and irrelevant features to analyse

In [12] Hybrid Cryptography, this reduces the network and routing overhead. This methodology is more powerful and secure than MANET in parameters like End-to-end delay, Battery life, Connectivity, Network capacity.

In [13] Uses real time discrete event system for PS-Poll DOS attack in 802.11 networks. To detect DOS attack, this approach uses RTDES on real time discrete event system. Even though loss of frames is a major setback, its high accuracy and detection rate tries to balance that. Encryption change in protocol is required to detect the PS-DOS attack.

In [14] discussed a Cluster-Based Intrusion Detection Framework for Monitoring the Traffic of Cloud Environments. This framework monitors network traffic. Few advantages are Telecommunication traffic, Monitoring, Intrusion detection, Virtual machining,

OBSERVATION

Based on various studies conducted by different authors, show different IDS approaches and algorithms through which intruder can be detected are surveyed in this paper. It is observed that in paper [7] [8] [14], shows 90.12%. has precision and identification rate. Another paper [5] in largescale network there is a certain amount of processing delays. From [4] we can observe that based on real time, intruders and their activities can be detected. So, when a new IDS are proposed, many such attributes may be used to recognize internal intruders in the real - time system. Based on real time, this study can be utilized by MNC's and different associations for securing their significant information by distinguishing the intruders precisely.

ACKNOWLEDGMENT
I would like to convey my gratitude for the support, guidance and research that my mentor, Jayalakshmi D.S, has given me in every way to research various IDS approaches.

REFERENCES

[1] Final Version of NIST Cloud Computing Definition Published. Available online http://www.nist.gov/itl/csd/ cloud-102511.cfm (accessed on 03 April 2015).

[2] Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, “a novel unsupervised Anamoly detection Approach for Intrusion Detection System”, 2017 IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.

[3] Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, “a novel unsupervised Anamoly detection Approach for Intrusion Detection

System”, 2017 IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.

[4] https://www.researchgate.net/figure/Threat-taxonomy-for-Cloud-Computing_fig1_308691801(accessed on 04 March 2019)

[5] Manoj s. Koli, Manik K. Chavan, “An Advanced method for detection of botnet traffic using Internal Intrusion Detection”, 2017 International Conference on (ICICCT), March 10-11, 2017, Sangli, India.

[6] Thabet Kacem, Duminda Wijesekera, Paulo Costa, Alexander Barreto, “An ADS-B Intrusion Detection System”, 2016 IEEE on ISPA, 2016, Fairfax, Virginia

[7] Shengyi Pan, Thomas Morris, Uttam Adhikari, “Developing a Hybrid Intrusion Detection System using Data Mining for power system”, IEEE Transactions on, vol. 6, issues. 6, Nov. 2015.

[8] Mehdi Ezzarii, Hamid Elghazi, Hassan El Ghazi, Tayeb Sadiki, "Epigenetic Algorithm for performing Intrusion Detection System",2016 International Conference on ACOSIS, Oct17-

[9] Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece

[10] Trae Hurley, Jorge E. Perdomo, Alexander Perez-pons, "HMM-Based Intrusion Detection System for software defined networking", 2016 15th IEEE Conference on Machine Learning and Application, Dec 18-20, 2016, Miami, Florida.

[11] Mariem Belhor, Farah Jemili, "Intrusion Detection based on genetic fuzzy classification system", 2016 IEEE 13th International Conference on Computer Systems and Application (AICCSA), Nov 29, 2016-Dec 2, 2016, Sousse, Tunisia.

[12] Sharad Awatade, Shweta Joshi. "Improved EAACK: Develop Secure Intrusion Detection System for MANETS using hybrid cryptography", 2016 International Conference on computing communication control and automation (ICCUBEA), Aug 12-13, 2018, Maharashtra, India.

[13] Mayank Agarwal, Sanketh Purwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in system”,

Real- Time Analysis And Simulation ofEfficient Public Transport Monitoring System

Masthan M[1],MdMosahid Raeen[2],Mirza NasimAkhtar Begg[3], Prof. Veeresh Patil[4]

[1]Department of Computer Science, East West Institute of Technology, Bengaluru, India

[2]Department of Computer Science, East West Institute of Technology, Bengaluru, India

[3]Department of Computer Science, East West Institute of Technology, Bengaluru, India

[4]Assistant Professor, Department of Computer Science, East West Institute of Technology, Bengaluru, India

E-mail: mmasthan126@gmail.com, mosahid835@gmail.com, mirzanasimakhtar@gmail.com,veereshpatil@ewit.edu

Abstract— This project proposed a method for safety measures which are necessary while driving vehicles. Road safety rules can be useful up to some extent to get away from accidents. If any misbehaviour occurs in vehicle due to driver, then a message will go to nearest police station along with the problem specification. That particular message includes the location of the bus where it is occurred. Our system can also detect alcoholic person who has been entering into bus. The alcoholic person may be driver or may be passenger. In our Project MEMS sensor is used to identify whether an accident takes place or not. If an accident takes place, a message will go to nearestpolice station.This project focuses on the implementation of a Real-Time bus Tracking System (RTBTS), by installing GPS (Global Positioning System)-module devices on buses which will transmit the current location on the GPS Receiver.We are using RFID based authentication for both passengersand driver. Fire sensor is used to monitor the fire in the bus if fire occurs in the bus send intimation to the owner and fire station. If this misconception occurs, then that message will go to nearest police station. In this way, we are indirectly providing safety to passengers and bus.

Keywords—RFID based authentication,real-time information, Real-Time Bus Tracking System (RTBTS), GPS module.

I. INTRODUCTION

The GPS based system combines GPS technologies. It is widely used in many applications and millions of users are benefitted by it every day. The product is mainly intended to increase the security and safety amongst the transportation system. This vehicle-tracking device can be installed in any vehicle to prevent thefts or to monitor the route of the vehicle. Whenever a vehicle is stolen or is lost, the device will send the coordinates of latitudes and longitudes that will help to locate the vehicle on user‘s mobile. The tracking system covers most of the highways, major cities, towns and most of the accessible villages and works efficiently in areas with better mobile connectivity. This paper explains an embedded system, which is used to know the location of the vehicle using the popular and readily available technologies like the Global Positioning System (GPS) and Global System for mobile communication (GSM). The main feature of our design is that it proposed to use a development board, which will have GPS and GSM module not as a separated module but closely linked with a microcontroller as in TanotisGboard Pro GSM/GPRS SIM900 Development Board ATmega328 Microprocessor. The advantage of using that development board is that it will reduce the size of whole system and it will reduce the power loss in terms of heat through external wirings used for the connection of GPS module with the microcontroller. Along with that, it will also increase the durability of the entire system. The ATMega328 microcontroller will provide the interfacing to various hardware peripherals. To know the location of vehicle, the mobile user has to click on the Track location button in the android app.

When it comes to public transportation, time and patience are essential. In other words, many people using public transport buses have experienced time loss because of waiting at the bus stops. Millions of Peoples need to travel from one place to another every day.

In this paper, smart bus tracking system has been proposed that when any Passengerenter into bus the RFID will check the Authentication massage will updates to IOT and also arrival times, buses current locations, and bus routes on a map can be easily found out with the help of IOT. GPS (Global Positioning System) and Google maps are used for navigation.

Each RFID tag has an information about and individual Passengers which was sensed by an RFID reader transmit the corresponding information to IOT. The outputs of this controller board are given to Wi-Fi module and LCD display. This Wi-Fi modem can sends the information to IOT according to the received data.

The proposed system shows that the RFID tracking technology is a practical option for monitoring and tracking the Passengers during their trip to and from source to destination. The GPS Module is used for Live Tracking of the Buses and alerting if fire accident occurs and send a link to android app.This system also gives an alert if fire occurs in the bus.

II. Related Work

Paper 1:"Real Time Bus Position and Time Monitoring System" IJSTE-International Journal of Science Technology Engineering, Volume 1, Issue 10, April 2015. Many passengers are usually late to work, students are late for classes as a result of they decide to anticipate the bus rather than simply merely using another alternate transportation. A variable message shown on the web which will be real time info regarding the bus showing the time of arrival at a particular bus stop might scale back the anxiety of passengers expecting the bus. With the advent of GPS and also the ubiquitous cellular network, real time vehicle tracking for higher transport management has become attainable. These technologies can be applied to conveyance systems particularly buses, which are not ready to adhere to predefined timetables owing to reasons like traffic jams, breakdowns etc. The increased waiting time and the uncertainty in bus arrival build conveyance system unattractive for passengers. The real-time bus position and time observance system uses GPS technology alongside totally different application to fetch knowledge and with code that displays the information online on with different buses on a special route to the user. When this info is conferred to the traveler by wired or wireless media or online internet media, they can use their time with efficiency and reach the stop simply before the bus arrives, or take alternate means of transport if the bus is delayed. They can even arrange their journeys long before they really undertake them. This will build the general public transport system competitive and passenger- friendly. The use of personal vehicles is reduced when additional individuals use transit vehicles, which in turn reduces traffic and pollution

.

Paper 2: M. B. M. Kamel, "Real-time GPS/GPRS based vehicle tracking system," International Journal Of Engineering And Computer Science, Aug. 2015"
The Real Time Bus Monitoring and Passenger Information bus tracking device will serve as a viable notification system that will effectively assist pedestrians in making the decision of whether to wait for the bus or walk. This device is a standalone system designed to display the real-time location(s) of the buses in Mumbai city. The system will consist of a transmitter module installed on the buses, receiver boards installed on the bus stops, LED embedded map of the BEST bus transportation routes at the centralized controller. It will also have passenger information system software installed at the bus stops, which will provide a user the relevant information regarding all the bus numbers going for his source to destination along with the route details and the cost. Assembly of these modules will enable the tracking device to obtain GPS data from the bus locations, which will then transfer it to the centralized control unit and depict it by activating LEDs in the approximate geographic positions of the buses on the route map. It will also transmit its bus numbers and route names continuously as soon as the bus comes within the range of the receiver at the bus stop. In addition, the device will be portable and sustainable; it will not require an external power source, which will eliminate long-term energy costs.

Paper 3:"Real Time Availability System" International Journal of Advanced Research in Computer Engineering Technology (IJARCET) Volume 4 Issue 3, March 2015, This Paper is a survey to implement a method that makes transport much convenient for individuals who commute daily using the public bus transport of the city, for effective time management and making it trouble-free, not just for the commuters but the Transport Department to create an efficient public transport system. There are applications available in the market today which specifies the route and the timings, predict arrival times of different buses But the survey presented here aims to build an application that takes it to the next step by making information about the vacant seats and the current location of any bus in Real-Time, accessible to the daily commuters with a novel and economical wireless system. These methodologies offer incremental improvements in bus system to meet the capacity requirements of different size cities and presents a review of strategies which can be employed to satisfy public transport demands of different city sizes. Their aim is to build a flexible, comfortable, easily available and reliable bus service which may encourage shift from private vehicles to public transport.

III. Methodology

The objective of this proposed system is to develop an application which will help to provide security for peoples. This allows relations to check the status of secure smart bus by using Iot. The proposed system will provide various facilities like check drunk and drive, Fire Detection, accident emergencies, panic button, logistic management etc.

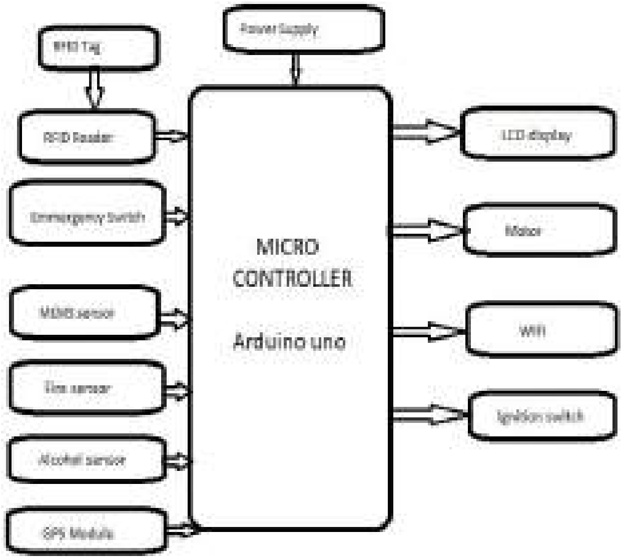


Fig1:Block Diagram

In this system Arduino microcontroller has been used. Bus unit consists of RFID Reader, different sensors and GSM module to issue the alert messages to parents when their children boards or leaves the bus. Fire sensor will be placed within the bus unit to detect fire and issues alert messages by giving the location of the bus using IOT.

In this system fire sensor is used to detect the fire accident. If there any fire accident occurs, the sensors receives a physical signal and transmit a digital signal to a Wifi module. The alert message will be send to the Relations with the help of IOT. Each Passenger consist of an individual RFID tag with the help of RFID tag, IOT.

The information of RFID tag is read by RFID reader. The reader transmits the corresponding information. RFID tag is used to send an alert message like the location of a person, speed of the bus to their respective Relations and departments.

LCD stands for Liquid Crystal Display is a flat panel display technology commonly used in TVs and computer monitors. It is also used in screens for mobile devices, such as laptops, tablets, and smart phones. The backlight in liquid crystal display provides an even light source behind the screen. This light is polarized, meaning only half of the light shines through to the liquid crystal layer. The liquid crystals are made up of a part solid, part liquid substance that can be "twisted" by applying electrical voltage to them. They block the polarized light when they are off, but reflect red, green, or blue light when activated.

A DC Power Supply Unit (commonly called a PSU) deriving power from the AC mains (line) supply performs a number of tasks: It changes (in most cases reduces) the level of supply to a value suitable for driving the load circuit. It produces a DC supply from the mains (or line) supply AC sine wave. It prevents any AC from appearing at the supply output. Power supplies in recent times have greatly improved in reliability but, because they have to handle considerably higher voltages and currents than any or most of the circuitry they supply, they are often the most susceptible to failure of any part of an electronic system.

GPS is a satellite navigation system used to determine the ground position of an object. Each GPS satellite broadcasts a message that includes the satellite's current position, orbit, and exact time. A GPS receiver combines the broadcasts from multiple satellites to calculate its exact position using a process called triangulation.

The proposed Bus Monitoring system has six modules are as follows:

1. Arduino UNO

The arduino Uno is a microcontroller board based on the ATmega328, It has 14 digital input/output pins, 6 analog input, a 16 MHZ crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button. The Uno differ from all preceding boards in that it does not use the FTDI USB to serial driver chip.”UNO” means one in Italian and is named to mark the upcoming release of arduino 1.0. The Uno is the latest in a series of USB Arduino boards and reference model for Arduino platform. The Arduino Uno can power via the USB connection or with external power supply. External power can come either from an AC to DC adapter or battery. The board can operate on an external supply of 6 to 20 volts. If supply with less than 7v, however, the 5v pin may supply less than five volts and the board may be unstable. The Ttmega328 has 32 KB of flash memory for storing code .It has also 2KB of SRAM and 1KB of EEPROM. The Arduino software includes a serial monitor which allows simple textual data to be send to and from the Arduino board, The RX and TX LEDs on the board will flash when data is being transmitted via the USB to serial chip and USB connection to the computer.

A Software Serial library allows for serial communication on any of the UNO’s digital pins, the arduino software includes a wire library to simplify use of the I2C bus. Arduino is open source hardware and software, which are license under the GNU lesser General public license, which is permitting the manufacture of Arduino board and software distribution by anyone.

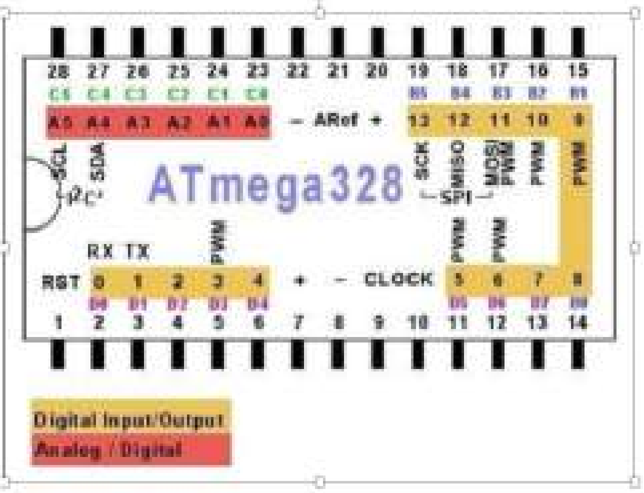


Fig 2:Atmega328 Micro controller

The Arduino are programmed using a dialect of feature from programming language C and C++. In addition to using traditional compiler tool chains, the Arduino provide integrated development environment (IDE) based on processing language project [1].

2. DC MOTOR:

DC motors are used to physically drive the application as per the requirement provided in software. The dc motor works on 12v. To drive a dc motor, we need a dc motor driver called L293D. This dc motor driver is capable of driving 2 dc motors at a time. In order to protect the dc motor from a back EMF generated by the dc motor while changing the direction of rotation, the dc motor driver have an internal protection suit. We can also provide the back EMF protection suit by connecting 4 diode configurations across each dc motor.

3. SENSOR:

MEMS Sensor gather information from the environment through measuring mechanical, thermal, biological, chemical, optical, and magnetic phenomena. The electronics then process the information derived from the sensors and through some decision making capability direct the actuators to respond by moving, positioning, regulating, pumping, and filtering, thereby controlling the environment for some desired outcome or purpose. MEMS Sensor is used to detect Earthquakes, to check whether the machine is working properly or not and gas shutoff.

In our Project MEMS sensor is used to identify weather the vehicle is safe side or met an accident. Private Travel Buses contains actuators. We will attach MEMS sensor to this actuators. When these actuators behavior is somewhat different from routine one then MEMS sensor will get activated and the corresponding status signal will goes to ARDUINO.

4.ALCOHOL SENSOR:

This alcohol sensor is suitable for detecting alcohol concentration on your breath, just like your common breathalyzer. It has a high sensitivity and fast response time. Sensor provides an analog resistive output based on alcohol concentration. The drive circuit is very simple, all it needs is one resistor. A simple interface could be a 0-3.3V ADC.

In our project Alcohol sensor detects, weather the driver is alcoholic or not. If driver is in alcoholic state then alcohol sensor sends this status to ARDUINO. Then ARDUINO sends corresponding error message to IOT through WIFI Module. This alcohol checking takes place after swiping RFID card.

5.RFID(Radio Frequency Identification):

Radio-frequency identification (**RFID**) is the wireless use of electromagnetic fields to transfer data, for the purposes of automatically identifying and tracking tags attached to objects. The tags contain electronically stored information. Some tags are powered by-electromagnetic induction from magnetic fields produced near the reader. Some types collect energy from the interrogating radio waves and act as a passive transponder. Other types have a local power source such as a battery and may operate at hundreds of meters from the reader. Unlike a barcode, the tag does not necessarily need to be within line of sight of the reader and may be embedded in the tracked object. RFID is one method for Automatic Identification and Data Capture (AIDC).

1) Tags

RFID tags having a factory-assigned serial number that is used as a key into a database, or may be read/write, where object-specific data can be written into the tag by the system user.

RFID tags contain at least two parts: an integrated circuit for storing and processing information, modulating and demodulating a radio-frequency (RF) signal, collecting DC power from the incident reader signal, and other specialized functions; and an antenna for receiving and transmitting the signal. The tag information is stored in a non-volatile memory.

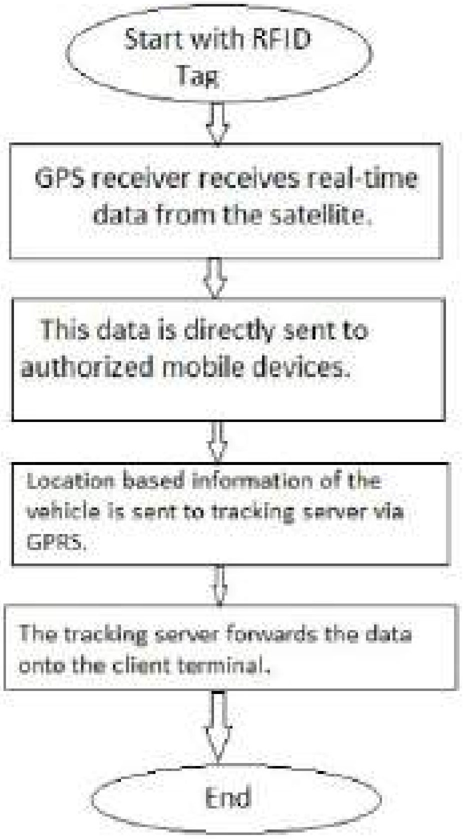
2) Readers

In our project, RFID tag is used for authentication purpose. To activate our entire system, we have to swipe RFID tag near RFID reader. If Tag details persists in the database of ARDUINO, then micro controller allows the driver to pass vehicle through 1st tollgate. After reaching 2nd tollgate, we have swipe the same RFID tag once again at Tollgate 2. Here, time difference between two stations will be calculated. If time difference between two stations is more than 1hour, then there is no problem. That means, driver is driving the vehicle with safe speed. If time taken from tollgate1 to tollgate2 is less than 1hour, then we can said that driver is driving vehicle with over speed. Like this the same process repeated for every 100km distance. Because of this, driver does not sleep while driving.

6. LCD Module:

LCD stands for **L**iquid **C**rystal **D**isplay. LCD is finding wide spread use replacing LEDs (seven segment LEDs or other multi segment LEDs) because of the LCD stands for **L**iquid **C**rystal **D**isplay. LCD is finding wide spread use replacing LEDs (seven segment LEDs or other multi segment LEDs) because of the following reasons:
The declining prices of LCDs. 2.The ability to display numbers, characters and graphics. This is in contrast to LEDs, which are limited to numbers and a few characters.

1. **Flow Chart:**


Start with RFID Tag
GPS receiver receives real-time data from the satellite.
This data is directly sent to authorized mobile devices.
Location based information of the vehicle is sent to tracking server via GPRS.
The tracking server forwards the data onto the client terminal.
End

Algorithm 1: Bus Tracking Algorithm

Input:

Routing Table entries:

Node ID:

Longitude:

Latitude:

Output:

step1:start

step2:Get longitude and latitude values using GPS module and send it to Microcontroller

step3:microcontroller send this information to GSM module.

step4:GSM module send this message to mobile device.

step5:stop.

Algorithm 2: Fire Sensor

- Input:

  Flame temperature:

  Environment temperature:

Output:

- Step 1: start
- Step 2: if(Ft > Et)
- Step 3: alarm on and send message to nearest police station and hospital through GSM module.
- Step 5: stop

IV. Results and Discussion

It should include important findings discussed briefly. Wherever necessary, elaborate on the tables and figures without repeating their contents. Interpret the findings in view of the results obtained in this and in past studies on this topic. State the conclusions in a few sentences at the end of the paper. However, valid colored photographs can also be published.

V. Conclusions

In this papers, we develop the “Real Time Analysis and Simulation of Efficient Public Transport Monitoring System” in that mainly we focused on the accuracy of location and calculations of time, coordinates and simple user interface. This system save the time and increase the work efficiency of end users because it reduces the user’s efforts to travelling for work and avoid the wastage of waiting time for bus. It also consider the points that is Robust, Reliable and efficient for travelling in city.

The functionalities are better more accurate than those provided by the arduino based systems as notifications can be instantly cleared whereas GSM systems spam the information.

VI. FUTURE SCOPE

This software could be modified and developed for future use. Provisions for detecting theft, restricting entry and verifying assigned passenger list on id can be added.

The RFID can be replaced with a better reader or more reliable identification methods like biometric identification.

Acknowledgment

Firstly, we express our sincere thanks to our guide Mr.VeereshPatil, Assistant Professor, Department of CSE, EWIT and Dr.ArunBiradar, Head of Department,Computer Science and Engineering for their moral support. We express our sincere gratitude to our principle Dr. K Chennakeshavalu for his constant support and encouragement; we also thank all the faculties of East West Institute of Technology for their co-operation and support.

References

[1] R.Ramani, S. Valarmathy, Dr. N Suthanthira, S. Selavaraju, M.Thiruppathi, R.Thagam, ―Vehicle Tracking and Locking Based GSM and GPS‖, Issue Date: Sept 2017

[2] F. M Franczyk, and J.D. Vanstone, ―Vehicle Warning System‖, Patent number: 73639, Issue Date: 22 Apr 2018.

[3] William Stalling, ―Wireless Communication and Networks‖, 2nd edition, prentice hall of India, 2017

[4] Krishna Kant, ―Microprocessor and Microcontroller‖, Eastern Company Edition, New Delhi 2018

[5] PranjaliJumle and Suresh Gohane “Intelligent Bus Tracking System based on GSM and GPS”IJCSN International Journal of Computer Science and Network, Volume 6, Issue 2, April 2017.

[6] Dhruv Patel, Rahul Seth and VikasMishra “Real-Time Bus Tracking System” International Research Journal of Engineering and Technology (IRJET) 2017.

[7] ManashPratimGohain, Speed Governors, GPS must forschool buses,The Times of India,February 24,2017

[8]Kumar, B. A., Vanajakshi, L., & Subramanian, S. C., "Bus traveltime prediction using a time-space discretization approach,"*Transp.Res. Part C Emerg.Technol.*, pp. vol. *79*, 308–332, 2017.

Authors Profile

[1] **Mr Masthan M** is pursuing his 8 semester B.E. in Computer Science & Engineering at East West Institute Of Technology, Bengaluru, India. His area of interest includes Internet of Things(IOT),Web Development,Big Data.

[2] **MrMd Mosahid Raeen** is pursuing his 8 semester B.E. in Computer Science & Engineering at East West Institute Of Technology, Bengaluru, India. His area of interest includes Internet of Things(IOT),Python and Java.

[3] **Mr Mirza Nasim Akhtar Begg** is pursuing his 8 semester B.E. in Computer Science & Engineering at East West Institute Of Technology, Bengaluru, India. His area of interest includes Internet of Things(IOT),Big Data and Machine Learning.

[4] **Mr, Veeresh Patil** got M.Tech degree in Computer Science, Bengaluru, India. He is currently working as Assistant Professor in the Department of CSE, EWIT.His area of interset includes Image processing,Machine Learning.