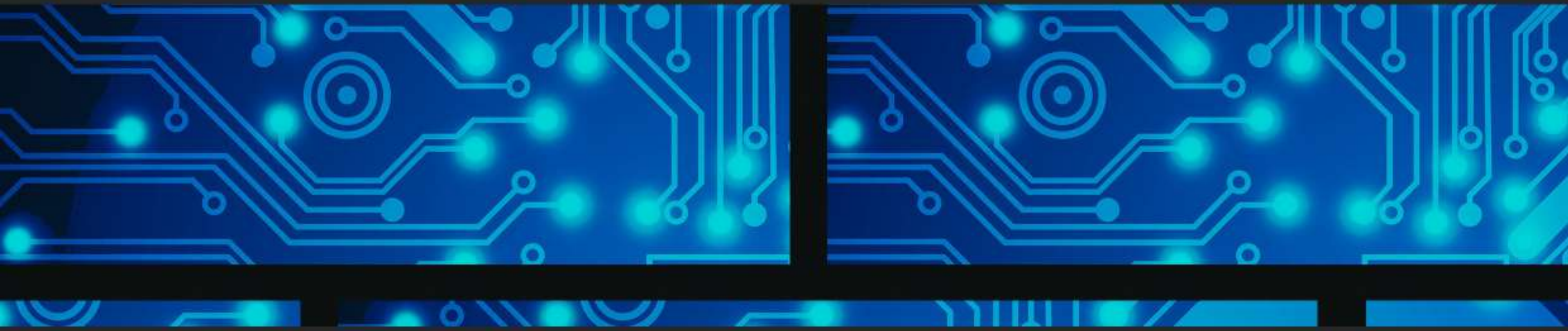
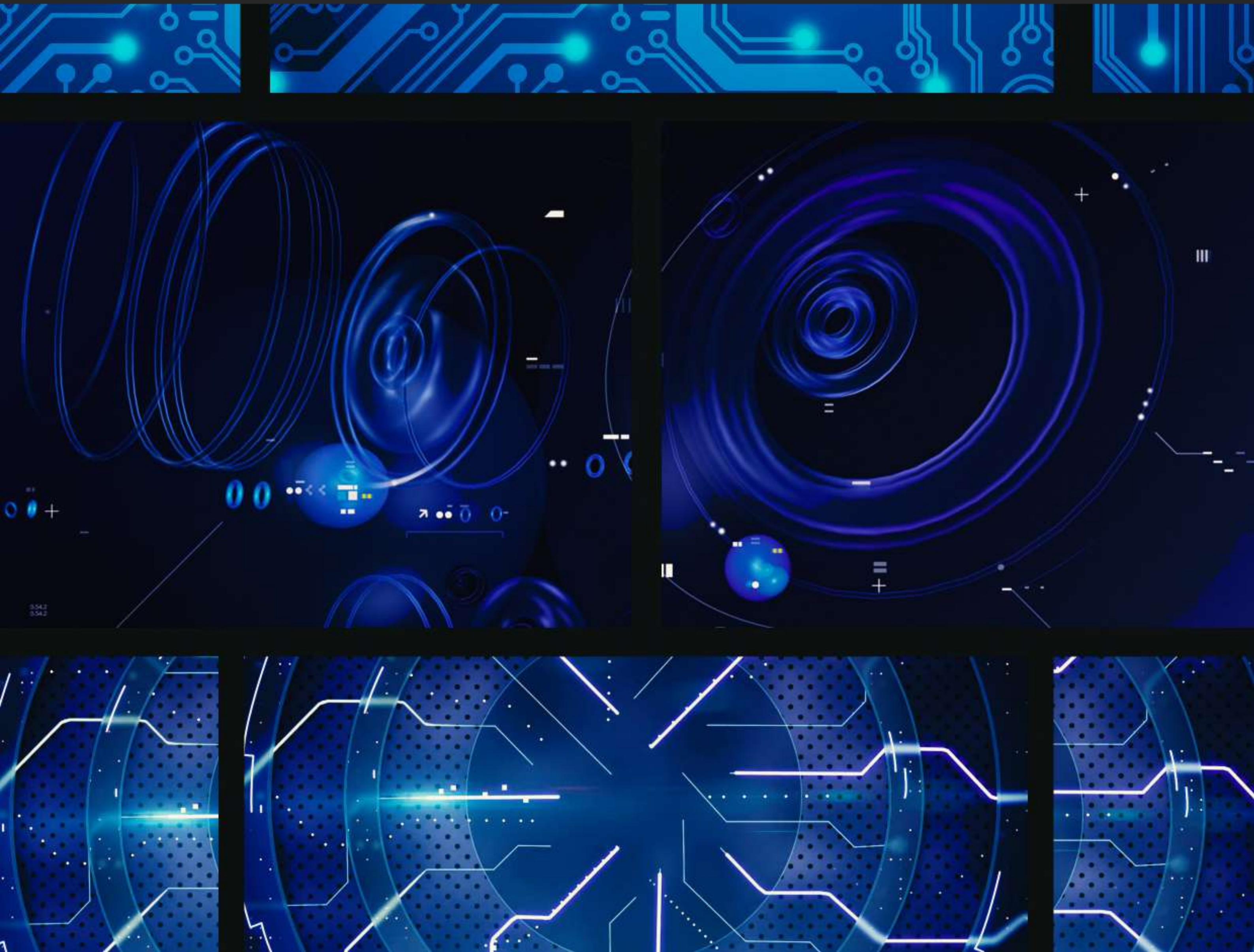
Mahatma Education Society's

Pillai College of Engineering



THE PCE JOURNAL OF COMPUTER ENGINEERING

ACADEMIC YEAR 2017-2018 | VOLUME 6, ISSUE 1



PILLAI COLLEGE OF ENGINEERING

Journal of Computer Engineering

PCE JCE_____________

Volume 6 Issue 1 2017-2018

Journal of Computer Engineering

Pillai College of Engineering

New Panvel

Table of Contents

Editorial

It takes immense pleasure in launching this issue of the Journal of the Computer Engineering Department, PCE. The journal is a forum for the students and faculty of the department to showcase their work in various imminent fields related to computer engineering and its applications.

This issue has 19 papers comprising the outcome of research work done by the students and the faculty of the computer department, exploring the various domains such as Human Machine Interaction, Machine Learning, Internet of Things, Natural Language Processing, Security, Mobile and Web technologies, Artificial Intelligence, Networking, E-Commerce and others.

I hope that this issue of PCE JCE will be helpful for the future aspiring computer engineers and the research students. I thank the editorial team for their efforts put in for the launching of this issue.

Dr. Sharvari Govilkar
Editor-in –Chief

An Ontological Interactive Personal Assistant based on Automation and Image Recognition

Susmitha Nair
Computer Department
Pillai College Of Engineering
Mumbai University, India
smurleedharan31@student.mes.ac.in

Jagdish Khaire
Computer Department
Pillai College Of Engineering
Mumbai University, India
jkhaire@student.mes.ac.in

Prenav Premkumar
Computer Department
Pillai College Of Engineering
Mumbai University, India
email address

Shruthi Srinivasan
Computer Department

Pillai College Of Engineering
Mumbai University, India
email address

***Abstract*—In this project,we present an all purpose personal assistant .Cognitive artificial intelligence, which is an emerging field, is used as a base for understanding and development.This allows the personal assistant to be more interactive through knowledge growth.The user interaction is further enhanced using the concepts of image processing and automation.Using image processing,the system analyses images given as input to it and produces a natural language description.Another concept used is Ontology reasoning which improves the scope of the system by giving it flexibility of ability selection,combining different hardwares to facilitate all user requirements.**

Keywords : Artificial Intelligence,Personal Assistant,Ontology,Image Processing,Speech Recognition,Voice Commands

I. INTRODUCTION

The mechanism that occurs within the brain that makes intelligence in human has been a mystery for a long time. the increasing interest in the research on human brain led to evolution of new field called Artificial Intelligence(AI).AI is an applied technology to produce an intelligent system that can mimic human intelligence. The scope of AI is expanded with introduction of image processing. Now a days all over digitization technology is used. Text recognition using image processing involves a system designed to translate images of type written text into machine editable text or to translate pictures of characters into a standard encoding scheme representing them. A personal assistant is an AI base technology which is a combination of several different technologies including speech recognition,language analysis, AI base natural language processing and image processing. It would be very interesting if the task that the user wants to perform can easily be carried out using a system i.e. the system interacts with the user.

An important issue on realising an autonomous agent is that, each agent's behaviour is constraint by its environment, i.e. the external resources, depending on the available functions and capabilities of the API and hardware. In order to resolve this issue, agents need a mechanism so that they can cooperatively execute task with the help of other agents on different environments. moreover by removing the need to use any other external peripheral devices to give input to a system it would be more convenient for the user to control a device by means of voice. thus the concept of ontology and automation comes into existence. Now a days a personal assistant proves to be more helpful if it has ability to process large amount of data and store it in desire format. Text recognition in image can be used in offices, banks and colleges.

To realise these mechanisms the system has to consider that a agent's behaviour can affect others agents behaviours and the voice input given by the user has to be interpreted properly so that the user's task is executed. here the main agent in coordination with other agents tries to find the most suitable agent that could perform a task at that particular instance based on it's scope and ability. To achieve this, an internal database is maintained that consist of all the abilities of each agent. An agent can give proper information of necessary abilities as well as the list of agents that have one of the ability of the search results. if the task to be performed is recognition of characters from an image that includes pre-processing, segmentation, feature extraction, classification and post-processing; this task can be carried out only by an agent that has the ability to capture images. In this way a personal assistant agent can use abilities of other agents using that information. An agent uses a microphone to listen to a user's request, converts it into data that can be analysed, compares it to a query plan and formulates a suitable response which is given as a verbal output through speaker.

II. LITERATURE REVIEW

The basic idea of this project came from several fields based on interactive personal assistant with ontology for ability selection of multiple agents involved. We have gone through several papers to gather information about various techniques for image analysis, speech recognition, pre processing, feature extraction and conversion.

For the Personal Assistant to be interactive and support Ontology, it needs to recognize the Speech input, use techniques to interpret the command and carry out tasks as per the results from the database. If the user requests consists of commands to deal with images, then image processing algorithm is required. In this paper we discuss an algorithm for solving the problem of character recognition. We give the input in the form of images. The algorithm is trained on the training data that is initially present in the database. We carry out preprocessing and segmentation to detect characters in images.The proposed method is extremely efficient to extract all kinds of bimodal images including blur and illumination. We also discuss a speech processing algorithm which is bi-directional. The speech initiation can either be from the system or the user.The user initiated input can be a dialogue based query or an executable query. Here, the interaction between the system and the user is enhanced by the implementation of system initiation of reminders i.e. speech system speaks on a reminder. The overall flow chart for both Speech processing and Image processing modules are as follows:

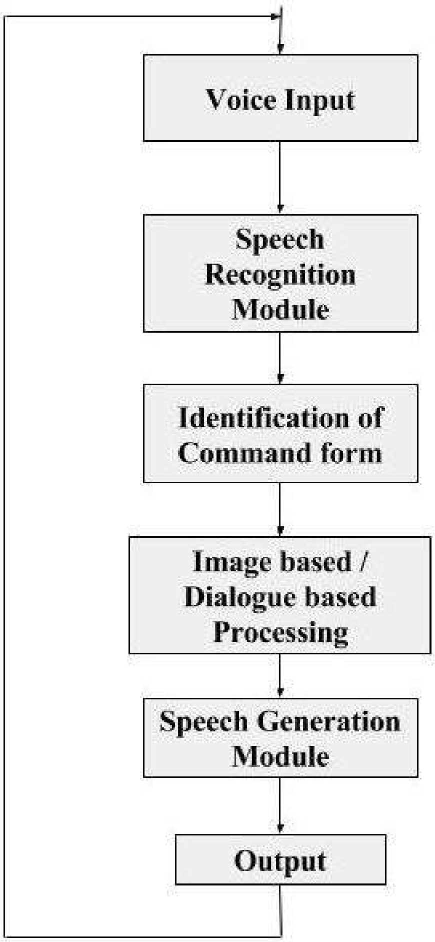

Voice Input
Speech Recognition Module
Identification of Command form
Image based / Dialogue based Processing
Speech Generation Module
Output

Fig. 1. Overall View of the system

III. THE CHARACTERISTICS OF SYSTEM DESIGN

A. Hardware Design

- LCD module: The module as a LCD controller, it supported 1024*1024 image of 15 gray-scale or 3375 colours.
- Keyboard module: It can be used for inputting passwords, further security details like ATM pins and other inputs which cannot be given by voice.
- Camera Module: Most probable and basic need for image recognition with minimum resolution or VGA Capturing image Camera.
- CPU Configurations: Minimum requirement for this project can be from Core i3 processor, AMD Ryzen 7 and above; including minimum drive space upto 500 GB with least RAM capacity upto 2 GB - DDR3 and cache memory in the range 5 - 7 MB.

B. Software Design

C. Design Of Algorithm

Scan the image in front of camera. Converting image from color to gray image. Conversion of image from grayscale image to binary image. Performing pre-processing to filter noise from image Performing skew correction to get possible binary image data. Segmenting each binary image characters. Extraction of the characters from image. Classifying each image characters with stored characters. Loading all characters in database matched. Appending characters into word line context. Forwarding processed data to text-to-speech module.

ACKNOWLEDGMENT

We are profoundly graceful to Prof. Krishnendu Nair for her expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. We would like to express deepest appreciation towards Dr. R.I.K. Moorthy, Principal PCE,New Panvel, Prof. Madhumita Chatterjee HOD Computer Engineering Department and Prof. Gaurav Sharma (Project Coordinator) whose invaluable guidance supported us in completing this project.

At last we would like to express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped us directly or indirectly during this course of work.

IV. CONCLUSION

REFERENCES

[1] Sho Oishi, Naoki Fukuta, ”Toward a Flexibility Ability Selection Mechanism for Personal Assistant Agent using Ontology Reasoning” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] Priyanka Jain, Priyanka Pawar, Gaurav Koriya, Anuradhal Lele, Ajai Kumar, Hemant Darbari, ”Knowledge acquisition for Language description from Scene Understanding” For , 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Arwin Datumaya Wahyudi Sumari[3], Adang Suwandi Ahmad ”Cognitive Artificial Intelligence” vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] Hugues Sansen[4], Shankaa ”The Roberta Ironside Project” unpublished.

[5] Ankush Bhatia, ”Making An Intelligent Personal Assistant” J. Name Stand. Abbrev., in press.

[6] Chowdhury Md Mizan, Tridib Chakraborty[6],Surparna Karmakar, ”Text Recognition Using Image Processing” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

AUTOMATED HOME SECURITY SYSTEM

Aravind Acharya-Student,PCE;Akhil Meleth-Student,PCE;Atish Mhatre-Student,PCE;Rohan Vadlamudi-Student,PCE;Deepti Lawand-Faculty,PCE

Abstract: Security has been one of the increasing concerns in our society. Nowadays, technology is being used to find efficient and user-friendly methods in order to overcome the problem of security. Newer and newer devices are invented that ensure the security of our houses, cars and other valuables. In this paper, we aim on developing a security system for our homes. The system enables a user to remotely access the security system. We also plan on developing an Android application which will act as an interface between the user and the security system. A more secure system is created as a two level bio-metric password is used for opening doors. Surveillance feed will be available on the go which can be viewed through the application. An image is captured on the ringing of doorbell as well as fingerprint mismatch which will be sent to the user via the application. The system also detects smoke, and in case of a heavy fire, it will automatically alert the emergency services along with the location info. The system alerts the user and the emergency services in case of burglary.

I. INTRODUCTION

Sensor based home security system are the high technology and methodical systems which connect wirelessly and ensure real time operation and indication of the threat to the house. The idea of comfortable living in home has changed since the past decade as digital, vision and wireless technologies are integrated into it. Nowadays internet plays a major role in every area, so integrating sensors technology within a wireless environment could resolve the security issues of society to a great extent. The various drawbacks of existing technologies are cost and range. In this paper a design and implementation of sensor based security system has been presented, which will resolve various security issues like unauthorized intruder entry, fire detection etc. Therefore, continuous monitoring of the home/apartment is possible. The system is cost effective, reliable and has low power consumption.[8]

II .LITERATURE SURVEY

From the research paper ‘Home Security System Based on Sensors and IoT

limitations of existing system are that most of systems established on Internet monitoring based systems require higher bandwidth, high data speed rates and high operational cost and hence more suitable for only in industry. ZigBee, Bluetooth based system has a geographical limitation. Data rate transfer rate is also low in ZigBee communication. Range is the biggest challenge in ZigBee and Bluetooth based systems. It is challenging to upgrade existing conventional control systems with remote control system capabilities. In cellular monitoring systems like GSM the long term operational cost is relatively high due to usage charges incurred in each message transaction. This system is concerned about overall security of the house and includes circuitry which in worst case (accidents) automatically sense the situation and sends the emergency message on the website, which is easily accessed by security guard/security firm/owner or individual. The end product will have a simplistic design making it easy for users to interact with.[2]

From the research paper ‘Security System using Arduino Microcontroller’ a security system is developed in which Arduino board is used which is considered as one of the modern programmable device and utilizes the speed dial function in mobile phones. This system is developed using PIR sensor, magnetic sensor, temperature sensor and all data from these sensors are continuously received and processed by Arduino Uno board. PIR sensor is used to detect human body that is a constant source of infrared radiation. Magnetic sensors are used to detect intrusion through doors and windows. Temperature sensors are used to detect temperature change for detecting accidents like fires. The communication between mobile phone and micro-controller is done using GSM shield. GSM shield uses sim card and due to range fluctuation or bad network, the GSM shield may not work properly. Android app will also be developed in which there will be direct buttons to control the system. Camera module can also be implemented on the system. [4]

III. PROPOSED SYSTEM

The System consists of various modules such as camera module, sensors module, servo motor etc. all these modules are connected to Arduino microcontroller.

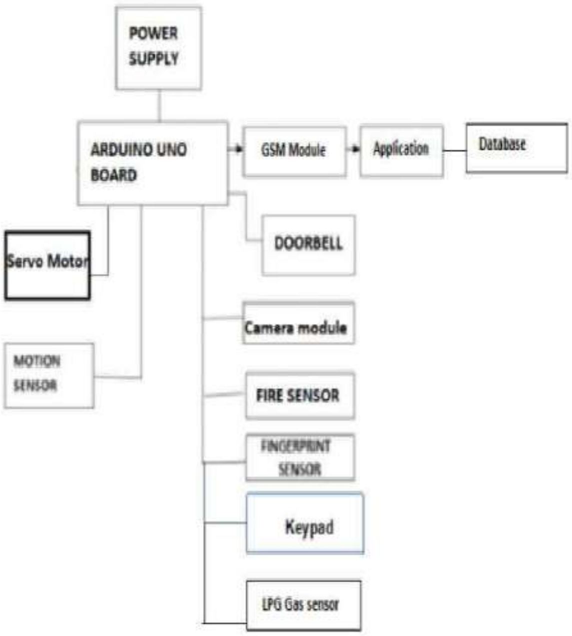


Fig: 3.1 Block diagram of the proposed system

The system can be divided into two sub parts consisting of:

Sensor subsystem: System consists of various sensors such as PIR sensor, lpg sensor, smoke sensor, fingerprint sensor etc and gives input to system and based on it following alerts are generated.

Software subsystem: Application is developed for interaction between user and hardware of system. Application is developed using Android studio.

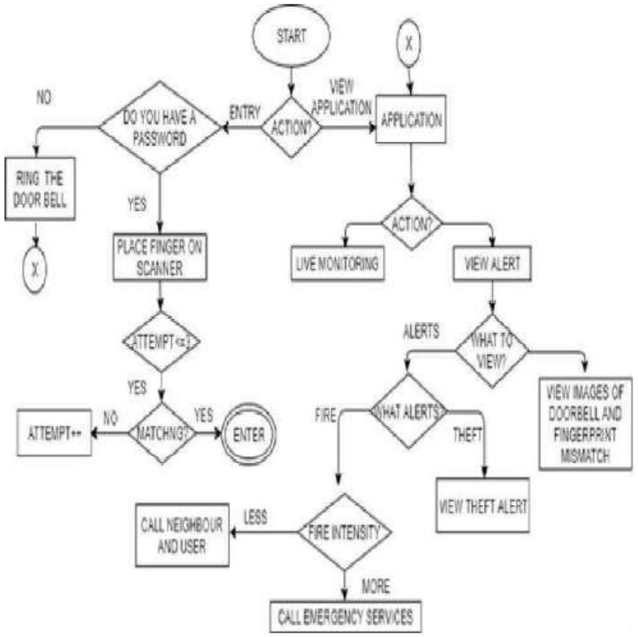


Fig 3.2: Flowchart of the system

When the user enters the system, it can perform two basic actions-opening the door lock or viewing the application. If the user has his fingerprint registered in the system then he can enter by placing his finger on the fingerprint sensor.. If you try to enter your fingerprint you have at most 3 chances to get it right, or else the system will capture your image and send it to the user. Once you are in the application, we have two options of monitoring our homes live and viewing the history of alerts. We have an option of viewing the fire alerts and theft alerts separately, the application also alerts the emergency services in case of fire. For this the system first checks the intensity of fire using its sensors and if the intensity is higher it alerts the emergency services and the registered neighbour as well. If the threat is mediocre, such as gas leakage, then the system alerts the neighbours in time who can act accordingly. The system also provides a feature of viewing the photos of all the images captured by the doorbell camera or the images of the person who tried to break into the system using improper fingerprints. In order to enhance security, a 2-level security feature can be added which will make use of the fingerprint sensor as well as the doorbell camera. The user must place his fingerprint on the sensor,if the fingerprint is authenticated then the camera captures the image of the person for face recognition.If the fingerprint is not authenticated then the user can have two more tries before the user is alerted via application. Once the fingerprint is authenticated, the camera captures the image and then checks with the image in its database for the image of the registered user.If both the images match,then the person can unlock the door.Else if the image does not match the user is alerted along with the image.However it may occur that due to technical issues the image may not be properly captured or there might be errors in image matching.In such cases,the system recognises this failures and an OTP is generated which will be sent to the user's device once fingerprint authentication is done. The security can be further increased by introducing OTP as a 3rd level authentication. after fingerprint and image recognition authentication.

WORKING CIRCUIT

After the activation the system will work as follows. Different Sensors Used are:

PIR Sensor

Fig 3.3: PIR Sensor

A PIR sensor stands for Passive infrared sensor it is as electronic sensor which measure infrared light.

A passive infrared sensor (PIR sensor) is an electronic sensor that measures infrared (IR) light radiating from objects in its field of view. .PIR sensor will detect the presence of human when someone enter the house.

LPG Sensor

This is a simple-to-use liquefied petroleum gas (LPG) sensor, suitable for sensing LPG (composed of mostly propane and butane) concentrations in the air. The MQ-6 can detect gas concentrations anywhere from 200 to 10000 ppm.[6]

LPG gas sensor will detect leaking of gas and gives a precautionary alert to the use.

Fig 3.4: LPG Sensor

Fire Sensor

The Fire Sensors is used to check if there is any fire presence in the room. It continuously check room temperature and send its value to micro-controller.

Fig 3.5: Fire Sensor

Biometric sensor:

A biometric sensor is a transducer that converts a biometric treat (fingerprint, face, etc.) of a person into an electrical signal. Generally, the sensor reads or measures pressure, temperature, light, speed, electrical capacity or other kinds of energies.[6]

Two component used are:

i. fingerprint sensor

ii. Camera for facial recognition

IV. ALGORITHM

Algorithm based on Tree Comparison using Ratios of Relational Distances:[9]

1. The direct Fourier transform is applied on the fingerprint so as to enhance the image as well as to obtain a binary image.

2. Thinning Algorithm is used to make the thickness of the edges as 1px.

3. We than isolate the end points of every edge,thus we get an image which consists of only end points of edges.

4.The 5 neighbors of the center most pixel is named as i1,i2,i3 and so on.

5.The Euclidean distance between this points from the center are calculated .

6.The following 10 ratios are calculated such as (i - i1): (i - i2), (i - i1): (i –i3), (i - i1): (i – i4), (i - i1): (i – i5) , (i – i2) : (i – i3),(i– i2) : (i –i4), (i – i2) : (i – i5), (i – i3) : (i – i4), (i – i3):(i – i5), (i – i4) : (i – i5) according to the following equation : (a – b): (a – c) =Max {(a-b), (a-c)} / Min {(a-b), (a-c)}.

7. A table containing these 10 ratios is built.

8. The database already contains a table of ratios of distances neighboring pixels from i.

9. A search is done to check whether the calculated 10 ratios are present in the database

10. If yes, there is a match else the finger print does not match.

ii face recognition:

Algorithm based on support vector machine:[10]

1 Representing images using vectors of size N2.

r1, r2, r3…rm

2 finding averaging set

3. $\psi = \left(\frac{1}{m}\right)\sum_{i=1}^{M} \Gamma i$

$\phi i = \Gamma i - \psi$

4. C=AT.A …….. Covariance matrix.

5. Finding Eigen vectors of M x M and finding Eigen vector of this small matrix.

6. V is non-zero vector and is a number, such that Av = λv, Then v is an Eigen vector of A with Eigen value.

7. $A^T A v_i = u_i v_i$

8. $AA^T(Av_i) = u_i(Av_i)$

.9. A face image can be projected into face space by $\Omega_k = U^T(\Gamma^k - \psi); k = 1 \dots M$

GSM (Global System for Mobile)

GSM module is an electronic device which is used to communicate with arduino board.

For our system we have use gsm 800 module for sending an sms alert to user.

Fig 4.1: GSM 800

Arduino Uno

Arduino is an open source computer hardware and software company, project, and user community that designs and manufactures single-board microcontrollers and microcontroller kits for building digital devices and interactive objects that can sense and control objects in the physical world [6]

In this system we required Arduino UNO, remote controller. By connecting all connection correctly apply a simple C or C++ code on Arduino sensor senses the motion so that it will get alert to user.

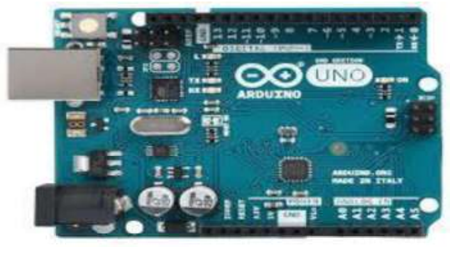


Fig 4.2: Arduino Uno

V. FUTURE SCOPE

A System can be developed in which the battery powered system kicks in as soon as electricity is shut off. Arduino microcontroller can be replaced with more advanced microcontrollers such as Raspberry pi. Cloud services with a very high memory capacity can be used for video backups. Variable sensitive gas sensors and smoke sensors can be used for gas and smoke detection respectively. The entire house can be made energy efficient with the automation of other electrical devices such as lights, fans, etc.

VI. CONCLUSION

Thus, we have successfully presented a low cost, reliable and safe home security system. The system can be implemented wherever the safety of the residents is a primary concern. Since the system uses batteries are a primary power source it can still work when power cuts occur. The system has undergone many testing processes and thus the chances of the system breaking down are very low. The system is designed in such a way that primary focus is on safety and reliability

VII. REFERENCES

[1] Home Automation and Security using Arduino Micro-controller- Viraj Mali, Ankit Gorasia, Meghana Patil, Prof. P.S. Wawage,NPCI-2016

[2] Home Security System Based on Sensors and IoT by Nidhi Sharma and Indra Thanaya,IJIRSET-June 2016

[3] IOT based Theft Preemption and Security System by Safa.H, Sakthi Priyanka.N, Vikkashini Gokul Priya.S, Vishnupriya.S, Boobalan.T

[4] Security System using Arduino Microcontroller by Priya H. Pande,Nileshwari N. Solanke,Sudhir G. Panpatte,IARJSET-Jan 2017

[5] Home automation and security system by Surinder KaurRashmi Singh Neha Khairwal and Pratyk Jain, (ACII), Vol.3,No.3,July 2016

[6]https://en.wikipedia.org/

[7]https://challenge.toradex.com/projects/10133-home-automation-system

[8]https://www.ijirset.com

[9] Abinandhan Chandrasekaran, Bhavani Thuraisingham 'Fingerprint Matching Algorithm Based on Tree Comparison using Ratios of Relational Distances'. Department of Computer science, .april 2007, INSPEC Accession Number: 9465252[online] Available:www.Ieeexplore.ieee.org

[10] K. Venkata Narayana, V.V.R. Manoj, K.Swathi .' Enhanced Face Recognition based on PCA and SVM'. Journal of Computer Applications. Volume 117,no 2,may 2015,pp. 975 – 8887[online] Available: www.ijcaonline.org

Automatic Accident Detection and Notification System

Payel Thakur
Assistant Professor
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - payelthakur@mes.ac.in

Sanjoli Singh
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - sanjolisingh19@gmail.com

Garima Shukla
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - garima.shukla17@gmail.com

Tanya Bhutani
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - tanyabhutani26@gmail.com

Sneha Negi
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - snegi.7995@gmail.com

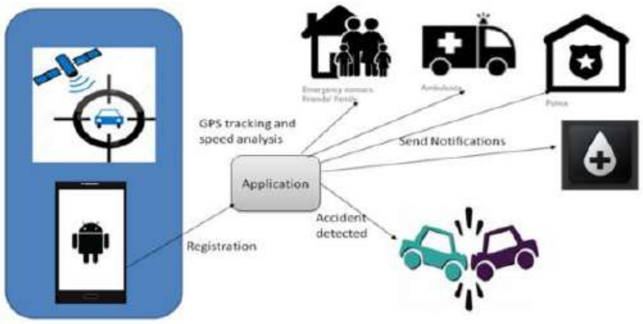
Abstract**—*****Vehicular Accidents are a major cause of concern in today's world.Safety of the driver and the co passengers can be threatened because of various reasons that lead up to an accident.And moreover there is a huge lag between the time of accident and time when emergency services reach ground zero.Many lives can be saved if proper emergency services reach the accident location at the right time.With the help of the proposed system not only accidents are detected but also notifications are sent to the nearest hospital,police station and emergency contacts.Accidents are detected using three sensors i.e,accelerometer,force resistive sensor and gyroscope so as to get accurate results.These sensors form the part of the embedded system which has an arduino and bluetooth module.The arduino constantly receives the sensor data and sends it to smartphone application via the bluetooth module.The smartphone detects whether an accident has occurred or not using the Accident detection algorithm.On detection of an accident,a message along with the gps coordinates(users current location),blood group and vehicle plate number(collected at the time of user registration) is sent to the nearest hospital,police station and emergency contacts.This process can significantly reduce the number of casualties because of delay in receiving proper medical care.Also in order to minimize false positives,an alarm system has been included which goes off as soon as accident has been detected.If the driver is safe,he/she can shut the alarm and cancel the sending of the message.The alarm rings for about 30 seconds after which it automatically forwards the message to emergency services and contacts.This application will help the service providers to reach on time and save valuable human life.***

Keywords—Accelerometer, gyroscope,bluetooth, nested if-else , Embedded Processor, gps, gsm ;

I. INTRODUCTION

In this day and age there is an extreme increment in the utilization of vehicles.Such substantial car use has expanded activity and along these lines bringing about an ascent in street accidents.This incurs significant damage on the property and additionally causes human life misfortune as a result of inaccessibility of quick well being facilities.Complete mishap aversion is unavoidable yet at any rate repercussions can be lessened. Proposed framework tries to give the emergency facilities to the casualties in the briefest time conceivable.

As human lives are in question, the discovery and reaction time are urgent factors for the victim(s) of a vehicle mishap and also the overseeing agencies.Indeed,even a slight decrease in the reaction time can diminish the number of fatalities and monetary loss by a huge factor.



The AADNS system uses the input from sensors and passes it to the smartphone via bluetooth.Using Accident detection algorithm,we can detect the occurrence of an accident with the inputs.

Registration includes user's personal info like blood group, etc. along with his photograph.In case of emergency, notification will be sent to nearest blood banks through mobile Search nearest Hospitals, police stations and blood bank.First user have to do registration to application then if any accident occurred then it is detected by GPS tracker and the personal details of those who have met with an accident that has been already stored in database is sent to nearby blood bank,hospital,friends,family members.Global Positioning System (GPS) is used to identify the location of the vehicle.GSM is used to inform the exact vehicular location to the emergency numbers.Message will give longitude and latitude values. From these values location of accident can be determined.Such a module works the same as a regular phone.

II. LITERATURE SURVEY

- ANDROID APPLICATION FOR AUTOMATED ACCIDENT DETECTION- This paper presents a system that uses smartphones to automatically detect and report vehicle accidents in a timely manner. Data is continuously taken from smartphone's accelerometer and analyzed using Dynamic Time Warping (DTW) to determine how badly the accident is happened.An e-Call System it automatically calls the nearest emergency Centre. Even if no passenger is able to speak, a Minimum Set of Data is sent, which includes the exact location of the Accident Site.

- CAR ACCIDENT DETECTION SYSTEM USING GPS AND GSM-The proposed system consists of two units namely, Crash Detector Embedded Unit and Android Control Unit. Crash Detector Embedded Unit is responsible for detecting the accident condition using three-axis accelerometer sensor, position encoder, bumper sensor and one false alarm switch. Bluetooth module (HC-05) is used to send the accident notification to the victim's android phone where an android app will get the GPS location of accident spot.

- REAL TIME TRAFFIC ACCIDENT DETECTION SYSTEM USING WIRELESS SENSOR NETWORK- This paper proposed the use of Wireless Sensor Network and Radio Frequency Identification Technologies. Sensors will be installed in a vehicle which will detect accident location and speed of the vehicle. These sensors will then send an alert signal to a monitoring station and monitoring station, in turn, will track the location where the accident has occurred.

- INTELLIGENT SYSTEM FOR VEHICULAR ACCIDENT DETECTION AND NOTIFICATION-Accident can be detected using flex sensor and accelerometer, while location of accident will be informed to desired persons such as nearest hospital, police and owner of vehicle through sms sent using GSM modem containing coordinates obtained from GPS along with time of accident and vehicle number. Camera located inside vehicle will transmit real time video to see current situation of passengers inside vehicle.

III. METHODOLOGY

A. Input Module

The Input Module peruses sensor information on increasing speed, turn and power and passes the gathered information to the Implanted Processor. The accelerometer is additionally utilized to compute the speed of the vehicle that is utilized as a part of the accident detection logic. The Gyroscope detects the rotation/tilt of the car and peruses the information in the wake of preparing in degrees every second. The four power sensors situated at each side of the car identify the effect power of the mishap.

B. Embedded Processor

The Embedded Processor assumes the part of an interpreter. It incorporates a flag handling module that specimens the adjusted information consistently, and a Bluetooth module that sends the adjusted information to the cell phone. What's more, utilizing the readings of the accelerometer, the speed of the vehicle is computed and utilized by the choice help segment in the cell phone.

C. Bluetooth Module

We have used two bluetooth modules i.e,one that is included in the embedded unit and the other is included in our smartphone.The one used in the embedded unit is HC-05 Bluetooth module.This module keeps receiving processed data from the arduino.On Accident detection the data is sent to the bluetooth module of our phone. As soon as we open the application in our smartphone,the bluetooth module is automatically switched on.The application runs in the background continuously.

D. Smartphone

The mobile phone application acts as the accident detection module as well as the way to send notification to emergency services. It had the accident detection algorithm, nested if else logic and the reaction module that enables sharing of accident data with user's emergency contacts and nearest hospital and police station. The Bluetooth module of the cell phone collects data from the embedded system.

E. Nested If-else

A nested function (or nested procedure or subroutine) is a function which is defined within another function, the enclosing function. Due to simple recursive scope rules, a nested function is itself invisible outside of its immediately enclosing function, but can see (access) all local objects (data, functions, types, etc.) of its immediately enclosing function as well as of any function(s) which, in turn, encloses that function. Suppose if an acceleration value is greater than or equal to the threshold value automatically a message is sent to the emergency contacts as "Accident Detected". If the acceleration value is less than threshold value then it means "No accident".

F. GPS Module

A GPS is a worldwide route satellite framework that gives geolocation and time data to a GPS recipient anyplace on or close to the Earth where there is an unhindered viewable pathway to at least four GPS satellites.The GPS framework does not require the client to transmit any information, and it works freely of any telephonic or web gathering, however these advances can upgrade the value of the GPS situating data. The GPS framework gives basic situating abilities to military, common, and business clients around the globe.

IV. ALGORITHM AND FLOWCHART

Accident Detection Algorithm:

Step 1) Setting up Threshold values for the different sensor values.

Step 2) Creating Rule base that should be satisfied for accident to be detected using Nested if-else.

Step 3) User phone number verification during first login using OTP.

Step 4) Once verified,user needs to fill registration form.

Step 5) Then the application simply runs in the background in correspondence with smartphone bluetooth.

Step 6) Collection of sensor data from embedded module

Step 7) Feeding the data to smartphone application AADNS.

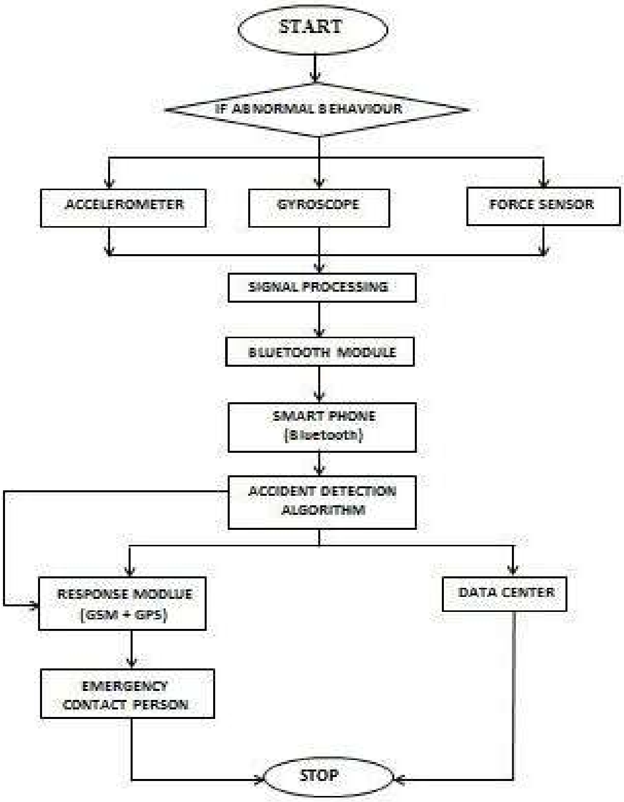
Step 8) Comparing the received values with the set threshold values.

Step 9) If the received values are equal to or greater than threshold values , then accident will be detected.

Step 10 On accident detection,alarm goes off to alert the driver that if he/she is safe they can shut the alarm.

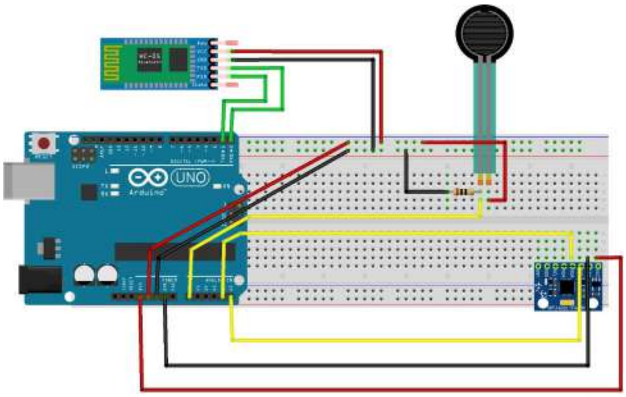
Step 11) On completion of 30 seconds the application automatically send a message to emergency contacts and emergency services.

Step 12) The message includes the current location acquired through gps system,the vehicle plate number and the blood group of user(collected during registration).



V. IMPLEMENTATION

The input module of the proposed system that comprises accelerometer (MPU-6050), gyroscope and force sensors (4-6) collect information from the vehicle. These input systems send information to microcontroller processor(Arduino uno). It transfers the information to the bluetooth module which then sends data to the android application. This application is run on a smartphone and it takes the location details from Network provider and sends message to concerned authority.



Accelerometer: This 3-axial component acquires the data about the current acceleration of the car along three orthogonal axes. The accelerometer is also used to calculate the speed of the vehicle that is used in the Accident Detection module.

Gyroscope: The Gyroscope senses the rotation/tilt of the car and reads the data after processing in degrees per second. This rate of rotation is used for evaluating if the car has rotated to its side or flipped completely.

Force Sensor: The force sensor located at front side of the car detects the impact force of the accident.

VI. TEST CASES AND RESULTS

CASE 1: When the car collides with any object with great impact - In this case the car is travelling with an average speed and then collides with another object with great impact, the resultant output would be that an accident has been detected and the alarm begins to ring for 30 seconds. If the alarm is turned off before the timer goes off i-e the traveller is safe and does not need emergency services.Hence the SMS won't be sent to the emergency services.Otherwise the SMS will be sent to the Emergency services for help.

CASE 2: When the car experiences collision from the sides or back - In this case the car is travelling or is at halt and experiences a collision from the sides or back of the car. If the collision is with great impact i-e higher than the threshold value ,the alarm begins to ring. If the alarm is not turned off, emergency services are contacted through SMS.

CASE 3: When the car collides with any object but with less force - In this case the Car is travelling with an average speed and then collides with another object with less force/impact. The impact experienced by the car is very less i-e less than the threshold value for an accident to be detected. Hence no accident is detected.

CASE 4: When the car rolls over in an accident - In this case the car while travelling meets with an accident in such a way that it experiences a roll over. The orientation of the car changes along with an impact experienced on it.Hence an accident is detected. This is assumed to be a critical situation, therefore no alarm will ring and the message to the emergency contacts and services will be sent for immediate help without wasting a second.

CASE 5: When the car experiences sudden deceleration - In this case, when driver of the car suddenly applies brakes,the car experiences a drop in acceleration. Since no impact or roll over is detected , we can conclude that no accident has occurred.

CASE 6: When the car is travelling at an elevated path - In this case , the car is travelling on an elevated platform. Example - Hilly areas, where the roads are steep and the car makes certain angle with the ground. This changes the orientation of the car but accident is not detected.

VII. Acknowledgment

We remain immensely obliged to my project guide Prof. Payel Thakur , for her valuable guidance, patience, keen interest and constant encouragement and for her invaluable support.

I would like to thank my college Pillai College of Engineering and Dr.Madhumita Chatterjee, H.O.D. of Computer Department and Dr. Sharvari Govilkar, H.O.D of Information Technology Department for their invaluable support.

I would also like to thank Dr. R.I.K Moorthy, Principal for his invaluable support and for providing an outstanding academic environment. I would also like to thank all the staff members of the department of Computer and Information Technology.

VIII CONCLUSION

Accident information would reach the emergency services within seconds.Significantly improves the time gap for rescue operation and save the life of huge number of victims. Victims personal details would be easily obtained through his registration with this application. Alert messages are send through GPS. Accelerometer and gyroscope is used here in order to detect the plausibility of an accidents.

IX FUTURE SCOPE

This report presents the techniques and algorithm that will be used to develop AADNS system. The comparative study of various other accident detection approaches being used elsewhere in the world is presented in this report. And also how our system is preferable to those mentioned. The use of GPS/GSM module in the embedded system will help locate the victim in case the mobile phone gets damaged. Use of commercial sensors will help bring more accuracy.

X. References

[1] "Auto Security | Car Safety | Navigation System | OnStar." OnStar. N.p.n.d. Web. 15 June 2014.

[2] "Vikram Singh Kushwaha , Deepa Yadav, Abusayeed Topinkatti , Amrita Kumari"-"CAR ACCIDENT DETECTION SYSTEM USING GPS AND GSM"-International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS) ISSN (Online) 2349-6967 Volume 2 , Issue 1(Jan-Feb 2015), PP12-17

[3]"Dnyanesh Dalvi,Vinit Agrawal,Sagar Bansod,Apurv Jadhav, Prof. Minal Shahakar"-"Android Application for automated accident detection"

IJARIIE-ISSN(O)-2395-4396-Vol-3 Issue-2 2017

[4]Kajal Nandaniya, Nadiad Viraj Choksi,Ashish Patel Assistant professor, Nadiad M B Potdar- Automatic Accident Alert and Safety System using Embedded GSM Interface -International Journal of Computer Applications (0975 – 8887) Volume 85 – No 6, January 2014

[5] "Real Time Traffic Accident Detection System Using Wireless Sensor Network"-"M.Amer Shedid,Hossam M. Sherif,Samah A. Senbel"-International Conference of Soft Computing and Pattern Recognition

Recommendation System using Jaccard Indexing

Gaurav Biswas[1], Shailesh Kotian[2], Vikas Singh[3] , Madhu Nashipudimath[4]
Department of Computer Science
Pillai College of Engineering
Panvel, Maharashtra, India
[1]gauravbiswas843@gmail.com, [2]shailesh.kotian3@gmail.com, [3]vikasingh96@gmail.com,
[4]madhumn@mes.ac.in

***Abstract*- With the rapid growth of internet, there are loads of data being generated which is very important for any online business. As a result of the E-commerce industry's growth there is a competition to create a better recommendation system in order to increase profit and retain buyers. Recommendation system helps users to discover products or contents that they may not have come across otherwise. The paper presents an Coextensive Jaccard Indexing algorithm for Book Recommendation. The System uses collaborative based filtering technique to recommend books for users. Recommendation is based on the ratings of K- Nearest neighbours. Also, this paper presents an experimental implementation of the proposed algorithm.**

Keywords- Book Recommendation, Jaccard Indexing, Similarity, Rating.

I. INTRODUCTION

In the last decade there has been a tremendous growth of technology. Nowadays we have better, faster and more effective means to connect to internet and world. Internet speed has exponentially increased and now almost everyone is connected to internet. This development played the main role in the growth of E-commerce and various online services. Internet is full of users structured and unstructured data. E-commerce requires to create virtual profile of users which help vendors to provide personalised experience to the users. There is huge amount of content/product that is being generated daily and it is not possible for users to manually search for these content. As a result e-commerce services do performs this search and provides personalised recommendation to the users. In order to survive in this market, vendors need to build better recommendation system which will provide relevant suggestions to the users. Recommendation System are mainly classified into two types Collaborative Filtering and Content-Based filtering. Recommendation system requires to find similarity between different attributes like user-user, item-item, user-item etc. There are various techniques to find this similarity such as Cosine Similarity, Pearson's Correlation, Jaccard Similarity etc. The choice of filtering technique and similarity measure varies depending upon the need and scope of the project.

In this paper we propose an algorithm that recommends books to readers. We developed a system which implements Hybrid filtering technique and uses Jaccard Similarity to find similarity between users. Hybrid filtering utilizes collaborative filtering to find similar users to predict the liking of the users and content based filtering to overcome cold start problem. We developed a system which learns users preferences based on the previous ratings and genre of interest. The system then generates the list of recommendation that the user most probably would like to read.

The paper is organised as follow. In Section II literature review on related researches is provided. Section III provides the detailed explanation about the implementation of proposed system. In Section IV we have conclusion that we have obtained from this system. Finally in Section V we have future scope of the project which includes ideas that can be included to improve the performance of the system.

II. RELATED WORK

Peter Bostrom and Melker Filipsson [1], proposed a paper on "Comparison of User Based and Item Based Collaborative Filtering Recommendation Services". The main intention of this work was to evaluate the performance of user based and item based collaborative filtering on sample dataset. Based on their observations they concluded that user based collaborative filtering is superior on all of the tested cases and also improves faster as the amount of training data is increased.

Nursultan kurmashov, Konstantin Latuta and Abay Nussipbekov [2], proposed a paper on "Online Book Recommendation System". They have used collaborative filtering method which provides fast recommendations to their users without need to be registered for a long time and have big profile information, browsing history and etc.

Praveena Mathew, Bincy Kuriakose and Vinayak Hegde [3], proposed a paper on "Book Recommendation System through Content Based and Collaborative Filtering Method". They have used association rule mining algorithm to find interesting association and relationship among large data set of books and provide efficient recommendation for the book.

Simon Philip, P.B. Shola and E.P. Musa [4], proposed a paper on "A Paper Recommender System Based on the Past Ratings of a User". They used content-based filtering technique to suggests or provides recommendations to the intended users based on the papers the users have liked in the past.

Mahmud Ridwan [5], proposed an article on "Predicting Likes: Inside a Simple Recommendation Engine Algorithm". In this article he explains about how to use Advance Jaccard Similarity measure to find the similar users and to predict the possibility value of users liking a book.

Madhuri Angel Baxla [6], proposed a paper on "Comparative Study of Similarity Measures for Item Based Top N Recommendation". In this paper they analyze different Similarity measures based on various range of users. They concluded that extended jaccard takes least time to recommend items.

Jian Li, Yajie Wang, Jun Wu and Fengmei Yang [7], proposed a paper on "Application of User-based Collaborative Filtering Recommendation Technology on Logistics Platform". The paper introduces user-based collaborative filtering recommendation technology on the logistics platform, to improve the operational efficiency of the logistics platform and to achieve the rational allocation of logistic resources.

Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton and Ashesh Mahidadia [8], proposed a paper on "Learning Collaborative Filtering and Its Application to People to People Recommendation in Social Networks". They have proposed an approach for people recommendation by collaborative filtering and Machine Learning. The proposed learning algorithm is able to rank the recommendations in order to further improve the success of predicted user interactions. The proposed algorithm outperforms all other methods including standard CF as measured on both Precision(SR) and recall.

III. METHODOLOGY

A. Collaborative Filtering

Collaborative filtering is commonly used to build personalized recommendations on web. Collaborative filtering methods are based on gathering and analyzing a large amount of information on users' preferences and predicting what users will like based on similarity to other users. The main advantage of collaborative filtering is that it does not require an understanding of items. Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future, and they will like similar kinds of items as they liked in past. Coextensive Jaccard Similarity is used to find the similarity between users and possibility value of user liking a book.

The user-based collaborative filtering book recommendation system can be divided into five steps: Data collection, Calculating related users' similarity, Selecting neighbour users, Calculating possibility value, Produce recommendation results.

Step 1: Data collection

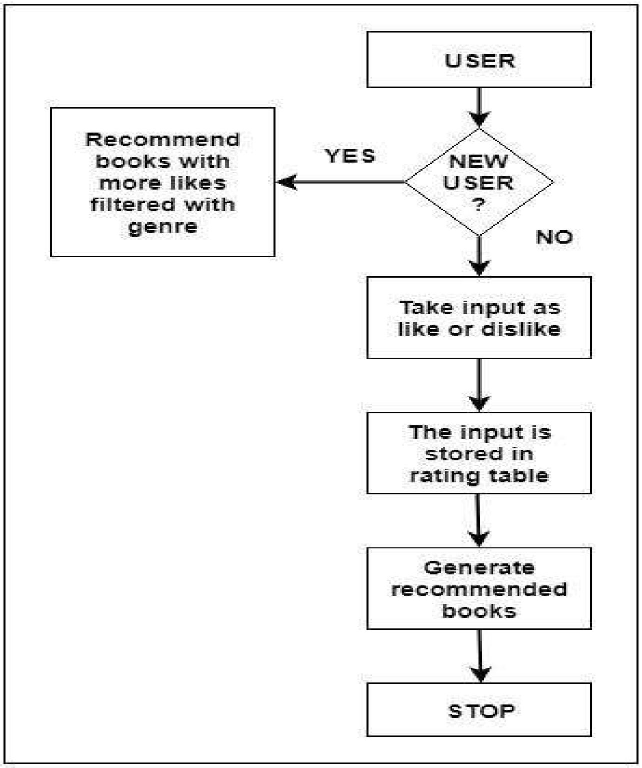

USER
NEW USER ?
YES
NO
Recommend books with more likes filtered with genre
Take input as like or dislike
The input is stored in rating table
Generate recommended books
STOP

Figure 1: Recommendations using ratings

Collaborative filtering produces recommendation based on the past evaluation of the users. The evaluation is stored in a rating table which consist of three fields namely user_id, book_id, rating.

Step 2: Calculating related users' similarity

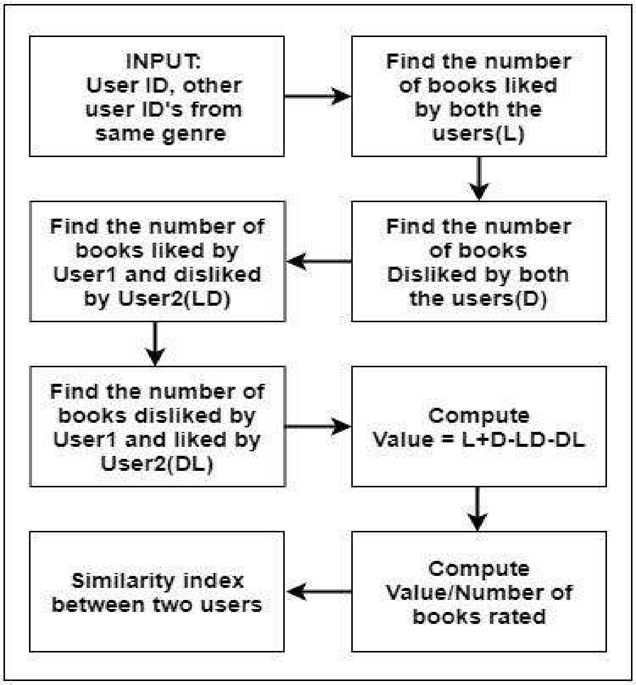

INPUT: User ID, other user ID's from same genre
Find the number of books liked by both the users(L)
Find the number of books Disliked by both the users(D)
Find the number of books liked by User1 and disliked by User2(LD)
Find the number of books disliked by User1 and liked by User2(DL)
Compute Value = L+D-LD-DL
Compute Value/Number of books rated
Similarity index between two users

Figure 2: Similarity between two users

To find the neighbour users we need to calculate users' similarity. We have used the Coextensive Jaccard Similarity to calculate similarity between users. Before calculating the similarity, users are filtered based on genre and then jaccard similarity is applied on users from same genre. The similarity between the user U_1 and U_2 is calculated by the formula as follows:

$$S(U_1,U_2) = \frac{|L1 \cap L2| + |D1 \cap D2| - |L1 \cap D2| - |L2 \cap D1|}{|L1 \cup L2 \cup D1 \cup D2|} \tag{1}$$

In the above formula L and D stands for likes and dislikes rated by users. The similarity between two users is represented using decimal number between -1.0 and 1.0.

Step 3: Selecting neighbour users

Similarity of related users are arranged in descending order and top K users are selected for further processing. So the users from same genre who have least similarities are filtered in this step.

Step 4: Calculating possibility values

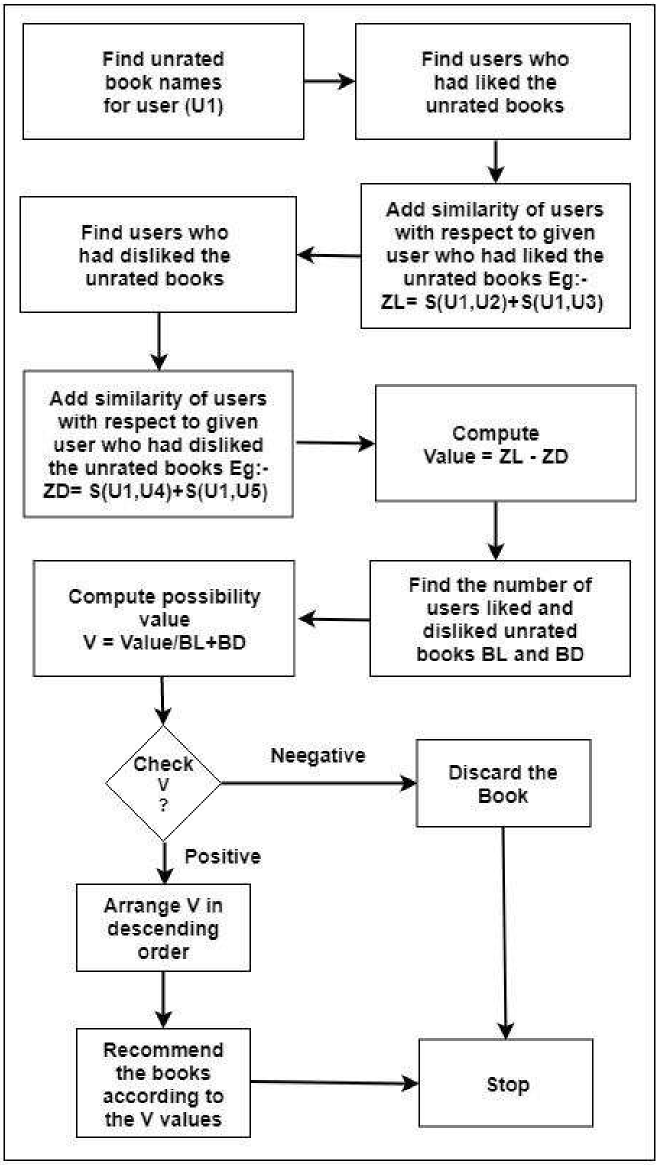

Find unrated book names for user (U1)
Find users who had liked the unrated books
Add similarity of users with respect to given user who had liked the unrated books Eg:- ZL= S(U1,U2)+S(U1,U3)
Find users who had disliked the unrated books
Add similarity of users with respect to given user who had disliked the unrated books Eg:- ZD= S(U1,U4)+S(U1,U5)
Compute Value = ZL - ZD
Find the number of users liked and disliked unrated books BL and BD
Compute possibility value V = Value/BL+BD
Check V ?
Neegative
Discard the Book
Positive
Arrange V in descending order
Recommend the books according to the V values
Stop

Figure 3: Computation of user possibility value

Now the similarity of K-nearest neighbours are used to find the possibility value. The possibility of user U liking book B is calculated by the formula as follows:

$$P(U,B)= \frac{ZL - ZD}{|BL| + |BD|} \qquad (2)$$

Z_L and Z_D are the sum of the similarity indices of user U with K-nearest neighbours who have liked or disliked the book B, respectively. $|B_L| + |B_D|$ represents the total number of users who have liked or disliked the book B. The result P(U,B) produces a number between -1.0 and 1.0.

Step 5: Produce recommendation results

After predicting the possibility value of books that are not rated by users, we rank those books in descending order of possibility value. The first P books will be recommended to the users.

B. Content based filtering

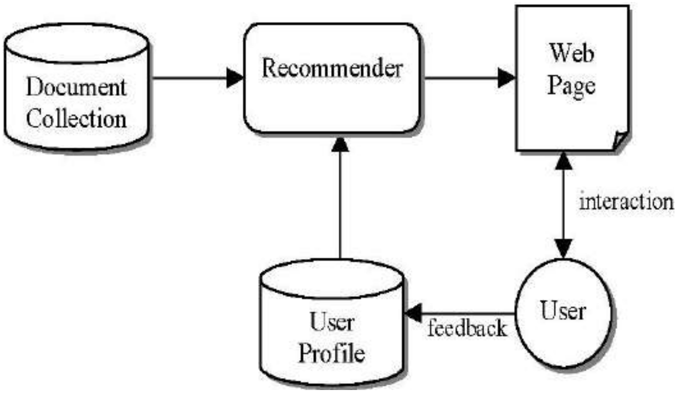


Figure 4: Architecture of Content Based Filtering

Recommendations based on collaborative filtering works fine for existing users who have rated some books in the past. But if we have a new user or a user who has never rated any book in the past then our system fails i.e system suffers from cold start problem.

In order to overcome this problem we can use content based filtering. This technique is implemented by creating profile of users. System generates a survey form for new users to get information about the user like their favourite genre. Now the books with most likes from user specified genre are recommended to user.

IV. RESULT ANALYSIS

Books are recommended using collaborative based filtering technique. The rating matrix is prepared with the conditions of users who has rated the highest number of books and books that have highest ratings. The first step is to find the genre-rating matrix with the relevant information i.e. users who have rated the highest number of books and books that have highest ratings. Now books can be recommended for the users.

For given user, the algorithm goes through genre table. For each user in genre table, it identifies all its common rating and compute the similarity between two user using Formula 1. Now possibility value for unrated book is calculated using Formula 2. Then the algorithm finds the top recommendations based on highest possibility value. The quality of a domain system can be evaluated by comparing recommendations to a test set of known user ratings.

Accuracy can be measured by comparing the recommendations with the likes. Each row of the table corresponds to a different split of recommendations. Column name in the below tables are defined as follows:

- **True Positives (TP):** These are the recommended books that have been liked.
- **False Positives (FP):** These are the recommended books that have been disliked.
- **Unknown Positives (UP):** These are the recommended books that haven’t been rated.
- **False Negatives (FN):** These are not recommended books that have been liked.

These systems are typical measured using precision and recall.

Table 1: Confusion Matrix for Basic Jaccard

Recommen-dations	TP	FP	UP	FN	Precision J	Recall J
5	1.4	0.8	2.8	32.6	0.280	0.0412
10	3.0	1.5	5.5	30.5	0.300	0.0896
15	5.1	2.3	7.6	33.1	0.340	0.1335
20	6.9	3.0	10.1	31.2	0.345	0.1811
25	9.0	3.7	12.3	29.9	0.360	0.2313
30	11.7	4.4	13.9	31.7	0.390	0.2696

Table 2: Confusion Matrix for Coextensive Jaccard

Recommen-dations	TP	FP	UP	FN	Precision CJ	Recall CJ
5	1.8	0.3	2.9	32.4	0.360	0.0526
10	3.9	0.7	5.4	30.2	0.390	0.1144
15	6.3	0.9	7.8	33.3	0.420	0.1590
20	9.1	1.3	9.6	31.4	0.455	0.2247
25	11.5	1.6	11.9	29.7	0.460	0.2791
30	14.1	1.7	14.2	31.9	0.470	0.3065

Ratings shown in table 1 and table 2 are obtained by experiment conducted (recommendation given to the readers) on 10 users. Performance of the recommendation system can be evaluated by comparing it with the results of existing system.

Precision(P): A measure of exactness, determines the fraction of relevant items retrieved out of all items retrieved. It is the proportion of recommended books those are actually good.

$$P = \frac{TP}{TP+FP+UP} \qquad (4)$$

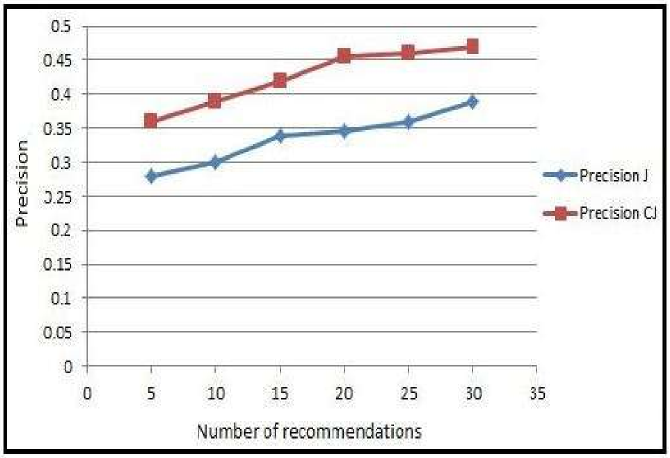


Figure 5: Comparison based on Precision value.

From above graph it can be concluded that Coextensive Jaccard has better precision than Basic Jaccard Indexing. Also, precision of the system increases with the number of books recommended.

Recall(R): A measure of completeness, determines the fraction of relevant items retrieved out of all relevant items. It is the proportion of all good books recommended.

$$R= \frac{TP}{TP+FN} \qquad (5)$$

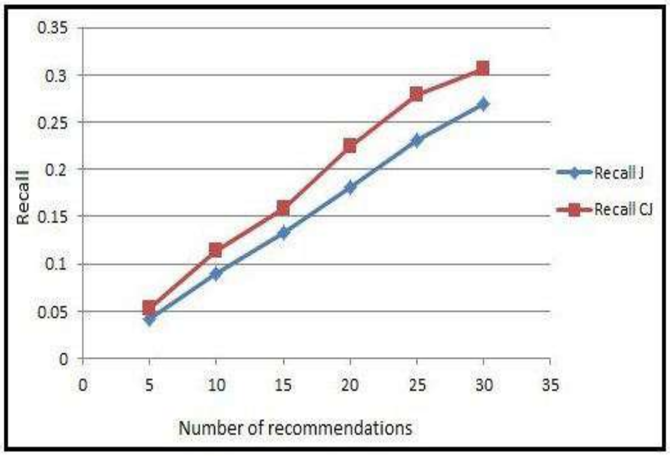


Figure 6: Comparison based on Recall value.

From above graph it can be concluded that Coextensive Jaccard has better recall than Basic Jaccard Indexing. Also, recall of the system increases with the number of books recommended.

V. CONCLUSION

Due to exponential growth of technology a huge amount of data is being generated which can be very useful if used properly. Recommendation system works on the data generated by the user and tries to find a pattern to predict the future interest of the users. Book Recommendation System is recommending books to users according to their past interest and stores recommendations in the users' web profile. It uses user-based collaborative filtering to find out the list of books based on ratings. The system provides better recommendation if we have sufficiently large datasets. This system gives better exposure to users about the new books which otherwise they would have never known about. The system can be used in library to to recommend books to their customers. Since the system is not content oriented it can be easily deployed in any other domain like Movie recommendation or Clothing recommendation etc.

There are many implicit data collection techniques such as Analyzing the books that the user views, Observing book view time, Keeping a record of read list of users, Monitoring the search history of users etc. It contains many vital information about the users that can be helpful in improving the performance of the system. Machine learning concepts can also be implement to build a model-based collaborative system that can give great recommendations.

VI. REFERENCES

[1] Peter Bostrom and Melker Filipsson. Comparison of User Based and Item Based Collaborative Filtering Recommendation Services. Examensarbete Inomteknik, Grundniva, 15 HP Stockholm, Severige 2017.

[2] Nursultan Kurmashov, Konstantin Latuta and Abay Nussipbekov. Online Book Recommendation System. Faculty of Engineering and Natural Sciences Suleyman Demirel University Kaskelen, Kazakhstan, 2016.

[3] Ms. Praveena Mathew, Ms. Bincy Kuriakose and Mr.Vinayak Hegde. Book Recommendation System through Content Based and Collaborative Filtering Method. Department of Computer Science Amrita Vishwa Vidyapeetham Mysuru Campus Mysuru, Karnataka, 2016.

[4] Simon Philip, P.B. Shola and E.P. Musa. A Paper Recommender System Based on the Past Ratings Of a User. International journal of advanced computer technology (IJACT), 2015 .

[5] Mahmud Ridwan. Predicting Likes: Inside A Simple Recommendation Engine's Algorithm[Online]. https://www.toptal.com/algorithms/predicting-likes-inside-a-simple-recommendation-engine

[6] Madhuri Angel Baxla. Comparative study of similarity measures for item based top n recommendation. National Institute of Technology Rourkela, 2014.

[7] Jian Li, Yajie Wang, Jun Wu and Fengmei Yang. Application of User-based Collaborative Filtering Recommendation Technology on Logistics Platform. Sixth International Conference on Business Intelligence and Financial Engineering, 2013.

[8] Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton and Ashesh Mahidadia. Learning Collaborative Filtering and Its Application to People to People Recommendation in Social Networks. University of New South Wales, Sydney NSW 2052, Australia, 2011.

Cashless Transactions Over Social Media Using Bots

Akshay Lanke akki95lanke@gmail.com

Nikhil Sahani nikhil@tuta.io

Prashant Arghode prashant.arghode28@gmail.com

Subodh Chalke subodh@keemail.me

Department of Computer Engineering, Mumbai University
Pce, New Panvel, India

Abstract:

In this modern era, cashless payments is the buzzword. Cashless payments allows one to send/receive money with ease. However most applications which enables a person to perform cashless transactions are either confusing or not compatible with each other. Social Money Bot will allow users to use their own choice of social media account to send/receive money. Using the Social Money Bot users can send money directly via their social media chat window or profile page. Thus there is no need to install a separate app in order to use the services provided by the Social Money Bot. Python's NLP library,

Natural Language Toolkit (NLTK) provides various functions to analyze and manipulate strings[1]. KnuthMorrisPratt pattern matching algorithm can be used to understand semantic meaning of strings[6]. The platform also provides security by using hashing for encrypting all the important details of the user[2]. Thus, providing security and simplicity to the user.

I.INTRODUCTION

Cashless transactions allows one to transfer money from one point to another with ease. Cashless transactions are the one where the payments are done by the means other than physical cash. Cashless transaction basically means that there will be no flow of physical cash among the people. Every transaction will be through electronic media or credit cards, bank transfers, checks etc. Compared to cash transactions, cashless transactions are less expensive to manage[6]. Social Money Bot Enables people to perform
cashless transactions using Social Media platforms such as Telegram,Whatsapp, Facebook etc. Thus eliminating the need of having a separate application for performing cashless transactions. Making the process simple and easy. To achieve the project makes use of string matching algorithms, API in order to support various platforms easily and hashing to protect users details.

II.LITERATURE REVIEW

[1]Survey on Chatbot Design Techniques in Speech Conversation Systems, Sameera A. Abdul-Kader, Dr. John Woods,Vol. 6 , No. 7, 2015
Sameera A. Abdul-Kader and Dr. John woods mentions all the techniques such as string matching, parsing, SQL and Database, chat script and AIML etc and tools such as Natural Language ToolKit (NLTK) which is free plugin for python to work with NLP. This paper presents a survey on the techniques used to design Chatbots and a comparison is made between different design techniques from nine carefully selected papers according to the main methods adopted. These papers are representative of the significant improvements in Chatbots in the last decade. The paper discusses the similarities and differences in the techniques. The paper also examines in particular the Loebner prize - winning Chatbots.
Techniques available for string matching are RabinKarp string search algorithm, Nave string search algorithm, BoyerMoore string search algorithm and KnuthMorrisPratt algorithm. The one which we are using is KnuthMorrisPratt algorithm because it is mor efficient then others.

[2]Method to Protect Passwords in Databases for Web Applications, Scott Contini, 2015.
Scott Contini proposes that the password should be stored as hash(s) where s = PPF (salt ,password , cost , misc). PPF is password processing function.The purpose of this research note is to present a solution with complete details and a concise summary of the requirements, and to provide a solution that developers can readily implement with confidence, assuming that the solution is endorsed by the research community. The proposed solution involves client - side processing of a heavy computation in combination with a
server-side hash computation.
Passwords can be hashed, encrypted or hashed with a salt in order to protect the passwords stored in the database. This system will use default password function of php to store the passwords in the database. The function creates a hash of the password with a random salt.

[3]A Survey of Methods for Preventing Race Conditions,Nels E. Beckman, May 10, 2006.
In this paper, Nels E. Beckman considers several different styles of software analysis, and their effectiveness at alleviating one very specific software defect: race conditions in concurrent software. In this paper, four different analysis styles were compared, all with the goal of detecting or preventing race conditions. Race conditions are a devious form of bug, and therefore the effectiveness of these techniques is of great interest. The techniques surveyed varied widely in the characteristics of their operation, but in the end, it seems as if a flow-based analysis would be the best tool for finding race conditions in an industrial
setting, at least at this point in time.
Paper mentions methods such as Flow-Based Race Analysis, Using Model-Checking to Detect Race Conditions, Dynamic and Hybrid Race Detectors and Race-Free Type Systems. The system uses simple locking mechanism in order to prevent race condition.

III. PROPOSED ARCHITECTURE

The main goal of this project was to enable people to perform cashless transactions over social media using their day to day social media applications. Social Money Bot has a website where once the user registers an account and link his/her social media account with the website he/she will be able to use that social media account to carry out cashless transactions or the user can interact with the bot itself to do the same.

The figure 3.1 depicts the overall architecture of the proposed system. It consists of a main server which uses API to communicate with bots of all the platforms. Systems website also use the same API.

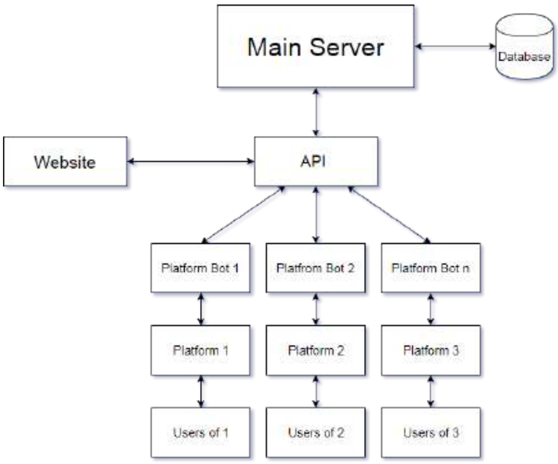


Figure 3.1: Block Diagram

IV. EXPERIMENTS AND RESULTS

4.1 Sample of Inputs/Dataset/Database Used/ and

[1] Sample Input: send test@test.com +1000

Social Money Bot breaks down the input while using algorithms for string searching to understand the input and perform the action. Here the action is send 1000 to user with email address test@test.com

[2] Sample Input: +1000

Replying +1000 to a message of an user will result in Social Money Bot retrieving user-name of the sender of the orignal message and send 1000 to that user.

[3] Sample Input: test@test.com

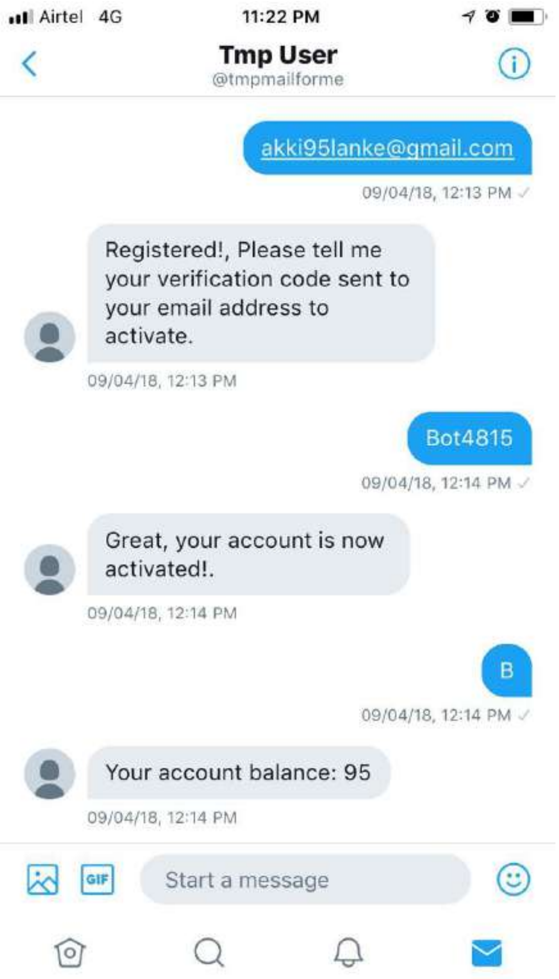
One can register an account by simply sending email address to the Social Money Bot

[4] Sample Input: Other commands such as ”balance”

Sending command or text ”balance” to the bot will result in bot retrieving the balance of the user from whom the text is coming from and responding him/her with the amount of money he/she has in his/her account.

[5] Sample Input: botxxxx

”bot” followed by four random digits is used as otp.



akki95lanke@gmail.com
09/04/18, 12:13 PM

Registered!, Please tell me your verification code sent to your email address to activate.
09/04/18, 12:13 PM

Bot4815
09/04/18, 12:14 PM

Great, your account is now activated!.
09/04/18, 12:14 PM

B
09/04/18, 12:14 PM

Your account balance: 95
09/04/18, 12:14 PM

Output

V. CONCLUSION

To give an overview of our project, it is basically about enabling people to send/receive money using social media platforms. To do so our platform must interact with the users on the platform and the platform itself for which it uses the API of the said platform.

Once it is connected to the platforms API, it uses its own API to process the data (Messages, Notification) received from the platform API. Projects API uses algorithm Such as KnuthMorrisPratt algorithm to process and take actions on data received from the users via bot on the social media platform. Our project removes the need of having an separate application to perform cashless transaction while improving the issue of compatibility (not being able to send/receive money to/from two separate applications).

VI. FUTURE SCOPE

The Social Money Bot can be further expanded various different field such as:

1. Cashless Transactions.

Designating or of financial transactions handled as by means of credit cards, bank transfers, and checks, with no bills or coins handed from person to person. Social Money Bot allows one to perform cashless transactions with ease.

2. Remittance, using Cryptocurrency.

A remittance is a transfer of money by a foreign worker to an individual in his or her

home country. Money sent home by migrants competes with international aid as one of the largest financial inflows to developing countries[6]. With the help of Cryptocurrencies and Social Money Bot this process can be eased.

3. Shopping.

Since Social Money Bot allows you to hold funds in its account, People can buy products such as Mobile airtime using the balance available in the system.

4. More Social Media Platforms.

Using Social Money Bot API, more platforms can be integrated into the system easily.

Thus expanding the ecosystem of the Social Money Bot.

REFERENCES

[1] (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.
6, No. 7, 2015 Survey on Chatbot Design Techniques in Speech Conversation Systems,
Sameera A. Abdul-Kader, Dr. John Woods.

[2] Method to Protect Passwords in Databases for Web Applications, Scott Contini 2015

[3] A Survey of Methods for Preventing Race Conditions, Nels E. Beckman, May 10,
2006

[4] ”ChangeTip” wiki available at https://en.wikipedia.org/wiki/ChangeTip

[5] ”Dogecoin” wiki available at https://en.wikipedia.org/wiki/Dogecoin

[6] ”Google” Search the world’s information, including webpages, images, videos and
more. Available at https://www.google.co.in/

[7] ”Stack Overflow” is the largest, most trusted online community for developers to learn,
share their programming knowledge, and build their careers. Available at https://stackoverflow.com

[8] ”Wikibooks” is a wiki-based Wikimedia project hosted by the Wikimedia Foundation for the creation of free content e-book textbooks and annotated texts that anyone
can edit. Available at https://www.wikibooks.org

[9] ”Python DevDocs” Python 3.6.4 API documentation with instant search, offline support, keyboard shortcuts, mobile version, and more.
Available at http://devdocs.io/python/

Credit Card Fraud Detection

Using Hidden Markov Model

Prof.Deepti Lawand
(dlawand@mes.ac.in)
Sayyed Shadab
(sayyedshadab65@gmail.com)

Sayyed Shazeb
(sayyedshazeb@gmail.com)

***Abstract*— Nowadays, the usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In this report, we model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time,our system will ensure that genuine transactions are not rejected. The proposed system examines the behavior of the user and calculates the threshold value of his purchase and if the user do any transaction valid user will receive a message to verify OTP (One Time Password) and that user will enter the OTP. If the current purchase value of transaction is below than threshold value then the user have to enter OTP as well as answer security question. If the answer to the security question is wrong then the card is blocked automatically. If the current purchase value of transaction is above the threshold value then the user have to enter OTP, answer security question and key logging with QR code. If any of the above security mechanism is not proved correctly then the card is blocked automatically.**

I. Introduction

This chapter introduces the currently existing techniques and an analysis of previous research related to our proposed methodology. The related research is described as a base for our approach. The chapter also describes features of software and hardware used in developing this report, what this re is all about, its objective and scope.Globalization and increased use of the Internet for Online Shopping has resulted in a considerable increase in Credit Card Transactions throughout the world. Credit card fraud is the criminal offence in which accused make use of others credit card in absence of the actual owner of the card to utilize or withdraw the money from the owner's account. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company and account holder.

The most efficient way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the “usual” spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc.

II. Existing System / Scenario and flaws

In case of the existing system the fraud is detected after the fraud is done that is, the fraud is detected after the complaint of the card holder. And so the card holder faced a lot of trouble before the investigation finish. And also as all the transaction is maintained in a log, we need to maintain a huge data. And also now a days lot of online purchase are made so we don’t know the person how is using the card online, we just capture the IP address for verification purpose. So there need a help from the cyber-crime to investigate the fraud. To avoid the entire above disadvantage we propose the system to detect the fraud in a best and easy way.

Following are the flaws of the existing system:-

- Detection of fraud is slower process. Due to this card holder has to suffer a lot before finishing of investigation.
- Chances of loss of data because there is need to maintain a huge data.
- Less chances to get information of the person who is doing a fraud transaction.
- The process gets slower because first image of IP address is captured and then help of cyber-crime is taken.

Steps to avoid all the above given flaws:-

- Check the withdrawal behavior of the person who is making trasactions.
- Detect fraud at the time of withdrawal.
- No need to maintain a log of data.
- At the time of withdrawal if necessary security blocking is done so we can catch the person who is making fraud.
- No need to capture IP address.

III. Proposed System

In proposed system, we present a Hidden Markov Model (HMM),which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. Card transaction processing sequence by the stochastic process of an HMM. The details of items purchased in Individual transactions are usually not known to any Fraud Detection System(FDS) running at the bank that issues credit cards to the cardholders. Hence, we feel that HMM is an ideal choice for addressing this problem.Another important advantage of the HMM-based approach is a drastic reduction in the number of False Positives transactions identified as malicious by an FDS although they are actually genuine. An FDS runs at a credit card issuing bank. Each incoming transaction is submitted to the FDS for verification. FDS receives the card details and the value of purchase to verify, whether the transaction is genuine or not.The types of goods that are bought in that transaction are not known to the FDS. It tries to find any anomaly in the transaction based on the spending profile of the cardholder, shipping address, and billing address, etc. then the application will for security questions and we propose the anti keylogging mechanisms like the virtual keyboards which are pertinent today. The server generates the QR code. Then the QR code is sent to the client. On client's terminal, the QR code is displayed. Now, the client has to take his smartphone in which the QR code scanning application is already installed. The QR code has to be scanned. After scanning the QR code, the decoded information will be displayed in the smartphone. The randomized keyboard which looks like a 6x6 matrix or 4x4 matrix with random arrangements of 0-9 digits and A-Z is displayed in the smartphone. On the client's terminal the password box is replaced with the 4x4 blank keyboard matrix. Now, the client has to just click on the rows or columns of the blank keyboard matrix by seeing where is password has been arranged in the smartphone. Through rigorous analysis, we verify that our protocols are immune to many of the challenging authentication attacks applicable in the literature. If the FDS confirms the transaction to be of fraud, then the account gets blocked and the issuing bank declines the transaction.

Advantages:-

1. The detection of the fraud use of the card is found much faster that the existing system.

2. In case of the existing system even the original card holder is also checked for fraud detection. But in this system no need to check the original user as we maintain a log.

3. The log which is maintained will also be a proof for the bank for the transaction made.
4. We can find the most accurate detection using this technique.
5. This reduce the tedious work of an employee in the bank
6. Preventing Keylogger.

Block Diagram

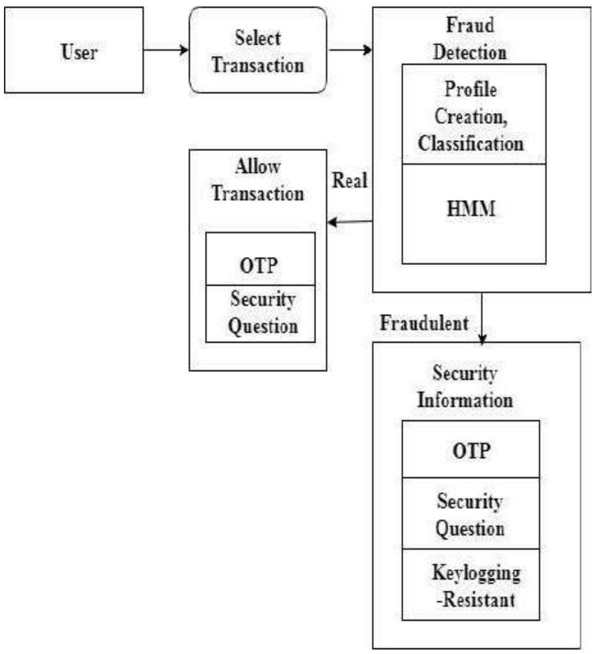


Fig.3.1 Proposed system block diagram

Fig.3.1 Shows the Fraud Detection Architecture, in this user performs an online transaction then it goes to the Fraud Detection System (FDS). In FDS users spending profile is checked with database and also HMM algorithm runs on user previous transactions. If user is authenticated user then FDS allow transaction or if user is unauthenticated user then FDS detects that transaction is fraudulent then it goes to the security system where HMM traces the IP address of the organization from where unauthorized user was trying to gain transaction and it also sends notification on authorized user's mobile number

A. Authorized User

In Fig 3.2, If an authorized user performs an online transaction then his spending profile is matched into our database and if it matches then the transaction is performed successfully and then user is notified that transaction is done successfully.

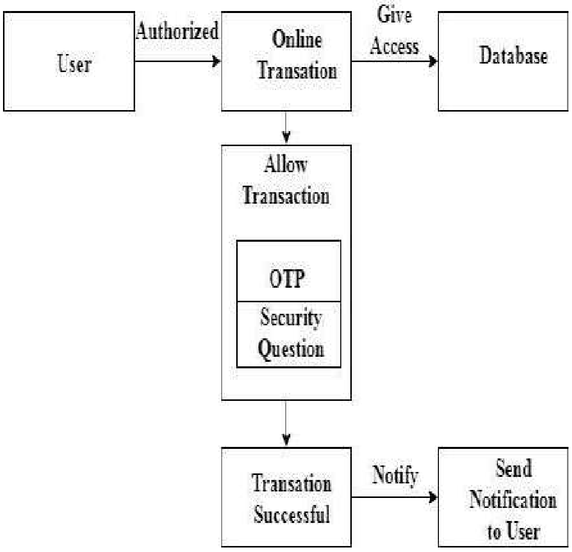
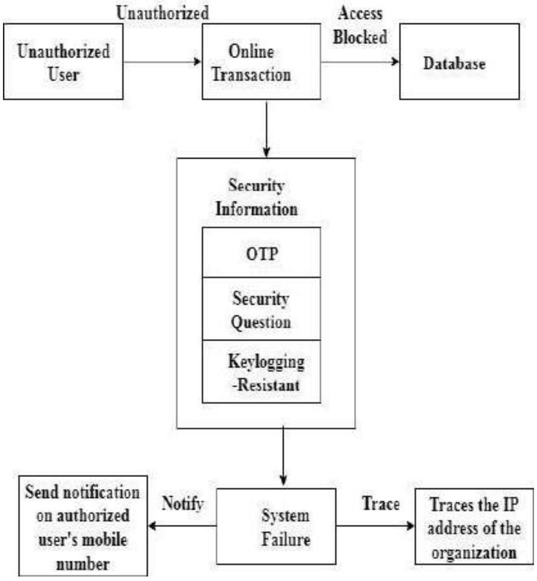


Fig 3.2 Authorized User Access To System

B. Unauthorized User

In Fig 3.3, If an unauthorized user tries to perform an online transaction and if the spending profile doesn't matches into the database then access is blocked to that user and system failure occurs. HMM traces the IP address of the organization from where unauthorized user was trying to gain transaction and it also sends notification on authorized user's mobile number and raises the alarm to Admin System.



Techniques and Algorithm Used

Keylogging-resistant

Remote Desktop Services, formerly known as Terminal Services, is one of the components of Microsoft Windows (both server and client versions) that allows a user to access applications and data on a remote computer over a network, using the Remote Desktop Protocol.

When Sumitomo Mitsui Banking Corporation discovered a keylogger installed on its network in London There have been other high-profile cases in keylogging attack. In 2003, t=he perpetrator installed the software at more than 14 Kinko locations in New York and using it to open bank accounts with the names of some of the 450 users whose personal information he collected [2]. Also in 2003, Valve Software founder Gabe Newell found the source code to his company's Half-Life 2 game stolen after someone planted a keylogger on his computer [3]. Some of the software-based keyloggers are hypervisor based, API-based, kernel-based, form grabbing based, memory injection based, packet analyzers, and remote access software keyloggers.

QR code is developed by Japanese Denso Wave corporation in 1994. It is a two dimensional barcode. There are 40 versions and four levels of error correction in QR code. The barcodes are attached to all sort of products for identification which is a optical machine-readable representation of data. Linear barcodes are one dimensional and have a limited capacity of coding 10 to 22 characters. The QR code has the high capacity which can hold 7,089 numeric, 4,296 alphanumeric, and 2,953 binary characters [1]. QR Code has been approved as an AIM Standard, a JIS Standard and an ISO standard

To record the credit card transaction dispensation process in conditions of a Hidden Markov Model (HMM), it creates through original deciding the inspection symbols in our representation. We Identity number (PIN) with database and account balance of user's credit card is more than the purchase amount, the fraud checking module will be activated. The verification of all data will be checked before the first page load of credit card fraud detection system. If user credit card has less than 10 transactions then it will directly ask to provide personal information to do the transaction. Once database of 10 transactions will be developed, then fraud detection system will start to work. By using this observation, determine users spending profile. The purchase amount will be checked with spending profile of user. By transition probabilistic calculation based on HMM, it concludes whether the transaction is real or fraud. If transaction may be concluded as fraudulent transaction then user must enter security information. This information is related with credit card (like account number, security question and answer which are provided at the time of registration). If transaction will not be fraudulent then it will direct to give permission for transaction. If the detected transaction is fraudulent then the Security information form will arise. It has a set of question where the user has to answer them correctly to do the transaction. These forms have information such as personal, professional, address; dates of birth, etc are available in the database. If user entered information will be matched with database information, then transaction will be done securely. And else user transaction will be terminated and transferred to online shopping website.

Flow chart

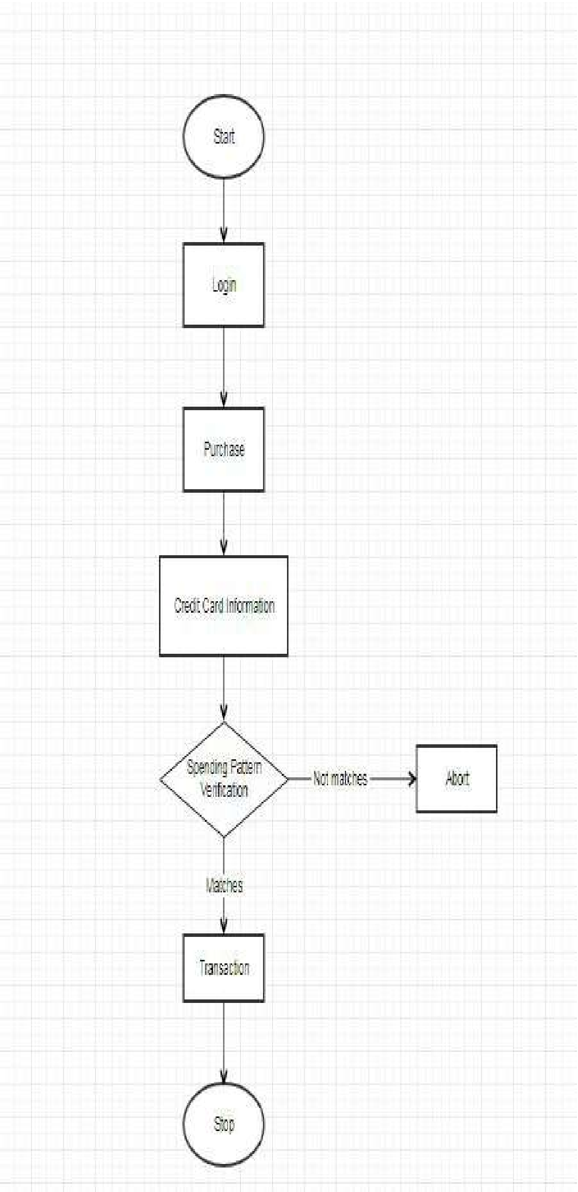


Fig. 3.4 Flow chart of HMM model

Conclusion

Credit card fraudulent detection which is done using HMM (Hidden Markov Model).This technique is used to detect various suspicious activities on credit card.It maintains a database,where past records of transactions are saved and any unusual transaction if carried out, which differs too much from the previous records, it tracks it.Let the user know by sending the details of the transaction on his mobile and hence prevent fraud.

Future Scope

After evaluation of well-known Hidden Markov Model it is clearly shown the various methods which can detect the Fraud efficiently and provide accurate secuirity. Speed of the software can be enhanced by implementation of algorithms of less complexity. Proper security provisions are made from malicious threats so that user account cannot be harmed intentionally or non-intentionally from frauds. Proper hierarchy of the users is maintained as per authority to access the data and use the services provided by the authority. Track all the necessary details during transaction process.

Acknowledgment

No project is ever complete without the guidance of those expert who have already traded this past before and hence become master of it and as a result, our leader. So we would like to take this opportunity to take all those individuals who have helped us in visualizing this project.

We express our deep gratitude to our project guide Prof. Deepti Lawand for providing timely assistance to our query and guidance that she gave owing to her experience in this field for past many year. She had indeed been a lighthouse for us in this journey.

We would also take this opportunity to thank our project co-ordinator Prof. Manjusha Deshmukh for her guidance in selecting this project and also for providing us timely help required for our project work

We are also great full to our HOD Dr. Madhumita Chatterjee for extending her help directly and indirectly through various channel in our project work.

We would like to express our special gratitude to Principal Dr. R.I.K Moorthy for giving invaluable support.

We extend our sincere appreciation to all our Professor from PILLAI COLLEGE OF ENGINEERING for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual.

References

[1] Credit Card Fraud Detection System using Hidden Markov Model and K-Clustering.abhinav srivastava,international journal of advanced research in computer and communication engineering vol. 3, issue 2, february 2014

[2] To secure online payment system using steganography, visual cryptography and hmm.amit r. bramhecha, prof. dinesh d. patil, international journal of innovative research in computer and communication engineering(an iso 3297: 2007 certified organization) vol. 3, issue 9, september 2015

[3]A survey on financial fraud detection methodologies,pankaj richhariya,prashant k singh,international journal of computer applications (0975 – 8887)volume 45– no.22, may 2012

[4]Credit Card Fraud Detection System using Hidden Markov Model, shailesh s. dhok,dr. g. r. bamnote,international journal of advanced research in computer science,volume 3, no. 3, may-june 2012

[5]Pratiksha l. Meshram, Parul Bhanarkar “Credit and ATM Card Fraud Detection using Genetic approach”, ijetae, vol. 1 issue 10, december- 2012

[6].Linda delamaire, hussein abdou, john pointon, “credit card fraud and detection techniques”,banks and bank systems, volume 4, issue 2, 2009

[7]. Krishna kumar tripathi, mahesh a. pavaskar “survey on credit card fraud detection methods”, ijetae, volume 2, issue 11, november 2012.

[8] Syeda, m., zhang, y. q., and pan, y., 2002 parallel granular networks for fast credit card fraud detection, proceedings of ieee international conference on fuzzy systems, pp. 572-577 ,2002.

Automated Question Paper Generator

Rupali Nikhare[1] Rhea Shetty[2] Shivam Singh[3] Shreya Nipanikar[4] Sujitha Sudevan[5]
[1]Professor, [2,3,4,5]Student, Department of Computer Engineering, Pillai College of Engineering,
New Panvel, Maharashtra, India

Abstract - Exams are a vital part of the current education system to test the student's knowledge of the subject. But creating a question paper is a very laborious and time consuming task. There are several factors that the faculty needs to take care of while making a question paper as per the university guidelines. The Automated Question Paper Generator (AQPG) is an intelligent system for simplifying the process of question paper creation. AQPG is a special web application, which stores the question bank related to a particular course and prints a question paper based on its syllabus and curriculum. The system also creates three sets of question papers simultaneously using Fuzzy Logic. AQPG is implemented using Java programming language and MySQL database. The question papers are generated based on the complexity and level of difficulty set for that particular test paper. In AQPG, Bloom's taxonomy is used as a standard measure for setting the difficulty level of questions. Shuffling algorithm is used to avoid repetition of the questions in the question papers. AQPG implements role based hierarchy to assign different access rights to different users, namely the admin, the sub-admin and the teacher. The admin manages all users by adding, updating or deleting sub-admin, teacher, and council members details. The sub-admin has privilege rights of adding, updating and deleting the teachers. Question paper is generated by the teacher and that generated question paper can be viewed by the teacher and the parent sub-admin. For additional security, before printing the generated question paper, an OTP is sent to the user trying to print the question paper to validate that the action is performed by legitimate user. This system also offers choices to select different templates of question paper and if the user is not comfortable with the predefined templates, customization option is also available. The teacher also manages subject as well as question details. Adding questions to the database is allowed only when entered total marks of the question and entered marking scheme distribution of the corresponding question tally. Whenever a new user is added, a confirmation email is sent to the user for the authentication purpose. AQPG implements Intrusion Detection System (IDS) which restricts unauthorized access and an alert is sent in the form of an email to Council Members, including the username of the user attempting to perform alleged illicit activity and the action performed. This enables an educational institute to generate question papers ensuring security and non repetitiveness in the question papers, while reducing human efforts and saving time as well as resources.

Keywords - Bloom's Taxonomy, Fuzzy Logic, Intrusion Detection System, Java, Shuffling Algorithm.

1. INTRODUCTION

1.1 Fundamentals

As manual generation of a balanced question paper by an individual is quite complex, the blending of technology into teaching and learning process is inevitable. Generating an effective question paper is a task of great importance for any educational institute. Hence, with the help of this technical paper we present the solution in form of Automated Question Paper Generator (AQPG).

1.2 Objectives

The objective of Automated Question Paper Generator is as follows:

- To automate the process of generating question papers without any repetition of questions and in doing so ensuring that the question papers are generated quickly and efficiently covering the entire syllabus.
- To generate the question paper as per the difficulty level chosen by the user using Bloom's taxonomy to select appropriate questions.
- To restrict unauthorized access by using Intrusion Detection mechanisms in order to provide security of the generated question papers.
- To create the marking scheme of the answers related to the generated question paper.

1.3 Scope

The Automated Question Paper Generator (AQPG) is a web based application which generates question papers quickly as per the difficulty level set by the user. This application can be used by educational institutions to create subjective examination question papers. AQPG develops three sets of question papers based on the chosen criteria while covering the entire portion. It ensures that questions are not repeated and also provides a marking scheme template for the answers related to the generated question paper. Authorization techniques are used to avoid unauthorized access of the system. In the unlikely event that an intruder were to enter into the system and perform malicious tasks, an alert would be sent to the council members in order to take remedial actions.

2. LITERATURE SURVEY

Rohan Bhirangi and Smita Bhoir, 2016 [4] have proposed 'Automated Question Paper Generation System'. The architecture of the system is as follows: An integrated Question Paper Generation System is needed with improvements in terms of speed, efficiency, controlled access to the resources, randomization of questions and security.

Surbhi Choudhary, Abdul Rais Abdul Waheed, Shrutika Gawandi, and Kavita Joshi, 2015 [5] have proposed a paper on 'Question Paper Generator System'. The working described is as follows: 1) Admin Login 2) Question Insertion 3) Difficulty Choosing 4) Random Paper Generation 5) Wide Chapter Coverage 6) Doc File Creation 7) Emailing 8) PDF Generation

Ashok Immanuel, Tulasi.B, 2015[7] have elaborated the categories of Bloom's taxonomy in the paper 'Framework for Automatic Examination Paper Generation System'. The categories in the cognitive domain of the revised Bloom's taxonomy include Remember, Understand, Apply, Analyse, Evaluate and Create. Bloom's taxonomy emphasises the need to identify the different types of learners based on the varied skill sets.

Kapil Nayak, Shreyas Sule, Shruti Jadhav, Surya Pandey, 2014[10] have proposed the paper for 'Automatic Question Paper Generation System using Randomization Algorithm'. With the help of this paper we present the solution in form of Automatic Question Paper Generator System (QGS) which makes use of shuffling algorithm as a randomization technique. This system includes several modules like user administration, subject selection, difficulty level specification, question entry, question management, paper generation, and paper management.

Suraj Kamya, Madhuri Sachdeva, Navdeep Dhaliwal, Sonit Singh, 2014[9] proposed the paper 'Fuzzy Logic Based Intelligent Question Paper Generator'. AQPG is based on the this paper. It uses a multi-valued membership function to denote membership of an object in a class rather than the classical binary true or false values. Fuzzy set is described by a membership function () that maps a set of objects onto the interval of real numbers between 0 and 1.

Aniruddha Joshi, Prathamesh Kudnekar, Mayur Joshi, Siddhesh Doiphode, 2014[8] published the paper on 'A Survey on Question Paper Generation System'. System Structure and Composition: 1) Structure of test question database 2) Structure of paper database 3) Structure of template database 4) System implementation Question Paper Manipulation. The process of Information extraction consists of following modules. 1) Construction of PDF files parser. 2) Construction of tag injector. 3) The process of tag preprocessor.

Kiran Dhangar, Deepak Kulhare, Arif Khan, 2013[6] published 'A Proposed Intrusion Detection System'. This paper is an intrusion detection system (IDS) proposed by analysing the principle of the intrusion detection system based on host and network.

3. AUTOMATED QUESTION PAPER GENERATOR

3.1 Overview

The Automated Question Paper Generator(AQPG) uses several algorithms in order to provide different features as mentioned in the objective of the system.

3.1.1 Existing System Architecture

The existing system is based on fuzzy logic for autonomous paper generation. In first phase, system requires four users to enter their choices for analytical, descriptive and easy, medium, difficult parts to provide some means for logical division of paper according to marks. In second phase, system provides a fixed skeleton along with various parameters on basis of input from all the four users. The third phase is not accessible by users, it is used at the examination end by authorized person only.

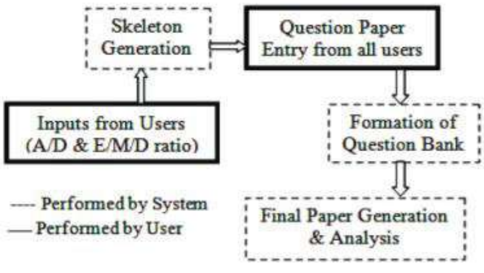

Skeleton Generation
Question Paper Entry from all users
Inputs from Users (A/D & E/M/D ratio)
Formation of Question Bank
Final Paper Generation & Analysis
---- Performed by System
—— Performed by User

Figure 3.1: Overview of Existing System Architecture.[9]

The high level architecture of a Fuzzy Logic based Question Paper Generator system is depicted in Figure 3.1 [9]. The question paper generation process is performed in three steps, each of which is handled by a separate component:

• Skeleton Generation: Based on the user input of the ratio of the difficulty level parameters, a blueprint of the question paper is created.
• Formation of Question Bank: This module collects data from the user in the form of questions and generalizes it to develop a question bank from which questions would be selected in later stages.
• Final Paper Generation and Analysis: This module uses the blueprint and the question bank, that was made in the previous modules, to select questions as per the requirements and generates the final question paper.

3.1.2 Automated Question Paper Generator Architecture

The architecture of AQPG is shown in Figure 3.2.

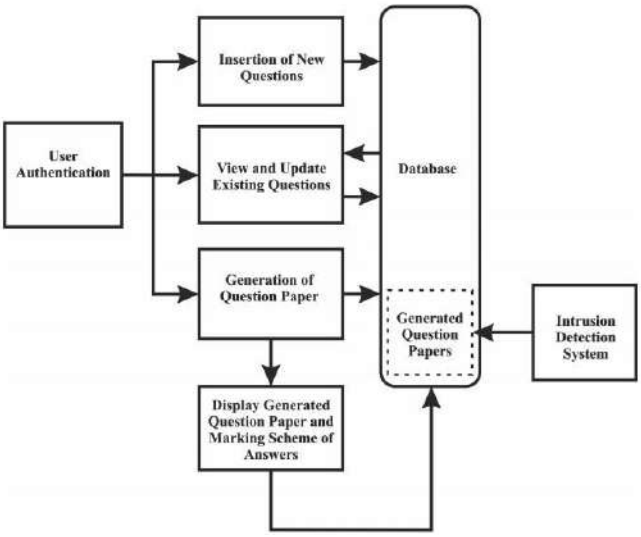

Insertion of New Questions
User Authentication
View and Update Existing Questions
Database
Generation of Question Paper
Generated Question Papers
Intrusion Detection System
Display Generated Question Paper and Marking Scheme of Answers

Figure 3.2: Architecture of AQPG.

The top level view of AQPG architecture shows six major components. The following list describes the various components of AQPG:
• User Authentication: A login module within this module is used to take the user name and password input from the user. Access to the system is granted only when correct user name and password is entered. IDS mechanism is used to alert the council members via email when wrong password is entered more than three times.
• Managing Users: AQPG implements role based hierarchy to assign different access rights to the users, which in turn provides security. The system contains three types of users as follows:

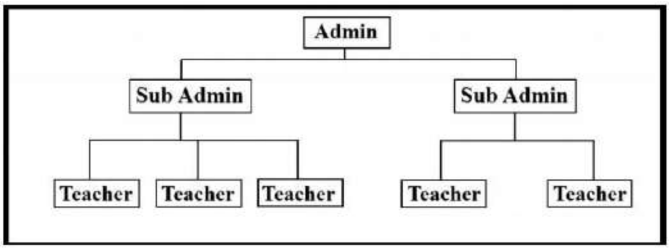


Figure 3.3: Role Based Hierarchy of Users.

1. Admin: The admin is the highest level of authority. The admin can add sub admins, teachers, and council members. The admin can also update details of these users or delete the entry of any user from the database.
2. Sub Admin: The sub admins are the next of kin of the admin. The sub admin can also view and print the question papers generated by the teachers that fall under the hierarchy of that sub admin.

3. Teacher: Teachers are the root nodes of the hierarchy tree. Teachers can add a new subject, update or delete the details of the subject added by that teacher. Similarly, teachers can add new questions by specifying all the required inputs such as the difficulty level of the question.

4. Council: The council is a response team responsible for executing corrective measures in the unlikely event of intrusion in the system. The details along with email addresses of the council members are stored in the database by the admin.

- Insertion of New Questions: This module facilitates the addition of new questions in the database, in case the syllabus is updated or the user wishes to add certain new questions from the syllabus. The difficulty level that is to be assigned to the new question is also taken as input from the user while inserting it into the database.
- Search and View Existing Questions: This module uses keyword extraction to fetch questions in the database which match the keyword given as input by the user. A list of all questions having the keyword is given as output by this module.
- Generation of Question Paper: As per the figure above, the steps involved in generating the question paper are as follows:

1. Selection of Template and/or Customization: The question paper generation process begins with the user choosing a predefined template of the question or opting for customization of the format of the question paper as per their requirements.
2. Input Difficulty Level: Once the framework of the question paper is ready, the user is prompted to set the difficulty level parameters for the questions as level as other parameters required for the question paper.
3. Skeleton Generation: With the blueprint and the computed values from the previous modules, a skeleton of the question paper is developed. It contains the number of questions required in the correct format as per the chosen template and the complexity of the questions to be picked from the database.
4. Assortment of Required Number of Questions: This module selects questions from the question bank using Bloom's taxonomy to check the required difficulty level. Randomization and Shuffling algorithms are used to pick questions from the entire syllabus.
5. Final Question Paper Generation: Once the questions have been selected, the final question paper is generated in the proper format with all the selected questions.

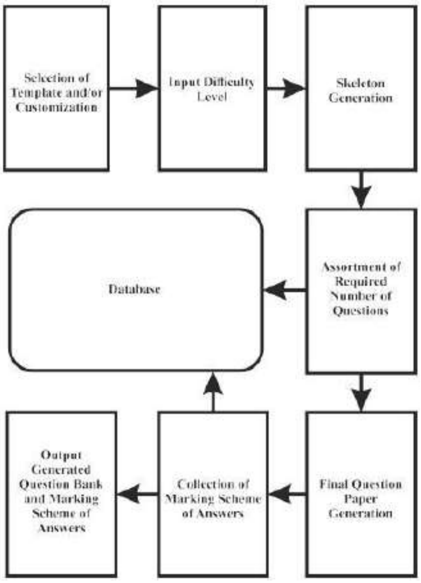

Selection of Template and/or Customization
Input Difficulty Level
Skeleton Generation
Database
Assortment of Required Number of Questions
Output Generated Question Bank and Marking Scheme of Answers
Collection of Marking Scheme of Answers
Final Question Paper Generation

Figure 3.4: Generation of Question Paper Module Internal Architecture.

6. Collection of Marking Scheme of Answers: On selection of the questions, the distribution of marks associated with the questions is fetched in the database. A marking scheme of the answers to the questions is then created.
7. Output Generated Question Bank and Marking Scheme of Answers: This marks the end of the Generation of Question Paper module and the question paper generation process is completed. The generated question paper and the associated marking scheme is given as input to the next module.

- Display Generated Question Paper and Marking Scheme of Answers: This takes the output of the previous module and presents it to the user. It provides options for discarding the generated question paper, and saving the question paper. On receiving the save command, this module saves the generated question paper in the database as a PDF file. The PDF file is encrypted using AES cryptographic hashing algorithm to ensure security.
- Database: The database consists of questions spanning over the entire syllabus of each course of the Computer Engineering department of Engineering. It also stores the respective marking scheme of answers related to each question. Once the question paper has been generated, it is saved in the database.

3.2 Flowcharts and Activity Diagram

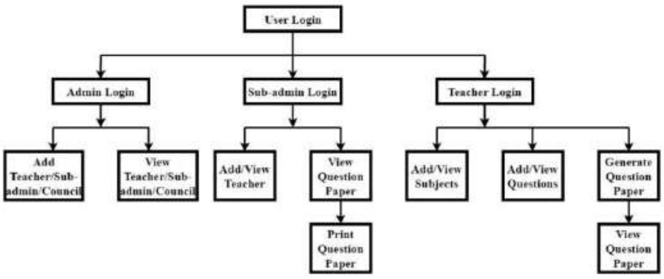


Figure 3.5: Overview of AQPG.

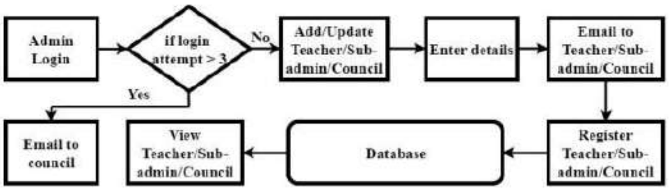

Admin Login
if login attempt > 3
No
Yes
Add/Update Teacher/Sub-admin/Council
Enter details
Email to Teacher/Sub-admin/Council
Email to council
View Teacher/Sub-admin/Council
Database
Register Teacher/Sub-admin/Council

Figure 3.6: Process of Adding a User.

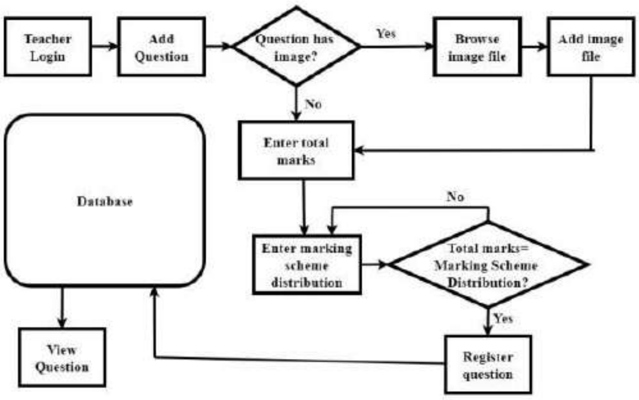


Figure 3.7: Process of Adding a Question.

3.3 Implementation Details

3.3.1 Algorithms/Techniques

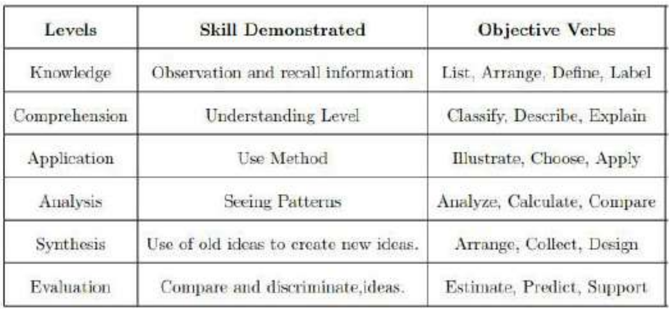
1. Bloom's Taxonomy: One of the most important aims in post primary education is the attainment of critical or higher-order thinking skills. Identifying how to encourage, teach and then assess these skills is important. Bloom's taxonomy is a classification system of educational objectives based on the level of student understanding necessary for achievement or mastery. Useful applications of the taxonomy include formulating questions to challenge your students in class tests, during class time and for homework assignments. A taxonomy is used to classify things. This taxonomy defines levels of objectives in 3 domains:

- Cognitive (knowledge based)
- Affective (emotive based)
- Psychomotor (action based)

The Cognitive Domain

This domain is mostly used. The objectives dealt with in the Cognitive domain place an emphasis on remembering or recalling information. Cognitive objectives vary from simple recall of material that was learned to highly original and creative ways of combining and synthesizing new ideas. The taxonomy is divided into six levels: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Bloom's Taxonomy is hierarchical; meaning that learning at the higher levels is dependent on having attained prerequisite knowledge and skills at the lower levels.[1]

There are "verb tables" to help identify which action verbs align with each level in Bloom's Taxonomy. Some of these verbs on the table are associated with multiple Bloom's Taxonomy levels. These "multilevel-verbs" are actions that could apply to different activities.



Levels	Skill Demonstrated	Objective Verbs
Knowledge	Observation and recall information	List, Arrange, Define, Label
Comprehension	Understanding Level	Classify, Describe, Explain
Application	Use Method	Illustrate, Choose, Apply
Analysis	Seeing Patterns	Analyze, Calculate, Compare
Synthesis	Use of old ideas to create new ideas.	Arrange, Collect, Design
Evaluation	Compare and discriminate,ideas.	Estimate, Predict, Support

Table 3.1: Verb table

2. Fuzzy Logic: The algorithm is as follows:

(a) A question paper can be categorized in two ways: Content of the paper (subcategory: Analytical(A),Descriptive(D)) and Difficulty Level of the paper (subcategory: Easy(E), Medium (M), Difficult(D)).

(b) Both the analytical and descriptive questions can be of any difficulty level, so both A/D and E/M/D parameters are considered as independent of each other.

(c) Users are allowed to choose any value for Analytical and Descriptive, both in range of 0-10 (10 being the highest value and 0 being the lowest value for analytical and vice-versa for descriptive). Users may also choose floating point numbers for A and D, irrespective of each other (Sum may or may not be 10).

(d) For E, M and D, user can give only integers values such that sum of all the three parameters must be 10, satisfying the following criteria: $1 \leq E \leq 5$, $4 \leq M \leq 10$ and $1 \leq D \leq 5$. [9].

Based on the input, using Fuzzy logic, calculations are performed as follows:

(a) Consider A as high and D as very high; find out the points having membership value, $\mu = 1$.

(b) In the set of high for A the candidates having $\mu = 1$ are 6, 7 (considering only integers) and for D, in the set of very high such candidates are 0, 1 and 2.

(c) Find out the average of every possible combination of these value s (6+0/2=3, 6+1/2=3.5, 6+2/2=4, 7+0/2=3.5 , 7+1/2=4, 7+2/2=4.5).

(d) Then these averages are take n in descending order of frequencies and are mapped on the output membership function.

(e) By tracing these value s on membership function for output 1 of FSK-K, find out which points carry maximum value of μ collectively (frequency of 3.5 and 4 is 2, sum of individual is 2 for both and they belong to same group in output; for A out it is high and D out it is low and can be confirmed from rule no.5).

Some of the rules are given below:

(a) If Analytical is very low and Descriptive is very low then Analytical Out is medium, Descriptive out is medium.

(b) If Analytical is low and Descriptive is very low then Analytical Out is high, Descriptive out is Low.

(c) If Analytical is medium and Descriptive is very low then Analytical Out is high, Descriptive out is Low.

(d) If Analytical is high and Descriptive is very low then Analytical Out is high, Descriptive out is Low.

(e) If Analytical is high and Descriptive is very high then Analytical Out is high, Descriptive out is Low. E.g. final values obtained from Fuzzy logics are A=7, D=3 and E=3, M=5, D=2.[9] This value is then used to create a framework of the question paper.

3. Randomization and Shuffling Algorithm: The main role of the shuffling algorithm is to provide randomization technique in AQPS thus different sets of question could be generated. A randomized algorithm is an algorithm that employs a degree of randomness as part of its logic.[11] An algorithm that uses random numbers to decide what to do next anywhere in its logic is called Randomized Algorithm. Both the algorithms work in combination as follows:

The number of questions that are required as per the template is stored in a variable, say w.The system randomly selects a question of a module from the database by comparing the difficulty level calculated using fuzzy calculations and the difficulty level assigned to the question using Bloom's taxonomy.The chosen question and the module are then locked. The system chooses the next question from another module, this ensures non-repetitiveness of questions.If number of questions, w, is not equal to zero but the number of module of the particular subject are all locked, then all modules are opened but the selected questions remain locked. This process is repeated until w becomes equal to zero.[9]

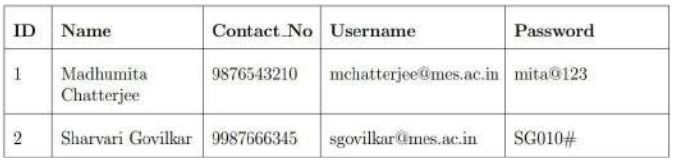
This process is repeated for three iterations to generate three sets of question papers. A minimum of 40 questions for each question having different weightage of marks of a subject is required for the algorithm to successfully fetch required amount of questions for the three sets of question papers. On an average, if there are 5, 10 and 20 mark questions for a 80 mark question paper, then 40 questions of 5 marks each, 40 questions of 10 marks each and 40 questions of 20 marks each are required. Therefore, $40 * 3 = 120$ minimum questions are required to generate a 80 question paper of a subject.

4. Intrusion Detection System: An intrusion detection system (IDS) is a device or software application that monitors a network or systems for malicious activity or policy violations.[3] Intrusion detection mechanism has been deployed on AQPG to send alerts to council members in case of suspicious activities. Alerts are sent in the form of emails.

Simple Mail Transfer Protocol (SMTP) has been used to send emails to council members. The activities which trigger the IDS to send emails are:
(a) Multiple Failed Login Attempts by Any User
(b) Any User Accessing a Generated Paper before the Set Time
The emails contain the username of the user and the type of incident occurred. Additionally, emails are also sent to new users upon registration as a confirmation with the username and password in the email.

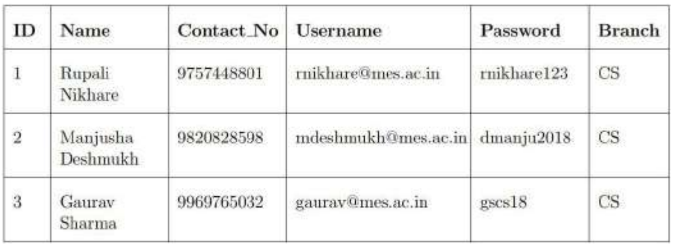
3.3.2 Sample Dataset for Experiment

User Details Tables: The primary key of each table is auto-incremented. User contacts such as name, username, contact number, password, etc are stored as per the requirements.



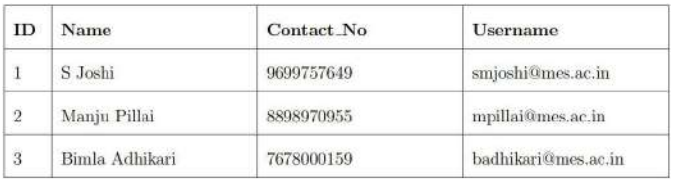
ID	Name	Contact_No	Username	Password
1	Madhumita Chatterjee	9876543210	mchatterjee@mes.ac.in	mita@123
2	Sharvari Govilkar	9987666345	sgovilkar@mes.ac.in	SG010#

Table 3.2: Sub-Admin Details Table



ID	Name	Contact_No	Username	Password	Branch
1	Rupali Nikhare	9757448801	rnikhare@mes.ac.in	rnikhare123	CS
2	Manjusha Deshmukh	9820828598	mdeshmukh@mes.ac.in	dmanju2018	CS
3	Gaurav Sharma	9969765032	gaurav@mes.ac.in	gscs18	CS

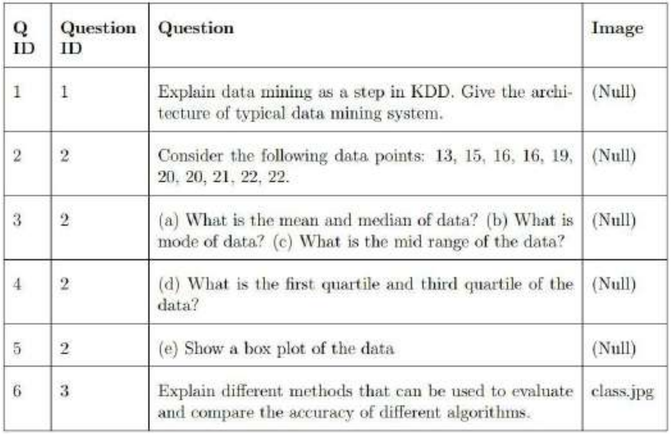
Table 3.3: Teacher Details Table



ID	Name	Contact_No	Username
1	S Joshi	9699757649	smjoshi@mes.ac.in
2	Manju Pillai	8898970955	mpillai@mes.ac.in
3	Bimla Adhikari	7678000159	badhikari@mes.ac.in

Table 3.4: Council Details Table

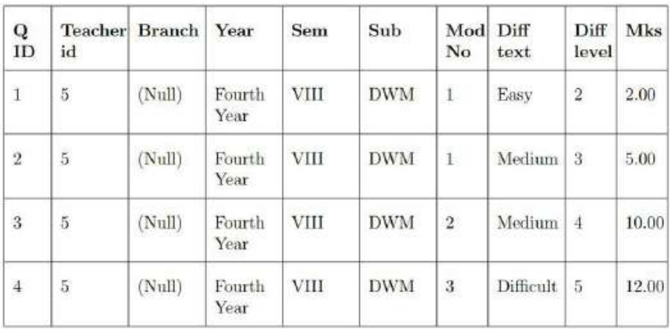
Questions Details Table: q id is the primary key of the table. question id is used to identify sub-parts of a question (if any). The question and images associated with the question are in the subsequent columns.



Q ID	Question ID	Question	Image
1	1	Explain data mining as a step in KDD. Give the architecture of typical data mining system.	(Null)
2	2	Consider the following data points: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22.	(Null)
3	2	(a) What is the mean and median of data? (b) What is mode of data? (c) What is the mid range of the data?	(Null)
4	2	(d) What is the first quartile and third quartile of the data?	(Null)
5	2	(e) Show a box plot of the data	(Null)
6	3	Explain different methods that can be used to evaluate and compare the accuracy of different algorithms.	class.jpg

Table 3.5: Question Details Table

The details associated with the questions such as subject, marks, difficulty level, etc are mapped in the table below:

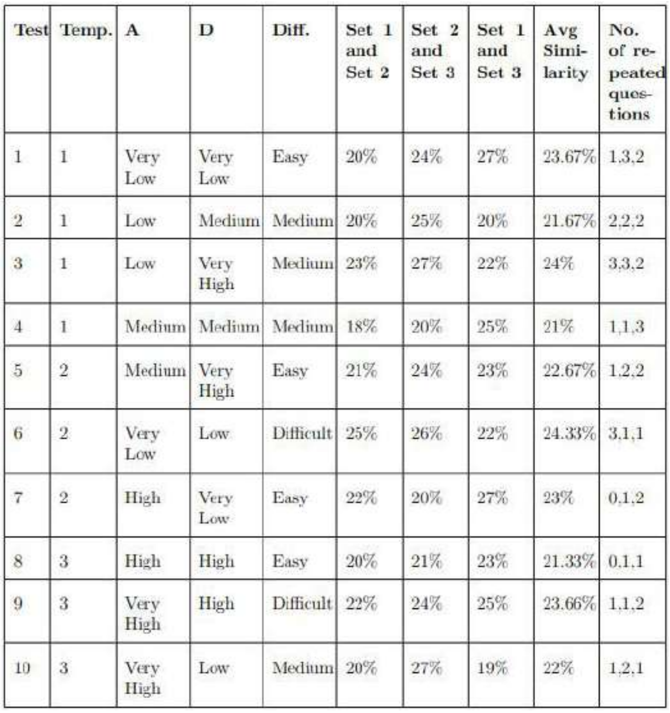


Q ID	Teacher id	Branch	Year	Sem	Sub	Mod No	Diff text	Diff level	Mks
1	5	(Null)	Fourth Year	VIII	DWM	1	Easy	2	2.00
2	5	(Null)	Fourth Year	VIII	DWM	1	Medium	3	5.00
3	5	(Null)	Fourth Year	VIII	DWM	2	Medium	4	10.00
4	5	(Null)	Fourth Year	VIII	DWM	3	Difficult	5	12.00

Table 3.6: Question Parameters Table

3.3.3 Performance Evaluation Metrics

We have tested the paper generation for various inputs, that is different templates and different difficulty levels. The experiment was carried out using a small dataset of 145 questions of a subject. Three sets of a 80 marks question paper was generated 10 times using different inputs. Based on our experimental analysis, we get the following results.



Test	Temp.	A	D	Diff.	Set 1 and Set 2	Set 2 and Set 3	Set 1 and Set 3	Avg Similarity	No. of repeated questions
1	1	Very Low	Very Low	Easy	20%	24%	27%	23.67%	1,3,2
2	1	Low	Medium	Medium	20%	25%	20%	21.67%	2,2,2
3	1	Low	Very High	Medium	23%	27%	22%	24%	3,3,2
4	1	Medium	Medium	Medium	18%	20%	25%	21%	1,1,3
5	2	Medium	Very High	Easy	21%	24%	23%	22.67%	1,2,2
6	2	Very Low	Low	Difficult	25%	26%	22%	24.33%	3,1,1
7	2	High	Very Low	Easy	22%	20%	27%	23%	0,1,2
8	3	High	High	Easy	20%	21%	23%	21.33%	0,1,1
9	3	Very High	High	Difficult	22%	24%	25%	23.66%	1,1,2
10	3	Very High	Low	Medium	20%	27%	19%	22%	1,2,1

Table 3.7: Performance Evaluation of AQPG

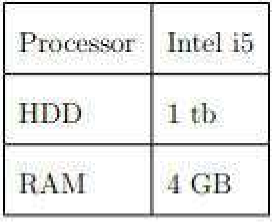
Here,
Column 1: Test Case Number
Column 2: Template Number
Column 3: Analytical Level
Column 4: Descriptive Level
Column 5: Difficulty Level
Column 6: Similarity between Set 1 and Set 2
Column 7: Similarity between Set 2 and Set 3
Column 8: Similarity between Set 1 and Set 3
Column 9: Average Similarity
Column 10: Number of Repeated Questions

The similarity quotient includes the header of the question paper, containing the details of the question paper, as well. Additionally, sub-stringing matching was also performed while testing the similarity between the different sets of generated question papers.

The average percentage of similarity between the different sets of generated question papers is approximately 22.73%. Whereas,the average number of questions repeated between the different sets of generated question papers based on all observations is 1.6, which can be rounded off to 2 questions.

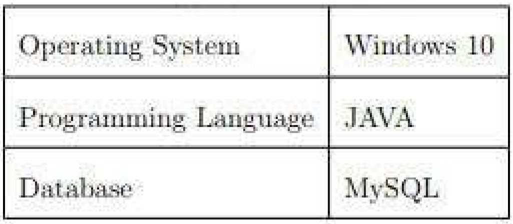
3.3.4 Hardware and Software specifications

The experiment setup is carried out on a computer system which has the different hardware and software specifications as given in Table 3.8 and Table 3.9 respectively



Processor	Intel i5
HDD	1 tb
RAM	4 GB

Table 3.8: Hardware details



Operating System	Windows 10
Programming Language	JAVA
Database	MySQL

Table 3.9: Software details

4. APPLICATIONS

The Automated Question Paper Generator(AQPG) can be widely used in educational institutions to develop subjective exam question papers without going through the hassle of the manual process. The application generates question papers efficiently by selecting questions from the entire portion of the course. The generated question papers need not be physically transported to their destination as the system provides facilities to access the question papers within the system to the desired recipients. The security mechanisms that is implemented in the system makes it secure by taking preventive as well as corrective measures to avoid leakage of question papers. Thus, the system excludes human efforts and saves time and resources.

5. CONCLUSION AND FUTURE SCOPE

In this report, the description of Automated Question Paper Generator is presented. The different algorithms such as Fuzzy Logic, Randomization, Shuffling algorithm, Intrusion Detection System, etc are explained with examples. The authentication technique is also described. The comparative study of various techniques used in existing systems is presented in this report. The AQPG has several modifications and combinations of features of the existing systems. The different standard datasets or variable inputs are defined that are be used for question paper generation. The datasets identified for experiments are user details and question details. The applications of this system are identified and presented.

With a few improvements, AQPG could also be used for objective exam question papers. AQPG is currently useful for generating question papers of pen and paper based exams, but it could be used to facilitate online tests too. Using the questions in the database of AQPG, students could be given practice papers before exams, containing questions of varying difficulty levels and the associated marks with each question. Upon updating the database with the answers, AQPG will be able to provide ideal answer key templates of the generated question papers. This would be helpful for both, students and teachers, alike. Teachers would have a guide while checking answer papers, whereas students would understand which points should be included in answers and subsequently help them write better answers for the given questions. A facility to schedule generation of question papers for weekly or monthly tests could also be introduced in the AQPG.

REFERENCES

[1] Bloom’s taxonomy. https://tips.uark.edu/using-blooms-taxonomy/. Accessed: 2018-03-15.

[2] Fuzzy logic systems. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_fuzzy_logic_systems.htm. Accessed: 2018-03-18.

[3] Intrusion detection system. https://security.stackexchange.com/questions/158893/question-about-ids-and-ips/158944#158944. Accessed: 2018-03-30.

[4] Rohan Bhirangi and Smita Bhoir. Automated question paper generation system. IJERMT, 2016.

[5] Surabh Chaudhary, Abdul Rais Abdul Waheed, Shrutika Gawandi, and Kavita Joshi. Question paper generator system. IJCST,, 2015.

[6] Kiran Dhangar, Deepak Kulhare, and Arif Khan. A proposed intrusion detection system. IJCA, 2013.

[7] Ashok Immanuel and Tulasi.B. Framework for automatic examination paper generation system. IJCST, 2015.

[8] Aniruddha Joshi, Prathamesh Kudnekar, Mayuri Joshi, and Siddhesh Doiphode. A survey on question paper generation system. IJCA, 2014.

[9] Suraj Kamya, Madhuri Sachdeva, Navdeep Dhaliwal, and Sonit Singh. Fuzzy logic based intelligent question paper generator. IEEE, 2014.

[10] Kapil Nayak, Shreyas Sule, Shruti Jadhav, and Surya Pandey. Automatic question paper generation system using randomization algorithm. IJETR, 2014.

[11] Manish Varshney Prabhakar Gupta, Vineet Agarwal. Design and Analysis of Algorithms. PHI Learning Pvt. Ltd., 2012.

ACKNOWLEDGEMENT

We remain immensely obliged to our project guide Prof. Rupali Nikhare, for her valuable guidance, patience, keen interest and constant encouragement and for her invaluable support.

We would like to thank Dr. Madhumita Chatterjee, Head, Department of Computer Engineering for her invaluable support.

We would also like to thank Dr. Sandeep M. Joshi, Principal for his invaluable support and for providing an outstanding academic environment.

We would also like to thank all the staff members of the department of Computer and Information Technology for their critical advice and guidance without which this project would not have been possible.

Last but not the least we would also like to acknowledge with much appreciation the crucial role of our family members especially our friends who have been a constant source of inspiration during this project work. The completion of this project would not have been possible without them.

We would like to say that it has indeed been a fulfilling experience working on this project.

Human Machine Interface for controlling a robot using image processing

Supervisor Prof.Rupali Nikhare

Abhijeet Patil
Computer Department
Pillai's College of Engineering
Mumbai University , India
abhijeetpatil014@gmail.com

Mangesh Nikam
Computer Department
Pillai's College of Engineering
Mumbai University , India
mangeshnikam71@gmail.com

Rohan Patil
Computer Department
Pillai's College of Engineering
Mumbai University , India
rohanpatil9106@gmail.com

Omprakash Pandit
Computer Department
Pillai's College of Engineering
Mumbai University , India
omprakashpandit99@gmail.com

***Abstract*—With the view of improvising human control over robots and other automated machines, a number of techniques have been devised. The aim is to make these machines more and more human friendly. It involves communicating with the robot through the users eye by making it follow the eye movement. In this paper a self- made Video Camera based Gaze tracking system has been discussed, whose output can be used to control an in house robot via Arduino Uno micro controller. The process involves image acquisition using a USB web cam mounted on the user's PC at a fixed position. The image frames obtained from video in real time undergo processing in MATLAB to provide necessary information regarding user's point of gaze. This information can then be used to control the movement of the robot.**

***Index Terms*—Real time pupil Tracking system, 2 wheel Robot, MATLAB image processing, Face detection, Iris/pupil center calculation, MATLAB-Arduino interfacing.**

I. INTRODUCTION

Eye Gaze tracking is a technique in which the eye movements of a person are recorded continuously so that the computer knows where a person is looking at any given time as well as the sequence in which their eyes are moving from one location to another. Eye movements may also be recorded and used in the form of control signals to enable people to interact with robots or other automated devices directly without the need for mouse or keyboard input. This can be a major advantage for certain users such as disabled people with non functional limbs or paralysis. The idea behind using eye movement as a control mechanism comes from the fact that eyes are the most extensively used sense organs. Even in disabled or paralyzed people, the eyes are mostly functional and can be effectively used to control devices such as a wheelchair. It is a human tendency as well as reflex to first look at the object of interest. Thus making use of this tendency directly can reduce the time required to convey the same to a robot. This system can not only be used to develop a robotic assistant for the disabled people with fully functional eyes (such as eye controlled wheelchair,a robot that gets water for the patient when indicated, an eye controlled television, etc.),its application can also be extended to industries where gaze tracking can be used for the development of a multi modal Human-Robot interface.

II. LITERATURE REVIEW

We got the idea of our project by available existing systems based on detection and tracking of Eye to give the directions to the Robot. We have gone through several papers together information about various techniques for Face analysis, Eye extraction,Pupil detection and Robot signals. In this chapter the relevant techniques in literature is reviewed. It describes various techniques used in the work. Identifying the current literature on related domain problem and Identifying the techniques that have been developed and present the various advantages and limitation of these methods used extensively in literature.

There are many different approaches for implementing eye detection and tracking systems. Many eye tracking methods were presented in the literature. However, the research is still on-going to find robust eye detection and tracking methods to be used in a wide range of applications.

A. Image Processing

1) Face Recognition:: Face recognition presents a challenging problem in the field of image analysis and computer vision. The security of information is becoming very significant and difficult. Face identification system is used in security. Face recognition system should be able to automatically detect a face in an image.

- In this paper, we are reading various facial images and storing them. Image test benches are read in our Verilog program and stored in memories. We compare images bit by bit and check if there is any mismatch. If image is matched then we display 'Match found' otherwise 'No match found'. In further study, we are obtaining special features of the face such as Lip portion or Eye portion. We subtract test images with stored images and compare the subtracted value with threshold limit for detection.

2) Computer-Vision-Based Eye Tracking: Most eye tracking methods presented in the literature use computer vision based techniques. In these methods, a camera is set to focus on one or both eyes and record the eye movement. The main focus of this paper is on computer vision based eye detection and gaze tracking. There are two main areas investigated in the field of computer vision based eye tracking. The first area considered is eye detection in the image, also known as eye localization. The second area is eye tracking, which is the process of eye gaze direction estimation. Based on the data obtained from processing and analyzing the detected eye region, the direction of eye gaze can be estimated then it is either used directly in the application or tracked over subsequent video frames in the case of real-time eye tracking systems.

- This paper proposes an eye state detection system using Haar Cascade Classifier and Circular Hough Transform. Our proposed system first detects the face and then the eyes using Haar Cascade Classifiers, which differentiate between opened and closed eyes. Circular Hough Transform (CHT) is used to detect the circular shape of the eye and make sure that the eye is detected correctly by the classifiers. The accuracy of the eye detection is 98.56 percent on our database which contains 2856 images for opened eye and 2384 images for closed eye. The system works on several stages and is fully automatic. The eye state detection system was tested by several people, and the accuracy of the proposed system is 96.96 percent.

3) Arduino: Arduino Uno is a micro controller board based on 8-bit ATmega328P micro controller. Along with ATmega328P, it consists other components such as crystal oscillator, serial communication, voltage regulator, etc. to support the micro controller. Arduino Uno has 14 digital input/output pins (out of which 6 can be used as PWM outputs), 6 analog input pins, a USB connection, A Power barrel jack, an ICSP header and a reset button.

- In this paper they developed a pupil direction observing system for anti-spoofing in face recognition systems using a basic hardware equipment. Firstly, eye area is being extracted from real time camera by using Haar-Cascade Classifier with specially trained classifier for eye region detection. Feature points have extracted and traced for minimizing person's head movements and getting stable eye region by using Kanade-Lucas-Tomasi (KLT) algorithm.
- After a few stable number of frames that has pupils, proposed spoofing algorithm selects a random direction and sends a signal to Arduino to activate that selected direction's LED on a square frame that has totally eight LED's for each direction. After chosen LED has been activated, eye direction is observed whether pupil direction and LED's position matches.

III. OVERVIEW

The system overview gives a brief description about the overall working of the system. Here, the user interacts with the system through voice input. The further processing is done as follows:

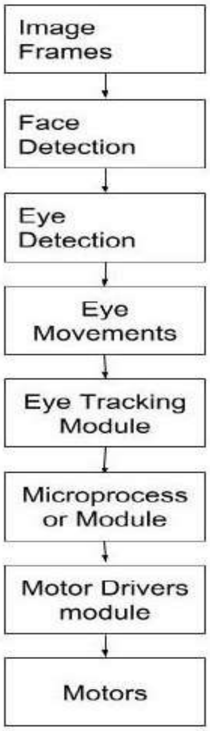


Fig. 1. Overview of the proposed system

- Eye movements and Camera module: Eye Gaze tracking is a technique in which the eye movements of a person are recorded continuously so that the computer knows where a person is looking at any given time as well as the sequence in which their eyes are moving from one location to another.
- Eye tracking module : The system (machine) identifies the orientation of the face movement with respect to the pixel values of image in a certain areas.eye area is being extracted from real time camera by using Haar-Cascade classifier with specially trained classifier for eye region detection and getting stable eye region by using Kanade-Lucas-Tomasi (KLT) algorithm.
- Arduino UNO and motors module: The gaze tracking system output can be used to control an in house Robot using Arduino UNO to make a move as pe the eye movements. Averages are useful for an overall sense of what the population feels. However these averages lack context during recommendations.

IV. CHARACTERISTICS OF THE SYSTEM DESIGN

A. Software Design

The tracking system is developed using c programming language along with the embedded c. Various libraries has been used in order to design the system. The working of the system has been discussed in the overview section.

B. Design and Implementation

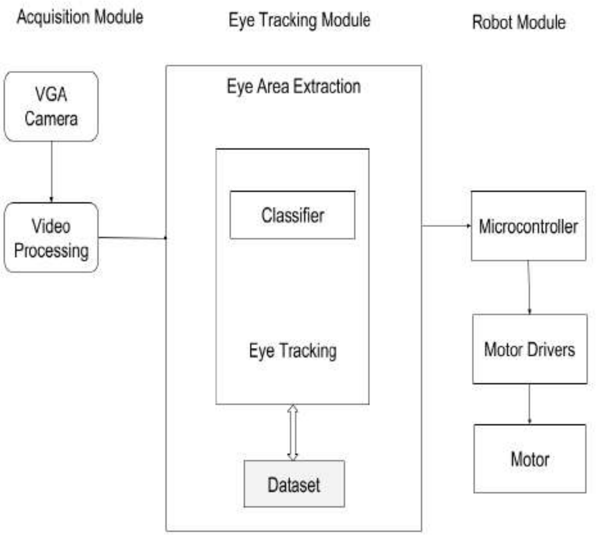


Fig. 2. Overall design of the system

1) **Video acquisition::** Video Acquisition is defined as the process of collecting visual information using a video camera by converting the analog video signals into digital form. It is a combination of video capturing, analog to digital conversion, encod- ing and color space conversion to generate data in any of the several color spaces available such as RGB, YCbCr, etc. as per requirement.

The real time video was captured using USB webcam (’VGA Webcam’). A video object obj was created to store the captured video using resolution ’640x480’ for high quality image. This configuration works at an average frame rate of 10 frames per second.

2) **Eye tracking module::** The image frames obtained from video in real time undergo processing in MATLAB to provide necessary information regarding user?s point of gaze. image. From this eye region, iris and pupil are located and their centers are calculated in real time to be stored in an array.

The Viola-Jones algorithm was implemented for face detection in the image. This algorithm not only detects the facial region in an image, but is also capable of finding the eye region accurately as it is a feature based detection algorithm. Thus it was used to locate both the user’s face as well as eye region in the video frames that were extracted continuously.

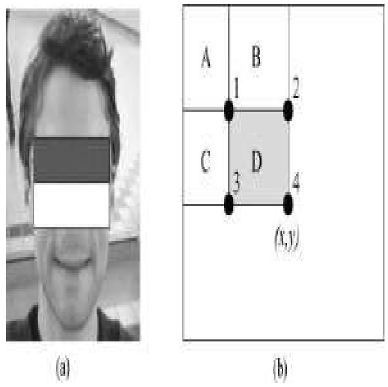


Fig. 3. Overall design of the system

Thus it was used to locate both the user?s face as well as eye region in the video frames that were extracted continuously. This was followed by putting a Bounding box around the face region and measuring the enclosed area. This area then finds the biggest eye between both to display the eye in a separate image. The image was simultaneously converted into a gray scale image.

- Edge Detection: In order to detect the iris region from the eye image, thresholding and edge detection was performed. Various edge detection techniques were applied on the image, of which the Haar Cascade Detection technique provided the best, most extensive results and was thus selected for the project. The threshold value depends upon the light intensity of the room as well as the image quality from the given camera.
- Hough Transform for Iris Center and Radius Calculation: After thresholding and edge detection, Circular Hough Transform was applied to the binary image to detect the dark circular region in the image and to calculate its center and radius. The radius range for search was defined to lie between 9 to 10 pixels. This Hough transform is highly optimized. It uses the midpoint circle algorithm to draw the circles in voting space quickly and without gaps. It also includes an option for searching only part of the image to increase speed if a rough estimate of the circle locations is known. Function imfindcircles uses a Circular Hough Transform (CHT) based algorithm for finding circles in images. This approach is used because of its robustness in the presence of noise, occlusion and varying illumination. The CHT is not a rigorously specified algorithm, rather there are a number of different approaches that can be taken in its implementation. However, by and large, there are three essential steps which are common to all.

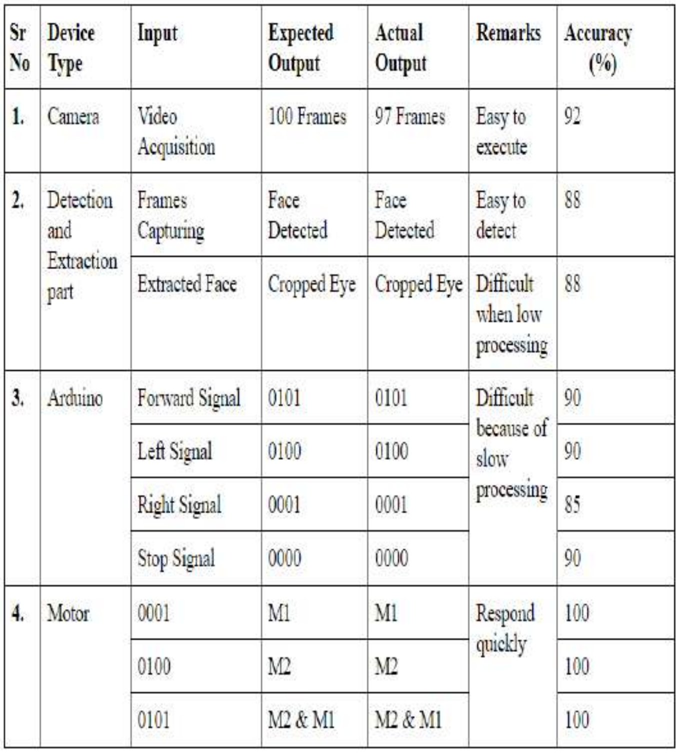
3) **Robot module::** This information can then be used to control the movement of the robot. Serial communication between matlab and 3R robot, 4 signals are going to generate based on the eye movements. After Arduino interfacing, this signals was used to send to the robot through Arduino controller. The DC motor were connected at the pins 10, 11, 12 and 13 of the Arduino board through the motor driver L293D. Power was supplied to the robot?s actuators (motors) through an 9V to 12V battery.

TABLE I
ARDUINO INPUTS AND OUTPUTS FOR MOTORS

Commands from Matlab	Arduino output pins	L239D output Pins
Straight	LHLH	LHLH
Left	LHLL	LHLL
Right	LLLH	LLLH

V. TESTING AND RESULT

We have tested this project in standard conditions and checked result for certain test cases those are shown in following table. How Man times this devices give correct outputs and how many times it doesn't.



Sr No	Device Type	Input	Expected Output	Actual Output	Remarks	Accuracy (%)
1.	Camera	Video Acquisition	100 Frames	97 Frames	Easy to execute	92
2.	Detection and Extraction part	Frames Capturing	Face Detected	Face Detected	Easy to detect	88
		Extracted Face	Cropped Eye	Cropped Eye	Difficult when low processing	88
3.	Arduino	Forward Signal	0101	0101	Difficult because of slow processing	90
		Left Signal	0100	0100		90
		Right Signal	0001	0001		85
		Stop Signal	0000	0000		90
4.	Motor	0001	M1	M1	Respond quickly	100
		0100	M2	M2		100
		0101	M2 & M1	M2 & M1		100

Fig. 4. Test Cases and Results

The above gives the performance of the system based on the types of devices specifying the total number of devices tested along with how many of them were correctly executed along with how many were wrongly executed.

The process involves image acquisition using a USB web cam mounted on the user's PC at a fixed position. The image frames obtained from video in real time undergo processing in MATLAB to provide necessary information regarding user's point of gaze. This information can then be used to control the movement of the robot.

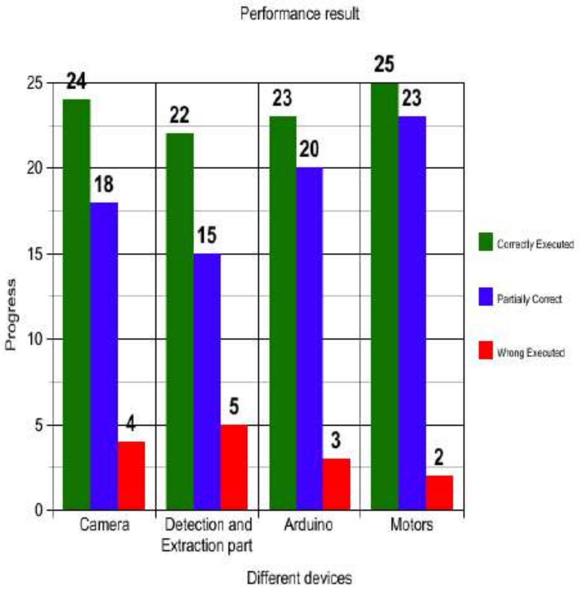


Fig. 5. Performance Evaluation based on types of devices

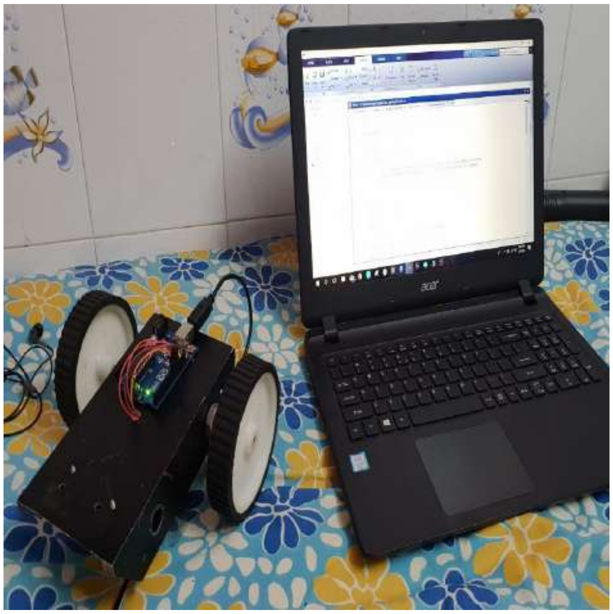


Fig. 6. Human based robot control

VI. FUTURE SCOPE

Currently this project is working with serial communication so later we'll try to implement for wireless communication which give more efficiency and less messiness to operate the wheelchair in better and fast way. Object detection and avoidance can be implemented in this existing system because while moving from one place to another if an object is come in the path of traveling then the wheelchair can detect the object that comes in between and it will make the alarm on which will alert the person that an object is detected.

A. CONCLUSION

In this report, the study of Image Processing techniques is presented. The different techniques such as Techniques for Eye Detection, Eye Tracking, Iris tracking, and Signal generation for robot is explained with examples.The comparative study of various techniques mentioned above is presented in this report. The hybrid approach is proposed with Eye tracking modification. The performance measures like detection and tracking are described in above chapters. The different variable inputs are denied that may be used in experiment for this domain systems. The applications of this domain is presented in the above chapter.

REFERENCES

[1] M. Taskiran, N. Kahraman, “Anti-Spoofing In Face Recognition with Liveness Detection Using Pupil Tracking,”IEEE 15th International Symposium on Applied Machine Intelligence and Information, January 26-28, 2017 Herlany, Slovakia

[2] Pankaj S Lengare and Milind E Rane,”Human hand tracking using MATLAB to control Arduino based robotic arm”,2015 International Conference on Pervasive Computing (ICPC).

[3] Norma Latif Fitriyani and Muhammad Syafrudin, “Real-Time Eye State Detection System Using Haar Cascade Classifier and Circular Hough Transform,”2015 Online International Confernece on Green Engineering and Technologies (IC-GET2015), 2015.

[4] Ambuj K Gautam, V Vasu, USN Raju, “Human Machine Interface for controlling a robot using image processing,” *M.Tech, MED, National Institute of Technology, Warangal, AP 506004 INDIA 2014.

[5] Lai Wei and Huosheng Hu, Senior Member, “A multi-modal human machine interface for controlling an intelligent wheelchair using face movements,”IEEE International Conference on Robotics and Biomimetrics, Volume 4, December 7-11, 2011.

[6] Ram Pratap Sharma and Gyanendra K. Verma, “”Human Computer Interaction using Hand Gesture,”National Institute of Technology Kurukshetra, Kurukshetra, Haryana 136 119, India,

[7] M.Carmel Sabia, V.Brindha, A.Abudhahir, ”Facial Expression Recognition Using PCA Based Interface for Wheelchair,” 2014 International Conference on Electronics and Communication System (lCECS -2014)

[8] Saumyarup Rana, M Prasanna Deepu, Sivanantham S and Sivasankaran K, ”Face Detection System Using FPGA,” 2015 Online International Confernece on Green Engineering and Technologies (IC-GET2015), 2015

[9] ”Cyberoam Web Application Firewall Brochure”, https://www.cyberoam.com/downloads/Brochure/CyberoamWAFBrochure.pdf, Sept 2017.

Smart Farm: An Automated Farming Technique Using Robot

Shashank Patil, Mefania Charles, Nikit Gondhali
Student, Department of Computer Engineering
PIIT, New Panvel, India

Dipti Patil [1], Tusharika Banerjee Sinha [2]
Professor, Department of Computer Engineering, PIIT, New Panvel, India [1]
Professor, Department of EXTC, PIIT, New Panvel, India [2]

***Abstract*—With India being an agricultural land, the need of automation in farming will always exist. The implementation of this system can be done through a robot equipped with various sensors such as humidity sensor, IR obstacle avoidance sensor. These components will be enabled with IOT to perform the task of automated farming. It will allow the robot to connect to measure the soil moisture and temperature and will give the results to the user, if all the parameters are suitable for harvesting. With IOT the robot will be able to perform ploughing, sowing, spraying pesticides over a selected area in the farm. The robot can accept the request from the user through a mobile application and will prepare a list of tasks to be performed. This would be stored in a database and the robot will perform all the operations of automated farming that haven't been thought of without any human efforts. ESP can be used to get the data and control the bot continuously. A camera to carry out the surveillance connected to the remote would be an addition to the system. Robot can be connected to the server through internet with a suitable protocol.**

I. INTRODUCTION

1.1 Fundamentals

The Internet of things (IoT) is the network of physical devices, vehicles, and other items embedded with electronics, and network connectivity which enable these objects to collect and exchange data. Each thing is uniquely identifiable through its embedded computing system but is able to inter operate within the existing internet infrastructure. Experts estimate that the IoT will consist of about 30 billion objects by 2020.

The IoT allows objects to be sensed or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems, and resulting in improved efficiency, accuracy and economic benefit in addition to reduced human intervention. When IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of cyber-physical systems, which also encompasses technologies such as smart grids, virtual power plants, smart homes, intelligent transportation and smart cities.

1.2 Objectives

In this project we are going to implement an AUTOMATED FARMING ROBOT. The main objective of the project is to focus on automation in farming so that robot performs the most possible tasks that are required for farming. The robot will perform the following task such as ploughing, sowing of seeds, soil moisture detection, spraying of pesticides and water irrigation. The robot will be controlled by an android application in our mobile via Internet. The IOT modules acts as a communication link between android application and robot, thus depending upon the input given the robot will perform the task.

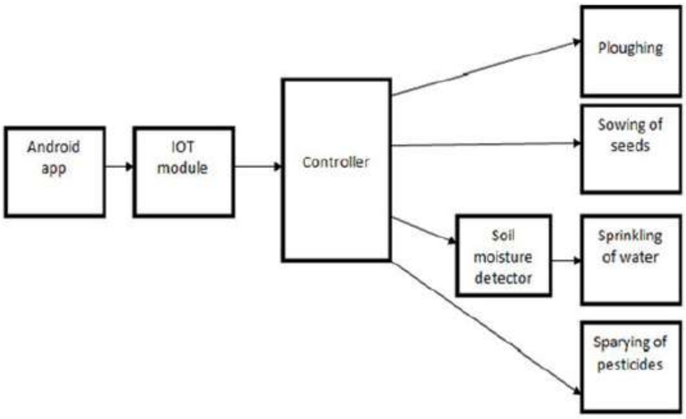


Figure 1: Generalized Block Diagram

1.3 Scope

Our farm equipment companies and researchers have developed a lot of small and heavy farm equipment for traditional farming needs but some kind of robotic and pneumatic mechanism are required in precision farming. The use of robots helps us in accuracy so that only particular amount of seeds is sowed and amount pesticides being sprayed and water is also conserved.

II. METHODOLOGY

A. Overview

The android-based Farming system is an automatic robot which performs multiple operations in the field of agriculture. In this project, we have implemented an automated farming robot. The implementation of this system is done through a robot equipped with soil moisture detector, camera, IR sensor and water spraying module. The components are enabled with Internet of Things that is through internet connectivity, the robot performs sowing, ploughing, spraying fertilizers and water over a selected area in the farm. The robot works on solar power. The robot accepts request from the user through a mobile application and executes the requested task. It performs all the operations without any human intervention. ESP is used to get the data and control the robot continuously. A camera to carry out the surveillance is connected to the robot.

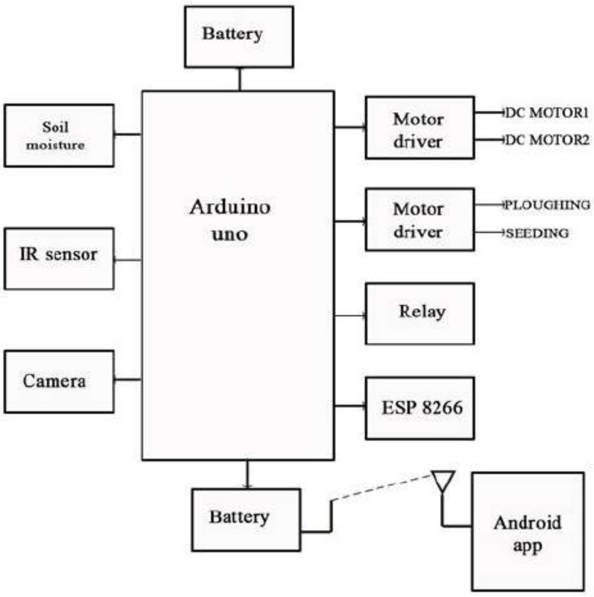


Figure 2. Block Diagram

It uses Arduino Uno which is programmed to receive the input signal of multiple sensors of the field. Once the controller receives this signal, it generates an output that drives a relay for operating the seeding and other circuitry which provides automatic control action on field. If the user sees the moisture level of every channel has sufficient amount then user can switch off the motor easily using GUI. An android mobile operating system application is interfaces with the microcontroller to control the action on the field. The soil moisture sensing arrangement is made by using two copper rods inserted into the field at a distance. Connections from the metallic rods are interfaced to the control unit. This signal is sensed to application which provides Graphical User Interface (GUI).

B. Techniques used

Spraying of Pesticide: The pesticide liquid present in a tank flows through a rubber pipe to the tip of DC motor, at that shaft of motor a fan blade is attached, which revolves at the delay time of robot or on front of crop. Due to revolution the liquid gets sprayed on the crops. The standard level is maintained by how much time delay we provide to the robot or the time in which the robot stands in front of crop.

Dropping of Seeds: The dropping of seed is done using the stepper motor mechanism. For that we are using the special mechanical head at the shaft of stepper motor. When the point on the farm where we want to drop the seed reaches, the stepper motor moves in a clockwise direction.

due to clockwise step angle change by stepper motor the tip of stationary as well as rotator container get match, due to matching of this tip`s the seed`s get path to dropped in the farm after very small delay of time the stepper motor moves in anticlockwise direction with same angle and the tips get close. In this way the controlling action of motor takes place at equal distance of farm, and also it dropped quantities seed`s on the farm.

Soil Moisture: The soil moisture sensor consists of two probes which are used to measure the Volumetric content of water. The two probes allow the current to pass through the soil and then it gets the resistance value to measure the moisture value. When there is water, the soil will conduct more electricity which means that there will be less resistance. Therefore, the moisture level will be higher. Dry soil conducts electricity poorly, so when there is less water, then the soil will conduct less electricity which means that there will be more resistance. Therefore, the moisture level will be lower.

Ploughing: This application is very easily achieved by attaching the attachment at the back side of the robot. For this application we require to give good mechanical strength to the robot, because it is quite heavy and when it is placed on soil for ploughing purpose, it require extra force to move forward. This is the initial operation in the farm. Once it is placed on the farm it continuously tracks the white line on the farm and does the ploughing through the attachment.

Power Supply: For becoming system echo friendly and beneficial for farmer we are going to provide the solar panel as a source power to the operation of whole process. Eco friendly in the sense as it doesn't require any fuel and source for operation, it saves electricity and fuel. Minimum pollution as well as saves the convention power. Due to open space of farming field it will be easily available, exception is the cloudy environment in rainy season. The solar energy is non-conventional source of energy so we can make system life longer.

III. HARDWARE SPECIFICATION

Arduino Uno: Arduino is a tool for making computers that can sense and control more of the physical world than your desktop computer. It's an open-source physical computing platform based on a simple microcontroller board, and a development environment for writing software for the board. The Arduino Uno is a microcontroller board based on the ATmega328. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with an AC-to-DC adapter or battery to get started.

Power Supply: The performance of the master box depends on the proper functioning of the power supply unit. The power supply converts not only A.C into D.C, but also provides o/p voltage of 5 volts, 1amp.

Motor Driver: Since motors require more current then the microcontroller pin can typically generate, you need some type of a switch (Transistors, MOSFET, Relay etc.,) which can accept a small current, amplify it and generate a larger current, which further drives a motor. This entire process is done by what is known as a motor driver. L293D is a typical Motor driver or Motor Driver IC which allows DC motor to drive on either direction. L293D is a 16-pin IC which can control a set of two DC motors simultaneously in any direction. It means that you can control two DC motor with a single L293D IC, Dual H-bridge Motor Driver integrated circuit (IC). The l293d can drive small and quiet big motors as well.

DC motor: In any electric motor, operation is based on simple electromagnetism. A current-carrying conductor generates a magnetic field; when this is then placed in an external magnetic field, it will experience a force proportional to the current in the conductor, and to the strength of the external magnetic field. The internal configuration of a DC motor is designed to harness the magnetic interaction between a current-carrying conductor and an external magnetic field to generate rotational motion.

Soil moisture sensor: Most soil moisture sensors are designed to estimate soil volumetric water content based on the dielectric constant (soil bulk permittivity) of the soil. The dielectric constant can be thought of as the soil's ability to transmit electricity. The dielectric constant of soil increases as the water content of the soil increases. This response is due to the fact that the dielectric constant of water is much larger than the other soil components, including air. Thus, measurement of the dielectric constant gives a predictable estimation of water content.

Obstacle Sensor: It consists of three major components. The first is an Infra-Red (IR) transmitter (usually an IR LED), the second is a TSOP (an Infra-Red receiver) and third IC 555.The main difference between LED and IR LED is that IR LED emits Infrared Radiations, which we cannot see by our naked eyes. TSOP requires the incoming data to be modulated at a particular frequency and would ignore any other signals. It is also immune to ambient IR light. They are available for different carrier frequencies from 32 kHz to 42 kHz.

Relay: A relay is an electrical switch that uses an electromagnet to move the switch from the off to on position instead of a person moving the switch. It takes a relatively small amount of power to turn on a relay but the relay can control something that draws much more power. A relay is used to control the air conditioner in your home. The AC unit probably runs off of 220VAC at around 30A. That's 6600 Watts! The coil that controls the relay may only need a few watts to pull the contacts together.

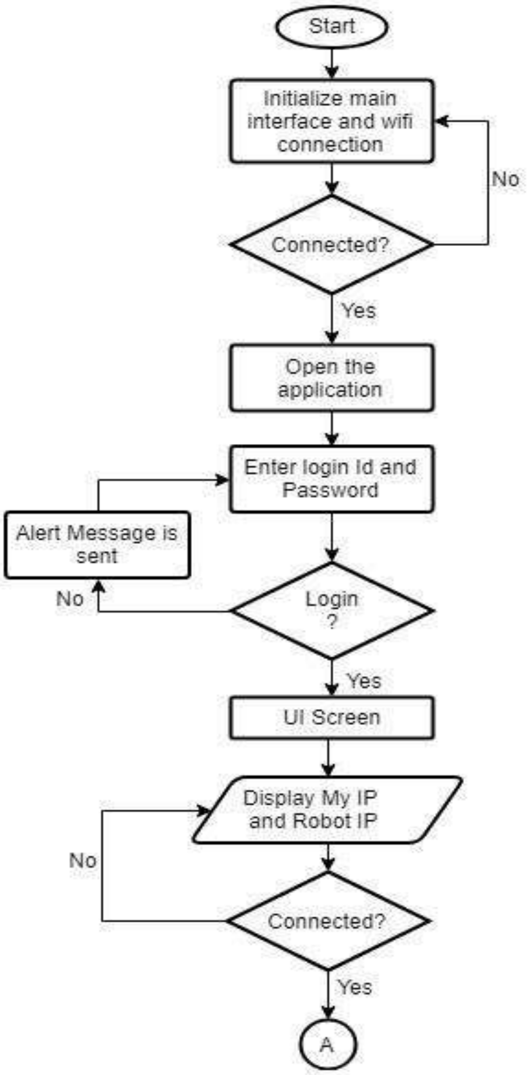
Solar Panel: Solar panels are devices that convert light into electricity. They are called "solar" panels because most of the time, the most powerful source of light available is the Sun, called Sol by astronomers. Some scientists call them photovoltaic which means, basically, "light-electricity." A solar panel is a collection of solar cells. Lots of small solar cells spread over a large area can work together to provide enough power to be useful. The more light that hits a cell, the more electricity it produces, so spacecraft are usually designed with solar panels that can always be pointed at the Sun even as the rest of the body of the spacecraft moves around, much as a tank turret can be aimed independently of where the tank is going.

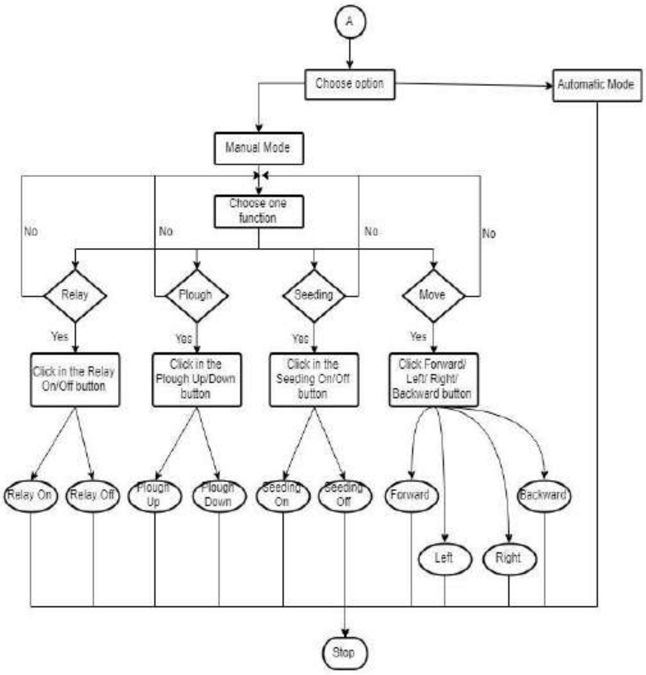
IV. SOFTWARE SPECIFICATION

Arduino IDE: A program for Arduino may be written in any suitable programming language for a compiler that produces binary machine code for the target processor. Atmel provides a development environment for their microcontrollers, AVR Studio and the newer Atmel Studio. The Arduino project provides the Arduino integrated development environment (IDE), which is a cross-platform application written in the programming language Java. It originated from the IDE for the languages *Processing* and Wiring. It includes a code editor with features such as text cutting and pasting, searching and replacing text, automatic indenting, brace matching, and syntax highlighting, and provides simple *one-click* mechanisms to compile and upload programs to an Arduino board. It also contains a message area, a text console, a toolbar with buttons for common functions and a hierarchy of operation menus. A program written with the IDE for Arduino is called a sketch. Sketches are saved on the development computer as text files with the file extension .ino. Arduino Software (IDE) pre-1.0 saved sketches with the extension .pde. The Arduino IDE supports the languages C and C++ using special rules of code structuring. The Arduino IDE supplies a software library from the Wiring project, which provides many common input and output procedures. User-written code only requires two basic functions, for starting the sketch and the main program loop, that are compiled and linked with a program stub main() into an executable cyclic executive program with the GNU tool chain, also included with the IDE distribution. The Arduino IDE employs the program to convert the executable code into a text file in hexadecimal encoding that is loaded into the Arduino board by a loader program in the board's firmware.

Basic4Android: Basic4Android (currently known as B4A) is a rapid application development tool for native Android applications, developed and marketed by Anywhere Software Ltd. B4A is an alternative to programming with Java. B4A includes a visual designer that simplifies the process of building user interfaces that target phones and tablets with different screen sizes. Compiled programs can be tested in AVD Manager emulators or on real Android devices using Android Debug Bridge and B4A Bridge. The language itself is similar to Visual Basic and Visual Basic .Net though it is adapted to the native Android environment. B4A is an object-based and event-driven language. B4A generates standard signed Android applications which can be uploaded to app stores like Google Play, Samsung Apps and Amazon App store. There are no special dependencies or runtime frameworks required.

V. FLOWCHART


Start
Initialize main interface and wifi connection
No
Connected?
Yes
Open the application
Enter login Id and Password
Alert Message is sent
No
Login ?
Yes
UI Screen
Display My IP and Robot IP
No
Connected?
Yes
A



V. CONCLUSION

Considering the decrease in the labour and with the increase in the population there is a need of automation in agriculture. This robot not only reduces the labour but increases the accuracy of seeding and ploughing. The farmers do not come in direct contact with poisonous pesticides due to spraying mechanism. It provides soil moisture which leads to reduction in the usage of water. There is a surveillance camera so that the farmer can have a view of his field always. Also, it reduces the labour cost as well as the total cost of this product is less and affordable.

REFERENCES

[1] S.S. Katariya, S.S. Gundal, Kanawade M.T and Khan Mazhar, “Automation in Agriculture”, International Journal of Recent Scientific Research, Vol. 6, Issue, 6, pp.4453-4456, June, 2015.

[2] Hemant M. Sonawane, Dr. A.J. Patil, “Overview of Automatic Farming & Android System”, International Journal of Engineering Trends and Applications (IJETA) – Volume 2 Issue 3, May-June 2015.

[3] Hariharr C Punjabi, Sanket Agarwal, Vivek Khithani and Venkatesh Muddaliar, “Smart Farming using IoT”, International Journal of Electronics and Communication Engineering and Technology (IJECET) Volume 8, Issue 1, January - February 2017, pp. 58–66.

[4] Abdulrahman, Mangesh Koli, Umesh Kori, Ahmadakbar, “Seed Sowing Robot”, International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 2, Mar – Apr 2017.

AUTONOMOUS NAVIGATION SYSTEM

Praveen Pandey,Kevin Mistry,Aiswarya kamble,Sushila Yadav
Department of Computer Engineering, University of Mumbai
PCE, New Panvel, India

Abstract: This paper present the simultaneous localization and map building (SLAM) problem asks if it is possible for an autonomous vehicle(drone) to start in an unknown location in an unknown environment and then to incrementally build a map of this environment while simultaneously using this map to compute absolute vehicle location.It is then shown that the absolute accuracy of the map and the vehicle location reach a lower bound defined only by the initial vehicle uncertainty. This paper also includes the solutions for obstacle detection and collision avoidance of UAVs exist, these solutions suffer from different drawbacks.In this study, an offline statistical estimation algorithm based on Extended Kalman Filter method is developed to solve the SLAM problem. For the application, a robot equipped with only simple and cheap sensors is used. Two of the most frequent problems in SLAM algorithms which are known as loop closing and data association are effectively solved by Extended Kalman Filter method.

Keywords:Navigation,Accessibility,Localization, ROS,RGBD-SLAM ,Flight Controller,RTAB MAP.

1. Introduction

In the past decade, the interest in UAVs and autonomy has constantly increased. Collision avoidance is an important requirement for autonomous flights. Although multiple solutions for obstacle detection and collision avoidance of UAVs exist, these solutions suffer from different drawbacks. To explore the capability and without any human interference system is designed.Anyone involved in mining knows that worker safety is of paramount importance. By allowing surveyors to collect accurate spatial data from above, drone or UAV technology can vastly reduce risk by minimising the time these staff spend on site.The challenges face in self-exploratory oriented autonomous mobile robot is the environment factors which have numerous complex geographical landmarks and also to detect an obstacles.

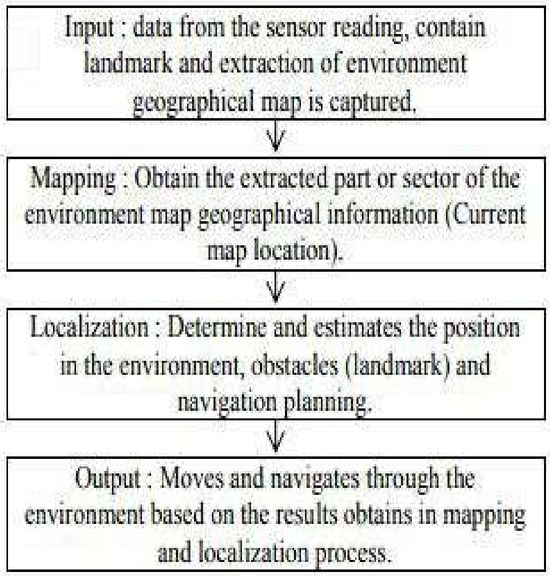
Autonomous drone is a drone that capable to act and perform the designated tasks itself

without the human interference. The autonomous drone or more scientifically called as artificial intelligence robot able to 'think' when making decision and 'act' based on the decision make. A key prerequisite for a truly autonomous robot is that it can simultaneously localize itself and accurately map its surroundings.

1.1.Simultaneous Localization And Mapping(SLAM)

SLAM can be applied to real-life problems such as natural disasters.During an earthquake, SLAM can be used to create a map that will allow a rescue agent to help victims find their way back or locate the right path. This method can also be used to find victims in a collapsed building. In the medical field, it can be used to create a map for endoscopy activities. It is implemented in some real-life applications, such as oil pipeline inspection, ocean surveying and underwater navigation, mine exploration, coral reef inspection, military applications, and crime scene investigation.

Solving the SLAM problem has become a popular area of research in the past years. SLAM problems generally include four major units, namely, sensor uncertainty, correspondence problem, loop-closing problem, and time complexity.problem, loop-closing problem, and time complexity (Begum, Mann & Gosine, 2008). Sensor uncertainty explains the noise of each instrument used.The correspondence problem is the difficulty of different viewpoints and the finding of a similarity between the same object from each view po int.


Input : data from the sensor reading, contain landmark and extraction of environment geographical map is captured.
Mapping : Obtain the extracted part or sector of the environment map geographical information (Current map location).
Localization : Determine and estimates the position in the environment, obstacles (landmark) and navigation planning.
Output : Moves and navigates through the environment based on the results obtains in mapping and localization process.

1.1.General Steps Perform In Autonomous Robot

1.2.RGB-D SLAM: We present an RGBD SLAM algorithm that uses geometric information provided by a 3D model to improve the camera poses estimation. Our algorithm relies on a local bundle adjustment; the cost function to be minimized is a combination of different types of residual errors: error based on visual features, error based on depth data and error based on geometric constraints provided by a 3D model of the environment. We demonstrate that this additional constraint in the bundle adjustment improves the accuracy and the robustness of the RGBD SLAM. This new solution is efficient for global localisation in indoor environments.

2.Problem Statement

Assume to have a mobile robot placed in an unknown environment, the robot is equipped with on board ultrasonic sensors able to provide the distance of the robot from the environment bounds. Since there is no a-priori knowledge on the environment, to yield the output equation related to the on board sonar sensors, a model for the environment boundaries is required. In the following subsections the robot and the environment models will be described.

3.Literature Review

In this section we cite the relevant past literature of research work done in the field of "Autonomous Navigation System" to avoid obstacle using various technique.

The paper addresses the use of Adaptive Kalman filter based method over standard EKF as it suppresses problem of filtering divergence and implements mapping effectively[1].

The development of a robot control system that uses Block Matching algorithm for mapping along with pattern matching and obstacle avoidance using openNi and openCV libraries[2].

To solve the SLAM problem the Rao-Blackwellised particle filter (RBPF) is used as discussed by LuigiD' Alfonso,Andrea Griffo ,Pietro Muraca,Paolo Pugliese .It also concludes that laser sensors and cameras are best way to represent SLAM problem[3].

The F. Pirahansiah and S.Saharan discussed significant issue in the field of robotics that mens SLAM.It addresses the problem of the possibility for a mobile robot to be placed in an unknown location and environment, where it will incrementally build a consistent map of the environment while determining its location within this map. They also introduced different types of SLAM application such as real time application .SLAM problems generally include four major units, namely, sensor uncertainty, correspondence problem, loop-closing problem, and time complexity (Begum, Mann & Gosine, 2008)[4].

Henning Lategahn, Andreas Geiger, Bernd Kitt present in their paper a technique which is dense stereo V-SLAM algorithm for 3D representation incoordinate systems spanned by past robot poses which is one of the technique.Iconic kalman filters were used for increasing reconstruction accuracy[5].

The authors, L. D. Perera and E. Nettleton in this paper show that SLAM initialized with a known vehicle pose can be considered as a problem of parameter identification in unknown environment. Using a rank test for nonlinear map state identification, they establish that all the map states in the SLAM problem are identifiable given the initial conditions of the vehicle pose with zero uncertainty.They provide simulations of a Kalman filter based SLAM algorithm to verify the theoretical results shown above on the parameter identifiability perspectives of the SLAM problem. Simulations are done in the 2D environment.[6].

The paper divided SLAM problem into five different parts as landmark extraction, data association, state estimation, state update and landmark update. This segment based SLAM algorithm used the currently acquired measurements to update the actual environment mapping[7].

The authors"Hugh Durrant-whyte And Tim Bailey" discuss the simultaneous localization and mapping problem asks if it is possible for a mobile robot to be placed at an unknown location in an unknown environment and for the robot to incrementally build a consistent map of this environment while simultaneously determining its location within this map.SLAM has also been implemented in a number of different domains from indoor robots to outdoor, underwater, and airborne systems.Also introduced probabilistic methods were only just beginning to be introduced into both robotics and artificial intelligence (AI)[8].

The idea in the paper was to match recent sensory information against prior knowledge of the environment, i.e. world model which in their case was an occupancy grid map.Also they used sensor fusion approach which was based on nonlinear model based estimators: extended and unscented Kalman filter (EKF and UKF)[9].

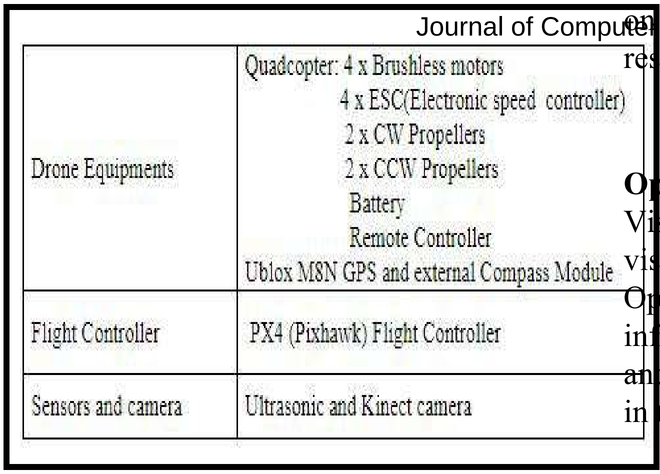
The authors S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte describe a scalable algorithm for the simultaneous mapping and localization (SLAM) problem. In the linear SLAM case with known data association, all updates can be performed in constant time; in the nonlinear case, additional state estimates are needed that are not part of the regular information form of the EKF[10].

The authors M. W. M. Gamini Dissanayake,Paul Newman ,Steven Clark ,Hugh F. Durrant-Whyte ,M. Csorba mentioned that the solution to the simultaneous localization and map building (SLAM) problem is, in many respects, a "Holy Grail" of the autonomous vehicle research community. The ability to place an autonomous vehicle at an unknown location in an unknown environment and then have it build a map, using only relative observations of the environment, and then to use this map simultaneously to navigate would indeed make such a robot "autonomous"[11].

4.Hardware And Software Specifications

The experiment setup is carried out on a computer system which has the different hardware and software specifications as given in Figure 4.1 and Figure 4.2 respectively.

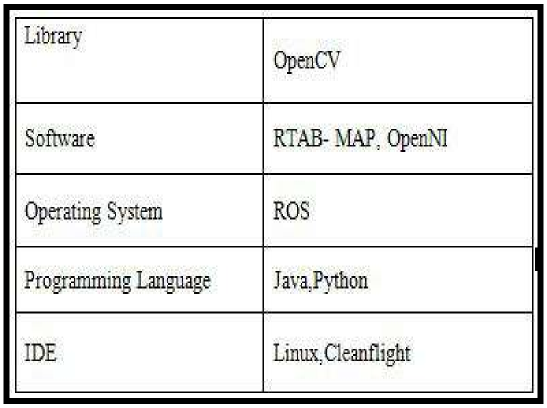
Hardware Details



Drone Equipments	Quadcopter: 4 x Brushless motors 4 x ESC(Electronic speed controller) 2 x CW Propellers 2 x CCW Propellers Battery Remote Controller Ublox M8N GPS and external Compass Module
Flight Controller	PX4 (Pixhawk) Flight Controller
Sensors and camera	Ultrasonic and Kinect camera

FIGURE 4.1

Software Details



Library	OpenCV
Software	RTAB- MAP, OpenNI
Operating System	ROS
Programming Language	Java,Python
IDE	Linux,Cleanflight

FIGURE 4.2

5. TECHNOLOGY

- **ROS:**ROS is an open source, meta operating system for robot. It provides the services including hardware abstraction, low-level device control, implementation of commonly-used functionality, message passing between processes, and package management.

- **RTAB MAP :**It is a Real Time Appearance Based Mapping (RTAB Map) which is RGB-D Graph-Based SLAM approach based on an incremental appearance-based loop closure detector. The loop closure detector uses a bag-of-words approach to determine how likely a new image comes from a previous location or a new location. When a loop closure hypothesis is accepted, a new constraint is added to the map's graph, then a graph optimizer minimizes the errors in the map. A memory management approach is used to limit the number of locations used for loop closure detection and graph optimization, so that real-time constraints on large-scale environments are always respected.

OpenCV: OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.

OpenNI: OpenNI or Open Natural Interaction is an industry led non profit organization and open source software project focused on certifying and improving interoperability of natural user interface and organic user interfaces for Natural Interaction(NI)devices,application that use those devices and middleware that facilities access and use of such devices

6.PROPOSED SYSTEM

For an autonomous navigation system, there are many agent travelling and mapping the outdoor environment.The objective is to find the technique or solution to makes the robot capable to autonomously navigate without any prior knowledge on the environment it explores. Analysing and examining different working projects in the domain various systems were found that met the interest the project.Among the found system,a autonomous system was discovered which had an autonomous drone which simply flies itself by detecting objects and mapping its surrounding with the help of sonar sensors.Along with it an additional system which had a depth camera with maps it surrounding more effectively compared to the sonar sensors was looked upon.So at the end inferring upon the systems analysed, using a combination of the systems together a system more efficient than existing systems was proposed to be developed.

AUDRONE ARCHITECTURE

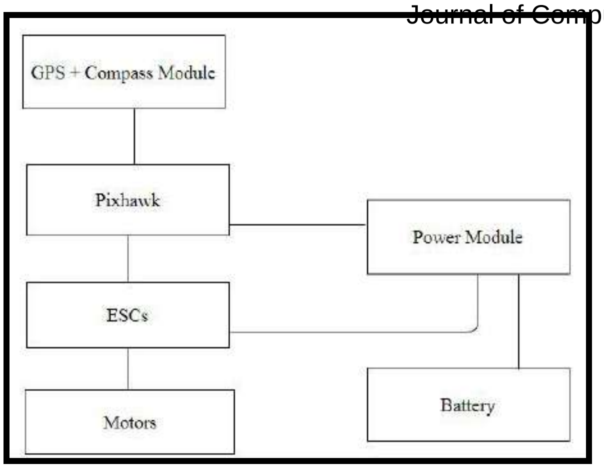


Figure 6.1

The figure above presents the architecture of AuDrone system which briefs us the steps of the working of the system from to data input to the defined output.

6.2 Working of kinect camera

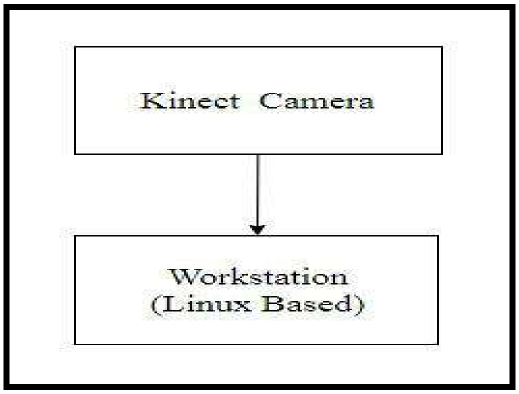
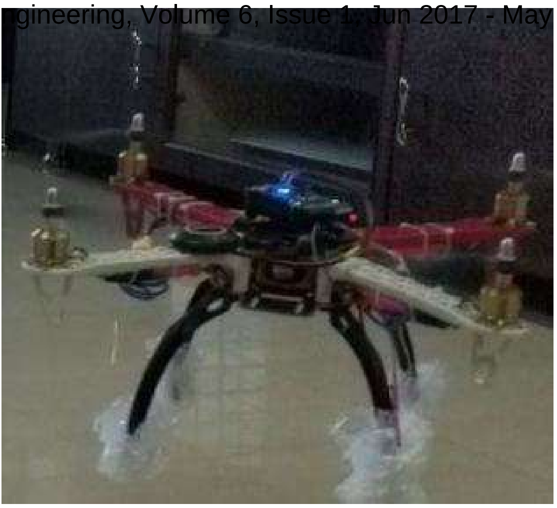


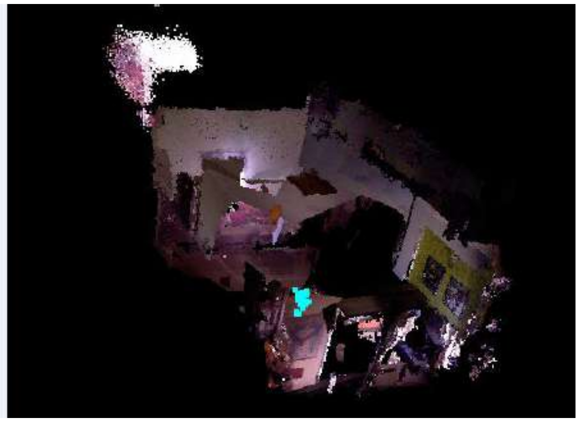
Figure 6.2

This figure above presents working of kinect camera.

7.Drone Structure and Observation



The above image shows design model of drone.



This image show map captured through kinect camera.

8.Conclusion

In this paper,the study of autonomous navigation system techniques is presented.The different techniques of SLAM used for localization ,mapping,object detection and avoidance of object.As with current graph-based RGB-D SLAM algorithms, our filter-based RGB-D SLAM in this paper does not depend on other sensors (such as gyroscope, encoder, *etc*). Our contribution consists of providing an appropriate observation model and motion model for the SLAM for a robot.The comparative study of various techniques mentioned above with robot is presented in this paper.

9.Reference

[1] Xiangyuan Jiang, Tingting Li and Yunhua Yu, A Novel SLAM Algorithm with Adaptive Kalman Filter, 2016.

[2] Putov Viktor Vladimirovich, Putov Anton Viktorovich, Ignatiev Konstantin Vasil'evich, Belgradskaya Elena Valer'evna, Kopichev Michael Mikhailovich,Autonomous Three-Wheeled Robot with Computer Vision System, 2015.

[3] Luigi D'Alfonso,Andrea Griffo,Pietro Muraca,Paolo Pugliese, A SLAM algorithm for indoor mobile robot localization using extended kalman filter and a segment based environment mapping, 2013.

[4] F. Pirahansiah and S.Saharan, Simultaneous Localization And Mapping Trends And Humanoid Robot Linkages, 2013.

[5] Henning Lategahn, Andreas Geiger and Bernd Kitt, Visual SLAM for Autonomous Ground Vehicles, 2011.

[6] L. D. Perera and E. Nettleton, The Simultaneous Localization and Mapping problem in a nonlinear parameter identifiability perspective, 2010.

[7] Bailey et Al , Simultaneous localization and mapping : part 1, 2006.

[8] Hugh Durrant-Whyte and Tim Bailey, Simultaneous localization and mapping : part 2, 2006.

[9] Edouard Ivanjko, Mario Vasak, and Ivan Petrovi, Kalman filter theory based mobile robot pose tracking using occupancy grid maps, 2004.

[10] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, Simultaneous localization and mapping with sparse extended information filters, 2004.

[11] M. W. M. Gamini Dissanayake,Paul Newman ,Steven Clark ,Hugh F. Durrant-Whyte ,M. Csorba, A solution to the simultaneous localization and map building (SLAM) problem, 2001.

Offensive Language Detection using AI Technique

Sneha Birajdar Shivani Dalvi Jagruti Dandekar Aishwarya Ganesan

sne.biraj@gmail.com dalvishivani06@gmail.com jagrutidandekar796@gmail.com gaishwarya58@gmail.com

Department of Computer Engineering, Mumbai University
Pce, New Panvel, India

Abstract:
Social Network has become a place where people from every corner of the world has established a virtual civilization. Text messaging through the Internet or cellular phones has become a major medium of personal and commercial communication. Such text may contain abusive words. Although a human could recognize these sorts of useless annoying texts among the useful ones, it is not an easy task for computer programs. We describe an automatic invective language detection method which extracts features and applies classification methods for invective language detection. The target of offensive document detection is to give output classification for a document provided by user using neural network. In this approach, classification is done by neural network.

Keywords:-- offensive document detection, neural network.

I. INTRODUCTION

An online social network (OSN) shall be defined as the use of dedicated websites and applications that allow the users to interact with other users, or to find people with similar interests to one's own. The social networking sites enable the people worldwide in stay in touch with each other irrespective of ages. The children in special are introduced to a bad world of worst experiences and harassments. The users of the social networking sites might be unaware of various vulnerable attacks hosted by the attackers in these sites.

Today internet has become a part of people's daily life. People use social network to share pictures, music, video etc., social network allows user to connect to various other pages on the web, including some useful sites like education, marketing, online shopping, business, e-commerce. Social networks such as Facebook, LinkedIn, MySpace, Twitter are more popular recently. Offensive language Detection is a natural language processing task that deals with finding whether any kind of abusive words(i.e related to religion,sex,racism,defecation,etc) are present in a given document and classify the document accordingly. The document which will be classified in OFLD is in english text format which can be mined from tweets, comments on social media, reviews on movies, political reviews, feedbacks.

II. LITERATURE REVIEW

[1]. The idea of creating such a system was implemented very early but many failed attempts occurred. This section consists of various works already been done on offensive or hate speech detection and techniques for classification of various documents using neural network.

[2]. This makes the study extensive, strong and objective. The pre processing task such as stop words removal is a very important task as it

does not play any important role in information retrieval. Stemming is used to remove all the possible suffixes from the keyword and gives stem of word. Vikas S Chavan,Shylaja S S worked on preprocessing and feature extraction[1]. The authors G.Vinodhini, RM.Chandrasekaran in [2] have specified that Back Propagation Neural Networks is supervised machine learning methods which analyzes data and recognizes the patterns that are used for classification.This work focuses on binary classification to classify the text sentiment into positive and negative reviews.

[3]. Here the authors Zhan Wang, Yifan He and Minghu Jiang in [4] have examined effectiveness of Radial Basis Function which complex than Multi-Layer Feedforward Neural Networks. Neural network using Multi-Layer Feedforward Neural Networks is presented in this paper for offensive language detection. It consists of an input layer, an output layer and several hidden layers. The hidden layer can be seen as a “distillation layer” that refines and extracts some of the important patterns from the inputs and passes it onto the next layer to see. It makes the network efficient and faster by identifying only the important information from the inputs leaving out the unimportant information. The Levenberg-Marquardt algorithm is used which is faster to converge than either the Gauss-Newton and Gradient descent on its own.

III. PROPOSED ARCHITECTURE

OFLD system provides accurate and precise offensive content detection associated with input document.The goal of any OFLD system is to detect offensive language associated with the subject.The output showing classification which demonstrates whether document associated with the subject is offensive or non-offensive.The objective of proposed OFLD system in this dissertation is to process the text file given as input and find the classification of the text file(i.e,offensive or non-offensive).The main purpose of this dissertation is to use one of the machine learning approaches which is better for offensive language detection giving more accurate results for classification of the documents based on training.

OFLD works in multiple stages and it uses neural network for classification. The system takes document as input. Input must be clear in formatting i.e. two words must be separated by white space and two sentences must be separated by punctuation mark. After taking input from user it will tokenize and remove stop words. After that stemming is performed on the output that is produced of the above step. Then the remaining words are sent to neural network for calculating feature values and thus sending it to neural network toolbox for simulation. After that, testing sample feature matrix data and the trained feature matrix data are used to find the class of the test sample input.The system will display the number of abusive words, class of the document and then the polarity in percentage.

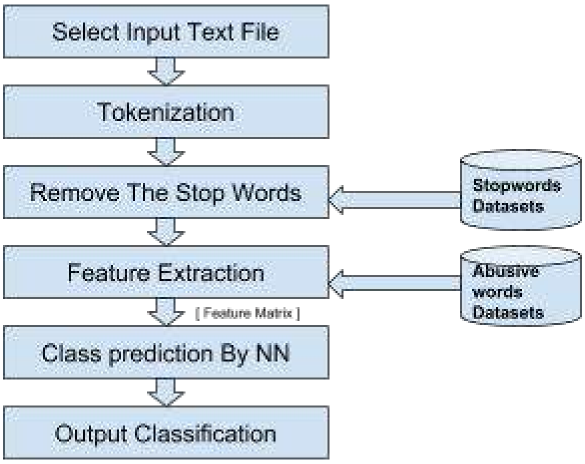


Figure 1. Block diagram

IV. EXPERIMENTS AND RESULTS

4.1. Select input text file

Sample input text : My god,the weird bitch is talking to me now. I think she has one of those disease where you lack social and interpersonal skills. These bitches dont care they just play their role. Dumb bitches do dumb things.

4.2. Pre-processing

This step involves processing of input provided by user to extract offensive words. As most of the reviews and comments that are available online contains unnecessary and unimportant data. So it is necessary to first clean and filter the input, so that unwanted content in the text are removed. Also reviews or comments may be single line or paragraph or complete document which needs to be broken down into individual tokens. Finally,there will be many words present in the input which mostly do not participate in calculation of overall polarity such as stop words are removed.

1) Stop Word Removal

Stop words are considered as the words in the documents which have no importance. These are mainly the words which are used for grammatical arrangement of a text. Stop words are common words that carry less important meaning than the other words. These words should be eliminated as they play no part in extraction of offensive words. We have gathered a set of stop words and each word is one by one compared with that set and when match is found, then those words are removed from the input text.Some examples of stop words are : a
About, above, after, again, against, all, am, an, and, any, are, as, at, be, because, been, before, being, below, between, both, but, by, etc.

Input : My god,the weird bitch is talking to me now. I think she has one of those disease where you lack social and interpersonal skills. These bitches dont care they just play their role. Dumb bitches do dumb things.

Output : My god, weird bitch talking. I think she one disease you lack social interpersonal skills. Bitches care play role. Dumb bitches dumb things.

2) Normalization

It is a process that chops off the end of the words in the hope of achieving the goal correctly often includes removal of derivational prefix and suffix. Porter's algorithm is an algorithm that is considered as the most effective and efficient algorithm for stemming.

Input : My god, weird bitch talking. I think she one disease you lack social interpersonal skills. Bitches care play role. Dumb bitches dumb things.
Output : My god, weird bitch talk. I think she on diseas you lack social interperson skill. Bitch care play role. Dumb bitch dumb thing.

4.3. Feature Extraction

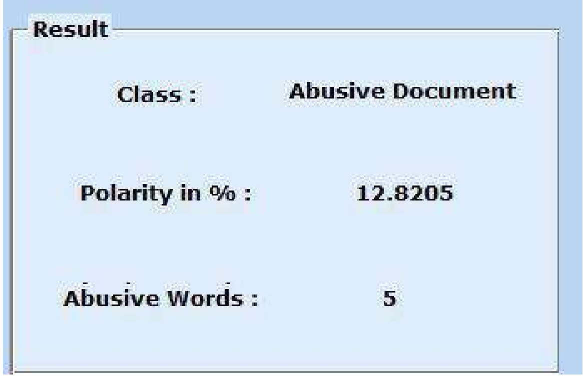
In this phase, features are calculated features such as term frequency, inverse document frequency, count, weight. Lexicon based are statistical feature selection methods can be used to select features from documents which treats document as bag of words(BOW) or string. Stemming and removal of stop words are mostly common feature selection step. Here 4 features are used.

4.4. Classification result by Neural Network

Result by NN classification

Input : My god,the weird bitch is talking to me now. I think she has one of those disease where you lack social and interpersonal skills. These bitches dont care they just play their role. Dumb bitches do dumb things.

Output :



Result

Class : Abusive Document

Polarity in % : 12.8205

Abusive Words : 5

Figure 2 : Output

This can be represented with the help of graph which displays the number of words after each step :
1.Number of words initial document
2. Number of words after stop words removal
3. Number of abusive words

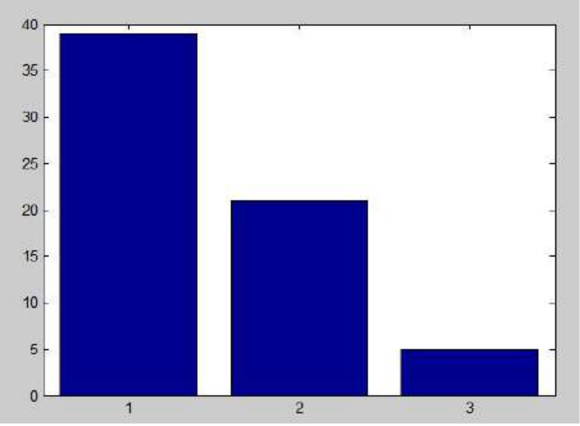


Figure 3 : Result

V. CONCLUSION

Offensive language detection has lead to identify offensive documents. Abusive words can be mined from blogs, texts, social media, news, articles, comments or any other source of information.

Offensive document detection has become quite popular with its application. This system allows users to find offensive word counts with the document and their overall polarity in percentage is calculated using classification by neural network.

The neural networks aimed at providing artificial intelligence to the system. The most helpful neural network in function approximation are Radial Basis Function (RBF) and Multi-Layer Feedforward Neural Networks networks. As Radial Basis Function is more complex here we focus on Multi-Layer Feedforward Neural Networks. Neural network using Multi-Layer Feedforward Neural Networks is presented in this paper for offensive language detection. It consists of an input layer, an output layer and several hidden layers. The hidden layer can be seen as a “distillation layer” that refines and extracts some of the important patterns from the inputs and passes it onto the next layer to see. It makes the network efficient and faster by identifying only the important information from the inputs leaving out the unimportant information.

VI. FUTURE SCOPE

The system can be implemented for different training functions, n number of hidden layers and n number of neurons.Also auto updating of input files using web crawlers, training files, auto training of input files with no human intervention. This system can also work on other indian languages if provided proper resources.

For improving the accuracy of the system, more feature can be added such as Part - Of - Speech (POS) i.e a POS tagger can be used for tagging noun, verb, adverb and adjective of each word. Accuracy can also be improved by considering huge amount of training and testing datasets. As Porter stemming algorithm used in matlab has less accuracy,it can be replaced by some other algorithm which will be more efficient and will give more precise result.

REFERENCES

[1] Vikas S Chavan, Shylaja S S "Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network ".

[2] G.Vinodhini, RM.Chandrasekaran "Sentiment Classification Using Principal Component Analysis Based Neural Network Model".

[3] Cheng Hua Li and Soon Cheol Park "Artificial Neural Network for Document Classification Using Latent Semantic".

[4] Zhan Wang, Yifan He and Minghu Jiang "A Comparison among Three Neural Networks for Text Classification".

[5] Chikashi Nobata, Joel Tetreault, Achint Thomas, Yashar Mehdad, Yi Chang "Abusive Language Detection in Online User Content".

[6] Zaidah Ibrahim,Dino Isa,Rajprasad Rajkumar and Graham Kendall "Document Zone Content Classification for Technical Document Images Using Artificial Neural Networks and Support Vector Machines".

[7] Fawaz AL Zaghoul and Sami Al-Dhaheri "Arabic Text Classification Based on Features Reduction Using Artificial Neural Networks".

[8] Theodora Chu, Kylie Jue, Max Wang "Comment Abuse Classification with Deep Learning".

[9] Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki "Hate Speech on Twitter : A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection".

[10] YangZhenYu ,JingHui "A Study on Text Classification Based on Stacked Contractive Auto-Encoder".

[11] Xi Ouyang, Pan Zhou, Cheng Hua Li, Lijun Liu "Sentiment Analysis Using Convolutional Neural Network".

[12] Quanzhi Li, Sameena Shah, Rui Fang, Armineh Nourbakhsh, Xiaomo Liu "Tweet Sentiment Analysis by Incorporating Sentiment-Specific Word Embedding
and Weighted Text Features".

Secure Transmission of medical images using watermarking and cryptography with improved quality

Amit Tupdale[1], Aniket Tapre[2,] Pushkar Mhatre[3], Vijay Pratap[4], K S Charumathi[5]
Department of Computer Science
Pillai College of Engineering
Panvel, Maharashtra, India

Abstract

Nowadays telemedicine is used to remotely diagnose a patient. For this doctors need to exchange the medical images as well as medical reports to the health care facilities. But there is a problem for transmission of these images over insecure channels such as internet or drives. Hence we propose a system/software which uses digital watermarking as well as cryptographic function to hide the patient's data in the cover image before transmission. We are using invisible digital watermarking to hide the patient's medical records in the images and we are using AES algorithm to encrypt the patient's data before transmission over the network or drives. The proposed algorithm uses discrete wavelength transform in the frequency domain for transmission of medical images. We are also using reversible watermarking to maintain the authenticity of the image and to have end to
end security. At the receivers side the image will be authenticated.

Index Terms - Authentication, cryptography, Telemedicine, watermarking, wavelet transform.

I. INTRODUCTION

Telemedicine enables expert diagnosis and better healthcare access to distant patients especially in remote or rural areas by allowing the transmission of medical images through telecommunication. It has been used to overcome distance barriers and to improve access to medical services that would often not be consistently available in distant rural communities. It is also used to save lives in critical care and emergency situations. One application of telemedicine is the exchange of medical images between remotely located healthcare entities. However, a major obstacle telemedicine faces are providing confidentiality, integrity, and authenticity to transmitted medical images. To provide security to telemedicine we used digital watermarking in medical images and encryption of medical images.

Digital watermarking is the process that hides watermark data into a multimedia object such that the watermark can be detected or extracted from the object to prove its ownership or validate its integrity.

II. DIGITAL WATERMARKING

Watermark is inserted into a digital document (image, video, audio) a different kind of watermark is used to ensure security services such as (copyright, authentication, integrity, etc.). It is important, because on the one hand the extraction or removal of this information document becomes difficult and on the other hand the distortion introduced by the mark is imperceptible.

The size of this watermark depends on the image size and it is related to the existing patient records. At first randomized cryptographic fusion watermarking system was proposed. The system operates by encrypting the patient information and embedding the encrypted data in the medical image by bit-wise operation.

III. LITERATURE REVIEW

Ali AI-Haj , Noor Hussein and Gheith Abandahs [1] proposed paper proves to be the base paper for further research as it presents various methods to secure the transmission of medical images. The main objective of this paper is to use hybrid algorithm which combines encryption and digital watermarking techniques. A cryptographic watermark and the patient's data are hidden in the cover image before being transmitted over vulnerable public networks.

S. Nithya, K. Amudhas [2] proposed paper uses an approach SHA 256, AES and arithmetic compression techniques. The ROI of Medical images is irregularly placed in the area where the information is placed. The whole image of SHA 256 is embedded in insignificant bits of ROI .

Jaskaran Singh, Anoop Kumar Patels [3] proposed paper discusses about the proposed algorithm in which, the medical images are embedded as watermark into a special cover image. In this process, the cover image is transformed by discrete wavelet transform (DWT) and the LL sub-band obtained, is then transformed by discrete cosine transform (DCT). Finally, inverse discrete cosine (IDCT) and discrete wavelet transforms (DWT) are applied on modified sub bands to obtain the watermarked image.

Ali Al-Haj ,Gheith Abandah and Noor Husseins [4] proposed paper discuss mainly on the secure transmission of medical images with special standards which deals with the medical data security data issues . One such standard is the digital imaging and communication in medicine (DICOM) standard. Unlike the DICOM standard and other crypto-based schemes, the proposed algorithms provide confidentiality, authenticity and integrity for both constitutes of the DICOM

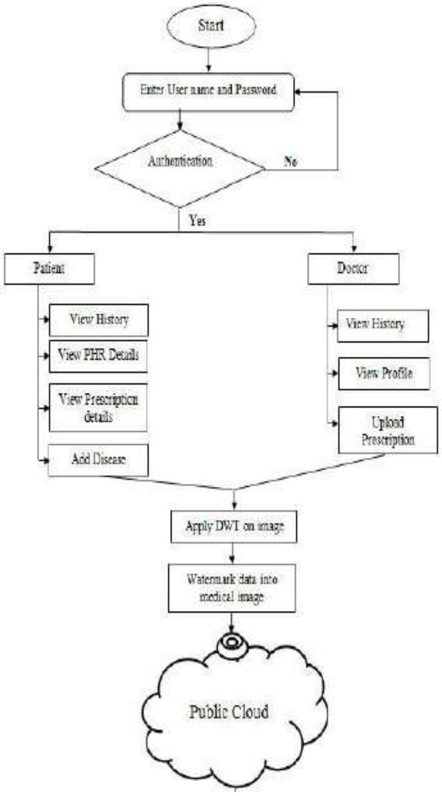
IV. METHODOLOGY

In this chapter we will be discussing about the proposed system architecture. The system will use a website as its graphical user interface. The system will provide login functionality for the doctor as well as the patient. A passkey will be generated for both the doctor and patient. The doctor will provide input to the system as uploading a medical image that is to be sent to the patient. At first part the system will generate a secret key using AES algorithm this will be used to encrypt the image that is to be sent. Then a digital watermark for the uploaded image will be created using discrete wavelet transform method. And then the watermarked image will be encrypted using a passkey and will be sent to the patient reverse method will be applied to get back the original medical image.

AES ENCRYPTION

AES is an iterative rather than Feistel cipher. It is based on substitution permutation network. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (per- mutations).

Interestingly, AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.

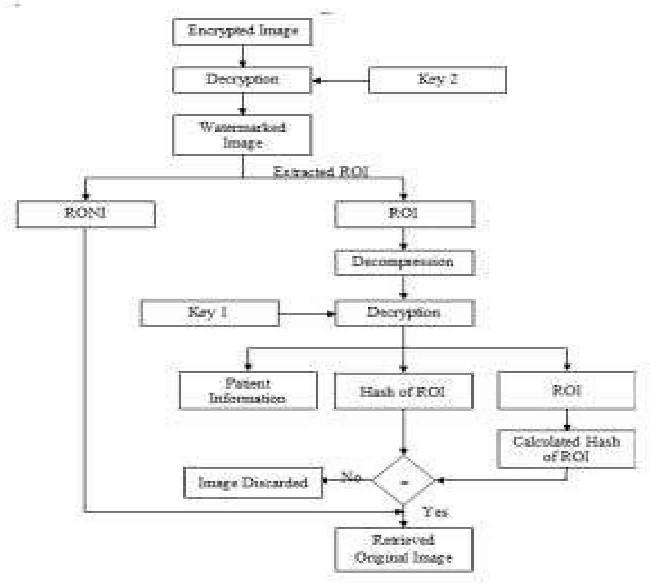


WATERMARK EMBEDDING:

The embedding procedure operates on the blocks of RoNI. The three security watermarks, which have been described in the previous section, are formulated in binary sequences, each of which is then embedded in selected DWT sub-bands in different multi-resolution levels. The procedure concludes by joining the un-watermarked ROI blocks and the watermarked RoNI blocks to form the overall watermarked image.

WATERMARK EXTRACTION PROCEDURE:

The extraction procedure is a direct reversal of the embedding procedure described above. A block diagram of the procedure is shown below.



RESULTS AND DISCUSSION:

1. Security: The proposed System uses AES algorithm which is one of the most secure algorithm in cryptographic science. AES supports larger key sizes than 3DES's 112 or 168 bits. AES is faster in both hardware and software. AES's 128-bit block size makes it less open to attacks via the birthday problem than 3DES with its 64-bit block size.

2. Performance: Our System uses concepts of cloud which is not susceptible to network traffic and has high performance.

3. Scalability: The system uses Dropbox for storing and uploading of Images transmitted by the doctors and the patient. The main advantage of Dropbox is that it is completely free. There are no upfront charges or any additional charges once you start using the service. When you register for a Dropbox account, you automatically get 2 gigabytes (GB) of storage space. This is a good amount of storage space.

4. Compression: The scheme of Reversible Watermarking is considered to be noise free and lossless compression. Also the storage of data done by Dropbox and the transfer of data by it is even more compression free. For the figure below x-axis indicates parameters which were used for the comparision of existing system to the system we have implemented and y-axis specifies the proportion.

CONCLUSION

In this project we have demonstrated through a proposed algorithm that combining encryption and watermarking techniques can provide secure transmission of medical images over vulnerable public networks. The algorithm is based on dividing the image into ROI and RONI regions and embedding three different watermarks in the RONI region. The watermarks were chosen and embedded in such a way to provide image integrity and authenticity, which are the two major requirements for secured medical image transmission. Based on the findings of this work, the proposed algorithm could open up a number of possibilities for the future work. For example, improvement on the quality of the extracted watermark bits can be achieved by applying different error correction schemes such as Hamming codes, turbo codes, Reed Solomon ECC code, and trellis codes. Another enhancement can be achieved by applying reversible watermarking techniques on the ROI region of the image.

FUTURE SCOPE

Future research can be done in two areas. First, in the respect of service similarity, semantic analysis may be performed on the description text of service. In this way, more semantic-similar services may be clustered together, which will increase the coverage of recommendations. Second, with respect to users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). By this means, recommendations can be generated even if there are only few ratings. This will solve the sparsity problem to some extent.

REFERENCES

[1] Ali AI-Haj, Noor Hussein, Gheith Abandah, 2016, Combining Cryptography and Digital Watermarking for Secured Transmission of Medical Images.

[2] S. Nithya, K. Amudha, 2016, Watermarking and Encryption in Medical Image Through Roi-Lossless Compression.

[3] Jaskaran Singh, Anoop Kumar Patel, 2016, An E ective Telemedicine Security Using Wavelet Based Watermarking.

[4] Ali Al-Haj ,Gheith Abandah,Noor Hussein, 2015, Crypto-based algorithms for secured medical image transmission.

[5] Anna Babu, Sonal Ayyappan, 2015, A Reversible Crypto-Watermarking System for Secure Medical Image Transmission.

[6] Quist-Aphetsi Kester, Laurent Nana, Anca Christine Pascu, Sophie Gire, Jojo Eghan, Nii Narku Quaynor, 2015, A Hybrid Image Cryptographic and Spatial DigitalWatermarking Encryption Technique for Security and Authentication of Digital Images.

[7] B.Nassiri, R.Latif, A.Toumanari, F. M. R. Maoulainine, 2012, Secure transmission of medical images by watermarking technique.

SARCASM DETECTION FOR ENGLISH TEXT

Riya Das, Shailey Kadam, Chetan Kalra, Vijeta Nayak and Dr. Sharvari Govilkar
Department of Computer Engineering, Mumbai University, PCE, New Panvel, India

Abstract— Sarcasm determines the mockery or irony used by that person to express his emotions. With the increase in the use of social medias which is mostly in the form of text, it becomes important to detect the sarcasm present in the sentences. So understanding the sentiments of the text becomes very important. In our previous paper* [14] *we proposed a conceptual framework for Sarcasm detection using three machine learning algorithms Viz. Random forest, Naive Bayes, SVM. Our training consists of Twitter dataset with emoticons, punctuations, hashtags and other dataset from different sites. This paper describes the processing steps and the actual workflow and compares the best algorithm among the three algorithms for future work purposes.

***Index Terms* — Hashtags, Punctuation Marks, Emoticons, Random Forest Classifier, SVM, Naive Bayes Classifier**

1. INTRODUCTION

Our objective is to use the concept of machine learning in order to train and test various sentences. Hence, this paper presents a method for detecting sarcasm in given text. Our dataset is a collection of tweets and various reviews with 46,000+ sentences.

Since our project mainly focuses on English text, the most important process is to remove all other mixed languages present in the given statement. This is done by script validation and filtering of pre-processing block. Before training any dataset first step is to clean the noise present in the dataset, which is done by preprocessor block by removing stop words and HTML tags. This cleaned data is then used to train classifiers such as Random Forest, SVM and Naive Bayes. The dataset is divided approximately into 70-30% in order to train and test data to get the desired result. A confusion matrix is then formed which helps us to understand the number of false positives and false negatives during the training part. This paper also deals with comparing these result to find out which classifier gives a better result and accuracy so that the best classifier can be used in social media analytics in order to improve the overall sentiment of these statements. The scope of the system would be to find the Sarcasm present in English Language Only.

The recipients of the system would be organizations which use social media monitoring such as public opinion, reviews and rating of the product which provide valuable information about emerging trends and what consumers and clients think about specific topics, brands or products.and also with the rapid development of craze TV series, use of sarcasm in daily life has become more common and prominent. Besides this, use of Hashtags and emoticons have rapidly been increasing. Therefore, it has become a need of an hour for all these companies to understand the progress of their products in the market and among their clients.

2. LITERATURE SURVEY

As discussed in our previous paper [14] we can conclude that though sarcasm can be determined with a lexicon based approach, but it would take more time for computation. While if we can obtain the features and store it in a file, we can reuse the same featured for determining sarcasm any number of times without actually performing all the processes. Therefore, our project mainly focuses on supervised machine learning approach as it is better to train and store the features, and use them for testing other sentences.

3. SARCASM DETECTOR

In this, we would be discussing about the system architecture. The input of the system would be reviews or simply some content from various Social Media Sites and tweets from twitter, etc.. The first step is to clean the raw input so that a standardized format of content is obtained. From the cleaned data, we have constructed our dataset which is used in training phase to train the various machine learning classifiers.

Few preprocessing of data is done like script validation, removal of URLs and HTML tags. This cleaned data is then converted into standard format i.e data matrix with reviews and labels. labels is of two types 0 and 1 indicating the sentence being not sarcastic and sarcastic respectively.

Training data consists of hashtags, emoticons, punctuation marks and too positive and negative sentences, therefore there is no need to handle them separately. The system uses three supervised machine learning algorithm, such as **Random Forest**, **Support Vector Machine (SVM)** and **Naive Bayes Classifier** to train and test the dataset.

In training phase the algorithm builds a classifier by analysing the training data and associated label with each class and creates a pickle file which consists of all the features extracted by the model in the training phase. From the data model created, a confusion matrix is generated which help us to find the number of true positives, true negatives, false positives and false negatives during the training phase to understand how accurately the data is being trained by each classifier. During the testing phase, the system accepts the input from the user and compares with the features stored in the pickle file and predicts whether the given input sentence is sarcastic or not.

The main aim of the system is to compare these algorithms to find which algorithm can be further used to detect sarcasm during text analytics.

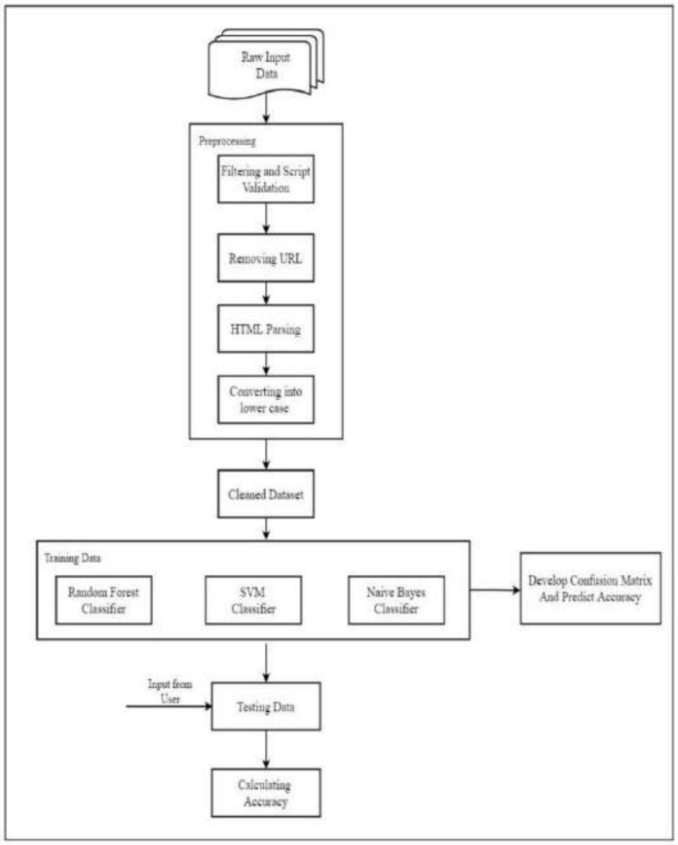


Figure 1 : Sarcasm Detector

3.1 Input Documents

The text will be in Romanized English format. The content would be collected from different social media domains like Twitter or from product based websites like Amazon, etc.

3.2 Preprocessing Block

The process of converting raw input data collected from various social media sites and twitter into standardized format of data matrix i.e label and review.

3.2.1 Filtering and Script Validation

The process of considering only English text by ignoring all the mixed language text so that processing of text can be made easier.

In this step, the given sentence is scanned character by character and compared with UTF-8. If character is present in the given list, then it does not belong to English Script and hence can be ignored.

3.2.2 Removing URLs

The process of removing all unwanted text such as URL so that more informative data can be stored in the dataset for training.

Algorithm :

```
a.  Input : The sentences only containing English Text
    and special characters like hashtags, emojis,
    punctuation marks, etc.
b.  Output : URL present in the sentence are removed.
c.  Steps :
    i.   START.
    ii.  Define a regular expression to identify the
         presence of https://www.abc.com
    iii. Scan the input document.
    iv.  Check for not End of file.
         1. Read a character from input file.
         2. IF character matches with regular
            expression then remove it.
         3. Display the text after removing text
            otherwise go to step 4.
         4.  Read the next input sentence.
         5. STOP.
```

3.2.3 Removing HTML Tags

The process of removing all unwanted text such as HTML tags so that more informative data can be stored in the dataset for training.

Algorithm :

```
a.  Input :  Sentences with no URLS.
b.  Output : Sentences without any HTML tags.
c.  Steps :
    i.   START.
    ii.  Identify all predefined HTML Tags by using
         predefined packages.
    iii. If the sentences contain any html Tags then
         remove it and display it otherwise go to next
         step.
    iv.  Read the next input sentence.
    v.   Presence of HTML tags can be compared by
         comparing the input and output string of this
         block.
    vi.  Repeat the same process until end of document
         is found.
    vii. STOP.
```

3.2.4 Converting into Lower Case

This block converts the input string into one standard format which is in lower case.

3.2.5 Clean Dataset

This block contains dataset free from all unwanted URL, HTML tags and converted into Lower Case.

Stop words are not removed during pre processing as it might contain some sentiments that would affect its meaning. In this blocks labels are assigned to each sentences and are stored into standardized format i.e review and its corresponding label. Labels are in form of 1 and 0 which represent sentences are sarcastic or non - sarcastic respectively.

3.3 Training Classifier

Data Classification is termed as the process that organizes data into categories so that it can be used efficiently and effectively. It basically has two phases :

a. **Training Phase :** At this phase, the classification algorithm uses the training data for analysing.
b. **Testing Phase :** In this phase, testing data are used to estimate the accuracy of the classifier. Testing data is the dataset used for evaluating the model in the training phase.

Based upon the data chunk the dataset is divided for training and testing. Ideally we used 70-30% to train and test data respectively.

3.3.1 Tf-idf

The TF (term frequency) of a word is the frequency of a word (i.e. number of times it appears) in a document.

The IDF (inverse document frequency) of a word is the measure of its importance in the whole corpus.

The formula for to measure Tf-idf is :

$$tfidf(t,d,D)=tf(t,d)*idf(t,D)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.1)$$

Where t denotes the terms; d denotes each document; D denotes the collection of documents.

3.3.2 Random Forest Classifier

Random forest algorithm is one of the supervised learning classification algorithm. This classifier generates large number of decision trees and randomly selects the best node from which features can be extracted and stored.

With increased number of trees for predication will automatically gives higher accuracy results. Hence, of our system we have generated maximum number of trees which help us to extract features for the classifier.

Algorithm for Random Forest can be divided into two phases :

i. Train the Dataset
ii. Random Forest Prediction

Algorithm :

i. Define parameters using TfidfVectorizer.
ii. Train the classifier with the parameters defined.
iii. Make predictions of data from training dataset.
iv. Find accuracy and confusion matrix for training and testing dataset.
v. Plot confusion matrix.

3.3.3 Support Vector Machine

A Support Vector Machine (SVM) is also one of the supervised machine learning algorithm that can be used for both classification and regression purposes. It is mainly used in classification problems.

In this algorithm, each data item is plotted against hyper plane in space with its feature extracted as the value od data item. The data points which are nearest to the defined hyper plane is called as support vectors.

i. CountVectorizer : It converts a collection of text documents to a matrix of token counts. This implementation produces a sparse representation of the counts.
ii. SGDClassifier : SGD stands for Stochastic Gradient Descent where the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing strength schedule.
iii. GridSearchCV : If it is not used we need to loop the parameters and run all the combination of parameters. For this we need to write the code manually which increases the time requirements.

Hence, for our system we have used GridSearchCV.

Algorithm :

i. Defining various parameters using SGDClassifier.
ii. Use GridSearchCV to iterate the parameters automatically.
iii. Train the classifier based upon parameters defined.
iv. Make predictions of data from training dataset.
v. Find accuracy and confusion matrix for training and testing dataset.
vi. Plot confusion matrix.

3.3.4 Naive Bayes Classifier

Naive Bayes Classifier is based on the Bayesian theorem. It is suitable where the dimensionality of the input attributes is high. In this model, parameter estimation is done by using maximum likelihood. It is used to find conditional probabilities.

P(X|Y) is the conditional probability of event X occurring for the event Y which has already been occurred.

$$P(X|Y)=P(X \text{ and } Y)/P(Y)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.2)$$

a. MutinomialNB Classifier : For our system we have implemented MultinomailNB which makes use of the Naive Bayes algorithm for multinomially distributed data. The parameters is estimated by a smoothed version of maximum likelihood, i.e. relative frequency counting.

Algorithm :

i. Define parameters using TfidfVectorizer and MultinomialNB.
ii. Training the classifier with the parameters defined.
iii. Make predictions of data from training dataset.
iv. Find accuracy and confusion matrix for training and testing dataset.
v. Plot confusion matrix.

4. Result Analysis

Training dataset is generated by cleaning the raw data collected from various social media sites like Amazon, Facebook, etc. and tweets from twitter. For the evaluation of our system, we have used 10,000 sentences of each type for each classifier model. The system extracts the features from the input sentence and compare it with the features stored in pickle file to detect whether the given input sentence is sarcastic or not.

Example 1 : Apparently I was not supposed to be happy :unamused_ face:
Random Forest : Yes
SVM : Yes
Naive Bayes Classifier : Yes
Expected Outcome : Sarcastic

Example 2 : I am going to take a leave from office today.
Random Forest : No
SVM : No
Naive Bayes Classifier : No
Expected Outcome : Non - Sarcastic

Example 3 : Whatever it is that is eating you, it must be suffering horribly.
Random Forest : No
SVM : No
Naive Bayes Classifier : No
Expected Outcome : Sarcastic

The efficiency of our system is based on the confusion matrix generated after training the classifier and number of correct output given by each classifier for input sentence during testing.

The following graph shows the accuracy obtained by the system during training phase.

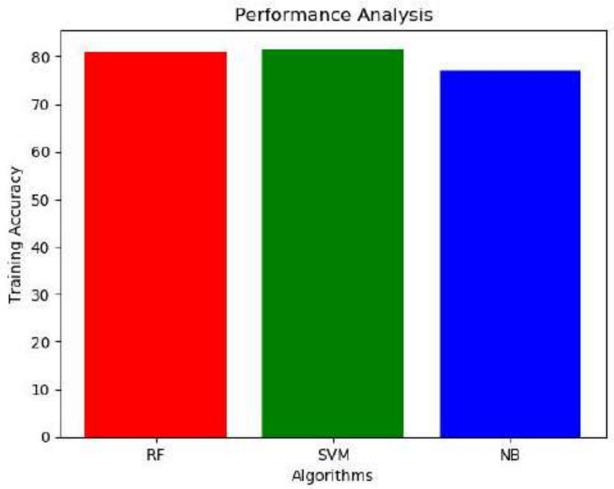


Figure 2 : Performance Analysis for Training Phase

Therefore the graph below helps us to compare which algorithm is best to classify the sentences into sarcastic and non-sarcastic respectively.

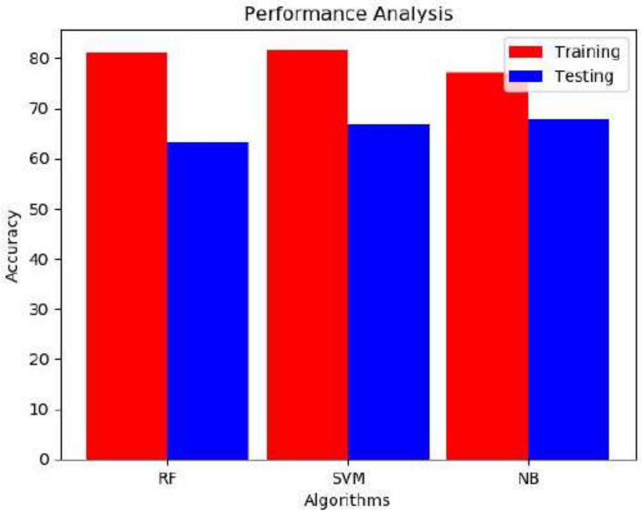


Figure 3 : Accuracy Comparison

5. Conclusion

Every algorithm has its own advantages and completely different process to identify patterns. The training accuracy obtained after training the three classifier is as :

Table 1 : Training Accuracy

Algorithms	Accuracy
Random Forest	81%
SVM	81.54%
Naive Bayes	76.99%

While testing accuracy obtained after evaluating 10,000 dataset of each is as :

Table 2 : Testing Accuracy

Algorithms	Accuracy
Random Forest	63.09%
SVM	66.74%
Naive Bayes	67.81%

The Naive Bayes algorithm performed better than the other two algorithm performed for identifying similarities between non sarcastic and sarcastic sentences respectively whereas by using Support Vector Machine, system has a slight edge for extracting sarcastic patterns.

Our System compares the best machine learning algorithm from the three algorithms viz. Naive Bayes, Random Forest and SVM to detect sarcasm present in the given text. It gives us the desired output from the features obtained during the training phase. But due to false positives and false negatives obtained while training, sometimes, this system predicts a wrong output. But this can be further improved by using deep learning techniques like Keras and Tensor-flow. Classifiers can be made more powerful by training more amount of dataset with emoticons which might increase the accuracy of the classifier.

Acknowledgment

We would like to thank Dr. Madhumita Chatterjee, Head of Computer Engineering Department for the invaluable support. We would also like to show our gratitude towards Dr. Sandeep M. Joshi, Principal, PCE, New Panvel for his invaluable support and for providing an outstanding academic environment. Last but not the least we would like to thank Prof. Dhiraj Amin and all other staff members of the Department of Computer Engineering for their critical advice and guidance without which this project would not have been possible. Also, we would like to say that it has indeed been a fulfilling experience for working out this project topic.

REFERENCES

[1] Whiting A and D Williams. Why people use social media: a uses and gratications approach. Qualitative Market Research: An International Journal, 2013

[2] Ilia Vovsha Owen Rambow Apoorv Agarwal, Boyi Xie and Rebecca Passonneau. Sentiment analysis of twitter data. In Proceedings of the ACL 2011 Workshop on Languages in Social Media, pages 30-38, 2011.

[3] Luciano Barbosa and Junlan Feng. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of COLING, pages 36-44, 2010.

[4] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. 2010.

[5] Bhyani R. Go, A. and L Huang. Twitter sentiment classification using distant supervision. Technical report, CS224N Project Report, Stanford, 2009.

[6] Daniel Neagu Haruna Isah, Paul Trundle. Social media analysis for product safety using text mining and sentiment analysis. 2015.

[7] Yulan He Hassan Saif and Harith Alani. Semantic sentiment analysis of twitter. 2011.

[8] P. Anderson J. Blackburn C. Borcea N. Kourtellis, J. Finnis and A. Iamnitchi. Prometheus user-controlled p2p social data management for socially aware applications. 2010.

[9] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2008.

[10] Ashwin Rajadesingan, Reza Zafarani Arizona, and Huan Liu Arizona. Sarcasm detection on twitter: A behavioral modeling approach. 2015.

[11] Hiroshi Shimodaira. Text classification using naive bayes. 2015.

[12] https://github.com/AniSkywalker: -Dataset

[13] http://scikit-learn.org/stable/

[14] Conceptual Framework For Sarcasm Detection for English Text - Riya Das, Shailey Kadam, Chetan Kalra and Vijeta Nayak (Department of Computer Engineering, Mumbai University, PCE, New Panvel, India).

Secure VPN Server Deployed on Raspberry Pi

Pooja Karan Bist, Akansha Santosh Mekade, Anurag Mohan Nair and Dr. Madhumita Chatterjee,
(Pillai College of Engineering, New Panvel)

*Abstract***—** ***With the increase in data accumulation, manipulation and the need for remote access, there is also a need for a secure network route or protocol through which users can access their data stored at a location far away from their current location. VPN Server is one of the most prominent and widely used network configuration aimed at supplementing the demand of remote access. The proposed system is focused on setting up a VPN server and securing the connection between the VPN host machine and the client that is accessing it remotely. The current VPN system though has security protocols deployed on it, it fails to comprehend the more advanced and complex threats to the system. The project aims at providing multiple layers of protection in the form of authentication during the connection establishment between a vpn client and the vpn server deployed on raspberry pi. The idea is to incorporate three layers of verification into the vpn authentication mechanism and eliminate any and all flaws that maybe present during the connection stages. The multiple layers includes the different modules and mechanisms like Pluggable Authentication Module (PAM), Client Specific Authentication (Private Key), Lightweight Directory Access Protocol (LDAP). In addition to deploying an advanced security mechanism, the project also focuses on converting the client machine into a mobile hotspot which will in turn, act as a Wi-Fi sources for other Wi-Fi enabled devices in the proximity, thus, extending the VPN connection to all devices and not just your desktop pc or laptop. Finally, to make this structure portable, the whole project is deployed on a Raspberry Pi environment. This enables the system to become extremely portable, reusable and user friendly, thus allowing the VPN to be set up whenever and wherever required. A user-end GUI implementation is the last stage of the proposed system where a simple and user-friendly GUI is designed to enable the user to navigate through the different actions possible on the VPN server.***

Index Terms— **Raspberry Pi; VPN (Virtual Private Network); OpenVPN; PAM (Pluggable Authentication Module); Client Specific; Mobile hotspot.**

I. Introduction

The constant and ever increasing need for remote access spotlighted the emerging era of VPN- a virtual remote networking module. Using VPN, users are not only able to access their data remotely but also in a secured way through a private virtual tunnel. But again the question arises, is the existing VPN system fully secured? Although the existing system does not have any fundamental flaws, there are few minor threats and vulnerabilities that can lead to unauthorized access to the server. As a consequence, it is equally important to deal with this minor threats and vulnerabilities in order to guarantee the user their privacy.

The proposed system gears up some additional features that resolves the minor threats and vulnerabilities with existing system. These additional features are the Raspberry Pi - an all time active device and a low power consumer; A Multi-tier Authentication Module - a high level authentication assurance; A Hotspot Module - VPN connection extender.

II. Literature review

Aparicio Carranza and Constadinos Lales, [1] give the theory of how data is insecure while accessing the public internet and how one can use the Raspberry Pi (A cheap microcomputer) as a VPN server to a home network; in order to create a VPN connection between a home network and the public internet.

Thomas Berger analyzed the current VPN technologies, [2], such as Internet Protocol Security (IPSec), Layer Two Tunneling Protocol (L2TP), and Point to Point Tunneling Protocol (PPTP). The analysis includes one significant drawback which concerns all tested technologies - the dramatic loss of performance and throughput. IPSec suffers from complex tunnel negotiation process, L2TP, when combined with IPSec, results in excessive data overhead whereas for PPTP, it's security level is not sufficient for critical applications. Hence, to enhance the security and reliability of a VPN, a strong authentication mechanism is required on top of the traditional username and password authentication credentials.

Anupriya Shrivastava, M.A. Rizvi proposed the concept of external authentication approach for VPN using LDAP protocol [3]. The advantage of this approach is that user information is stored in a dedicated authentication server which can have a large pool of organized, directory-based user data along with greater robustness and security. Hence this approach proposes to extend the functionality of LDAP server in order to strengthen the authentication process of VPN.

L. Caldas-Calle, J. Jara Member, M. Huerta and P. Gallegos [4] surveyed that the highest throughput is for RP3 and the lowest for RPB. Values indicate dependence latency buffer each model based on the average time RTT, it is more evident in RPZ and RPB. Packages with size beyond the fragmentation point suffer QoS decrease, due the need to fragment packets. The CPU power of each Raspberry Pi model is an important factor affecting the QoS parameters of a wireless VPN. Introducing VPN to secure communication implies more complex process in communication that requires more from hardware.

III. Existing system

In an existing VPN system, when a client requests for a connection the initial step taken is to match the certificate files. These certificate files contain the private key and the Signature Encryption Algorithm used to authenticate the client to the server. If the attacker is able to get this client file he can easily break into your private network and this can be a loophole for the existing VPN systems.

Where the VPN is used to protect your data transmission over the internet, there's a security protocol namely LDAP which is used for authentication purpose while accessing the directory files. Fundamentally, LDAP using operations such as "Bind" operation authenticates the user, willing to access the directory, through the username and password included in the Bind operation.

IV. Proposed system

The proposed system leaves the basic functioning of the VPN server untouched and adds on an extra feature for better usability and security. This extra layer of security is provided by a Multi-tier Authentication Module.

The Multi-tier Authentication Module provides three tiers of authentication.

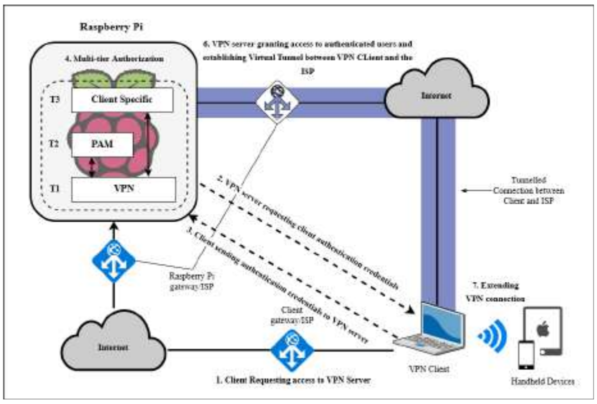

Raspberry Pi
4. Multi-tier Authorization
T3
Client Specific
T2
PAM
T1
VPN
6. VPN server granting access to authenticated users and establishing Virtual Tunnel between VPN CLient and the ISP
Internet
Tunnelled Connection between Client and ISP
2. VPN server requesting client authentication credentials
3. Client sending authentication credentials to VPN server
Raspberry Pi gateway/ISP
Client gateway/ISP
7. Extending VPN connection
Internet
1. Client Requesting access to VPN Server
VPN Client
Handheld Devices

Fig. 4.2. Proposed Architecture.

Tier 1: This tier comprises the basic functionalities of the VPN that includes authentication of client files.

Tier 2: This tier incorporates a PAM module - Pluggable Authentication Module, that uses low level authentication mechanisms to integrate different modules and use one simple authentication for all of them. This authentication is provided at the server side. PAM provides the same level of security as LDAP but in a more optimum way. Additionally, it also allocates a dedicated desktop for each client.

Tier 3: This tier generates a Client Specific Private Key that eliminates the possibility of multiple users using one client file to log in to the VPN server.

The connection once established, on the client machine, can be extended to other handheld devices using the Wi-Fi hotspot that is created on the client machine.

The user end machine has a simple GUI to operate and maneuver through the operations of the VPN server and its functionalities. The GUI contains three buttons: one to connect to the VPN server following up the entire authentication process, second button allows user to upload a file from the VPN server and the third button is used to download the shared files from the VPN server.

V. Implementation model

Implementing the proposed system focuses on collaborating different modules that works individually to function in a multi-tier architecture that ensure higher level of authentication for a VPN deployed on a Raspberry Pi. The modules incorporated in the system are:

PAM- Pluggable Authentication Module*:* PAM is widely used for authenticating users against a system that has accounts created on it. Each PAM authentication is done by comparing whether the entered credentials belong to a user account in the system OS. If so, access is granted. The PAM module also, by default, indicates that each user will have their own desktop on the server system. It is useful for a simple user credential authentication.

Client Specific Secret Key Authentication*:* This authentication layer lies above the PAM module. The client must first clear the traditional authentication after which he encounters the PAM authentication and finally after that, he will be reaching this last layer of authentication which is encoded to the client file and a secret key is attached. The key is generated at the time of '.ovpn' file creation on the server side.

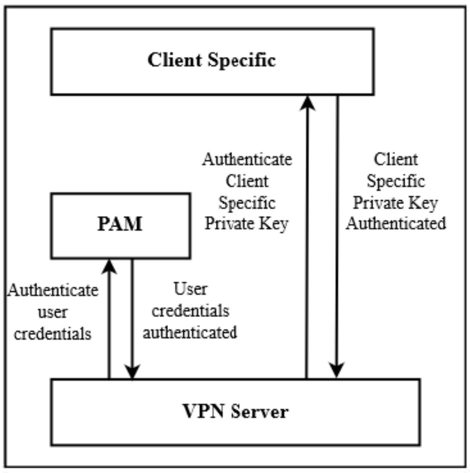

Client Specific
PAM
Authenticate Client Specific Private Key
Client Specific Private Key Authenticated
Authenticate user credentials
User credentials authenticated
VPN Server

Fig. 5.1. Multi-tier Authentication Module.

Shared Folder*:* There exists a shared folder between the VPN server and the client that is going to connect to it. This folder is hidden behind the different layers of authentication and can only be accessed once the VPN connection is established. For better understanding, consider a scenario:

Scenario: A client wishes to upload/download a file from the VPN server. For this, the client will have to follow the following sequence of actions:

1. Request to server for VPN access.

2. Pass through multiple layer of authentication

3. Establish a secure connection between the server and itself.

4. Then access the shared folder to perform desired file transfer.

Hence, even to share a file, the user must pass through the multi-tier architecture and connect to the VPN first.

Hotspot: Extending the VPN connection was our final step in the implementation module. The client has established its own secure connection with the VPN server and now wishes to connection their handheld devices to the same network. For this, we have implemented a hotspot module that will allow the user to extend the connection to nearby devices using SSID-Password method.

***ALGORITHM* :**

Here :

RPI	: Raspberry Pi
$Client$	: VPN Client
VPN_S	: VPN Server
PAM	: Pluggable Auth. Module
CSM	: Client Specific Auth. Module
VPN_H	: VPN Hotspot
$Client_F$	: Client File
CA	: Certificate Authority
$Cert$	: Client Certificate
SK	: Secret Key
PK	: Private Key
H_{Device}	: Handheld Devices
$H_{credentials}$	: Hotspot Credentials

Raspberry Pi consists of three authentication modules:

RPI [VPN_S, PAM, CSM]

Step 1 *:* Client requesting VPN_S for Access by Sending its $Client_F$.

Client ⟶ VPN_S : $Client_F$ [CA, Cert,{SK}]

Step 2 *:* VPN authenticates the File received as an access request by the Client.

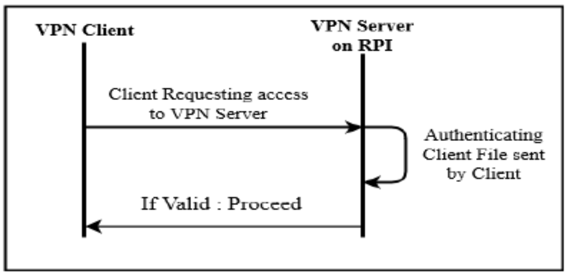

VPN Client
VPN Server on RPI
Client Requesting access to VPN Server
Authenticating Client File sent by Client
If Valid : Proceed

Fig. 5.2. Phase I.

***Step** 3*: If $\mathtt{Client_F = Valid}$; VPN_S requests client to provide the PAM credentials.

***Step** 4:* Client sends the PAM credentials that consist of a username and a password to PAM module.

$$\mathtt{Client} \longrightarrow \mathtt{PAM : Credentials\ [\ username\ ,\ password\ ]}$$

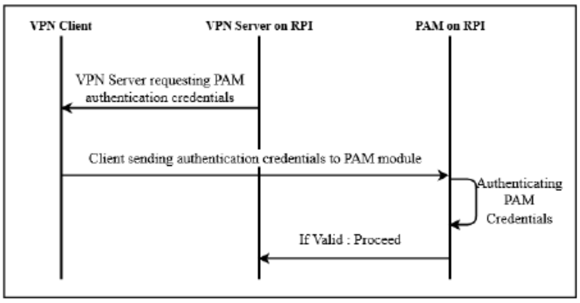

VPN Client
VPN Server on RPI
PAM on RPI
VPN Server requesting PAM authentication credentials
Client sending authentication credentials to PAM module
Authenticating PAM Credentials
If Valid : Proceed

Fig. 5.3. Phase II.

***Step** 5*: If $\mathtt{Credentials = Valid}$; VPN_S r requests client to provide the Client Specific Private Key.

***Step** 6:* Client sends the Client Specific Private Key to CSM.

$$\mathtt{Client} \longrightarrow \mathtt{CSM :\ [\ \{PK\}\ ]}$$

***Step** 7*: If $\mathtt{\{PK\} = Valid}$; the CSM permits VPN_S to grant access to client.

***Step** 8*: The VPN grants access to the Client and a Virtual Tunnel is established through ISP.

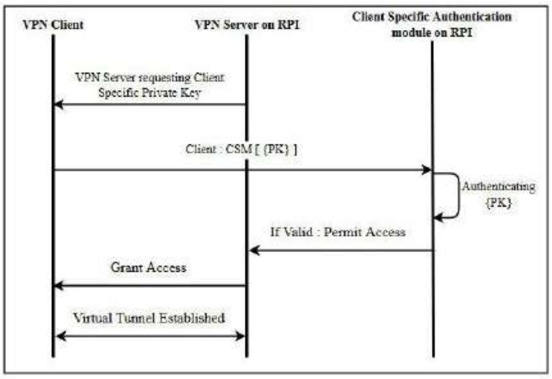

VPN Client
VPN Server on RPI
Client Specific Authentication module on RPI
VPN Server requesting Client Specific Private Key
Client : CSM [{PK}]
Authenticating {PK}
If Valid : Permit Access
Grant Access
Virtual Tunnel Established

Fig. 5.4. Phase III.

***Step** 9*: Once the VPN connection is established, the Client turns itself into VPN_H, in order to extend the VPN connection to $Handheld_D$

$$\mathtt{Client\ :=\ VPN_H}$$

***Step** 10:* To connect to the VPN_H, H_{Device} provides it's $H_{credentials}$ that consist of SSID and password to the VPN_H.

$$\mathtt{H_{Device}} \longrightarrow \mathtt{VPN_H\ :\ H_{credentials}[SSID,password\ ]}$$

***Step** 11*: If $\mathtt{H_{credentials} = Valid}$; the VPN connection is extended to $H_{Device.}$

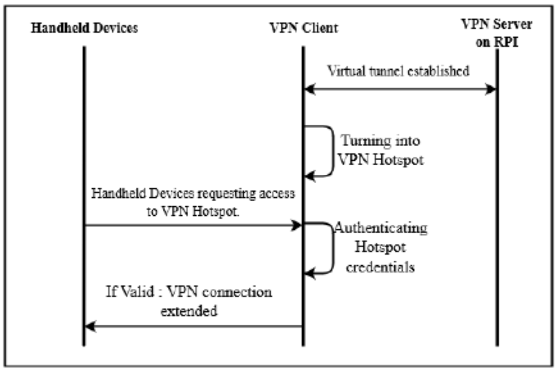
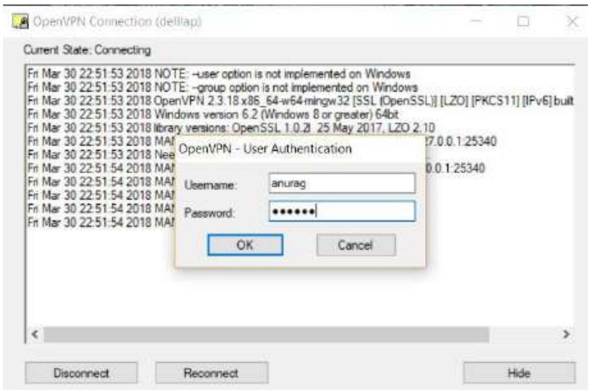

Handheld Devices
VPN Client
VPN Server on RPI
Virtual tunnel established
Turning into VPN Hotspot
Handheld Devices requesting access to VPN Hotspot
Authenticating Hotspot credentials
If Valid : VPN connection extended

Fig. 5.5. Phase IV.

The result of implementation of the individual modules was a sophisticated and secure 3-tier authentication system that enables user to connect securely to the VPN server.

Output Screenshots:



Current State: Connecting

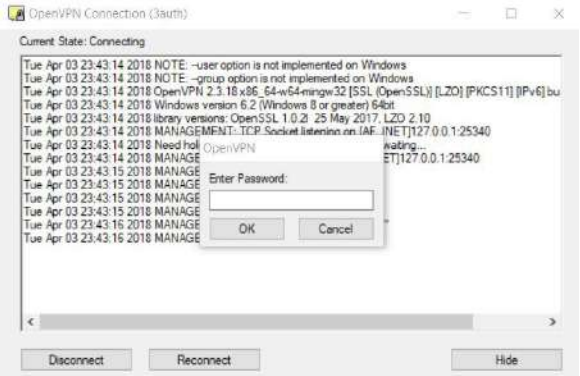
```
Fri Mar 30 22:51:53 2018 NOTE: --user option is not implemented on Windows
Fri Mar 30 22:51:53 2018 NOTE: --group option is not implemented on Windows
Fri Mar 30 22:51:53 2018 OpenVPN 2.3.18 x86_64-w64-mingw32 [SSL (OpenSSL)] [LZO] [PKCS11] [IPv6] built
Fri Mar 30 22:51:53 2018 Windows version 6.2 (Windows 8 or greater) 64bit
Fri Mar 30 22:51:53 2018 library versions: OpenSSL 1.0.2l  25 May 2017, LZO 2.10
Fri Mar 30 22:51:53 2018 MAN                                              27.0.0.1:25340
Fri Mar 30 22:51:53 2018 Nee
Fri Mar 30 22:51:54 2018 MAN                                              0.0.1:25340
Fri Mar 30 22:51:54 2018 MAN
Fri Mar 30 22:51:54 2018 MAN
Fri Mar 30 22:51:54 2018 MAN
Fri Mar 30 22:51:54 2018 MAN
```

OpenVPN - User Authentication

Username: anurag

Password: ••••••

Fig.5.5. Post file verification, PAM authentication prompt.

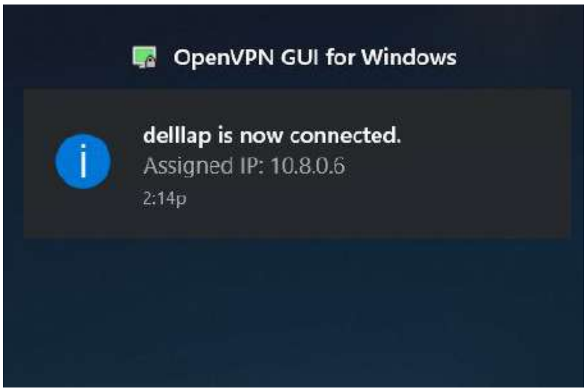


Current State: Connecting

```
Tue Apr 03 23:43:14 2018 NOTE: --user option is not implemented on Windows
Tue Apr 03 23:43:14 2018 NOTE: --group option is not implemented on Windows
Tue Apr 03 23:43:14 2018 OpenVPN 2.3.18 x86_64-w64-mingw32 [SSL (OpenSSL)] [LZO] [PKCS11] [IPv6] bu
Tue Apr 03 23:43:14 2018 Windows version 6.2 (Windows 8 or greater) 64bit
Tue Apr 03 23:43:14 2018 library versions: OpenSSL 1.0.2l  25 May 2017, LZO 2.10
Tue Apr 03 23:43:14 2018 MANAGEMENT: TCP Socket listening on [AF_INET]127.0.0.1:25340
Tue Apr 03 23:43:14 2018 Need hol                                   waiting...
Tue Apr 03 23:43:14 2018 MANAGE                                     ET]127.0.0.1:25340
Tue Apr 03 23:43:15 2018 MANAGE
Tue Apr 03 23:43:15 2018 MANAGE
Tue Apr 03 23:43:15 2018 MANAGE
Tue Apr 03 23:43:15 2018 MANAGE
Tue Apr 03 23:43:16 2018 MANAGE
Tue Apr 03 23:43:16 2018 MANAGE
```

Enter Password:

Fig 5.6. Post PAM, client specific secret key prompt.



OpenVPN GUI for Windows

delllap is now connected.
Assigned IP: 10.8.0.6
2:14p

Fig 5.7. Post 3-tier auth. VPN connection established.

VI. Applications

- ***When you want to share file remotely and securely:*** The security of the files while sharing it on an open network is always at risk. Also a question arises of access the files when you are away from home. The proposed system resolves these issues, since the VPN enables the user to access their files remotely and through a secured virtual tunnel.

- ***When You Want Privacy and Advocacy:*** When you are away from home and you wish to connect to a secure network, VPN is the way to go about it. It not only strengthens your connection but also increase privacy and advocacy by encrypting all the data transferred whether at home or abroad.

- ***When you want a secure wifi-connect:*** Once connected to the system, the client can extend its secure VPN connection to the nearby devices by converting itself into a wifi hotspot. This enables the non-client devices to use the secure network of the VPN.

VII. Conclusion

The VPN server deployed on Raspberry Pi enables user to access their virtual connection to home network at any time and on a low power consumption. The multi-tier authentication assures the security of connection establishment between the server and the client, while the hotspot module extends the VPN connection to the handheld devices.

VIII. References

[1] Constadinos Lales Aparicio Carranza. Using the raspberry pi to establish a virtual private network (vpn) connection to a home network. International Conference on Portable devices, 2014.

[2] Thomas Berger. Analysis of current vpn technology. First International Conference on Availability, Reliability and Security, 2012.

[3] M.A. Rizvi Anupriya Shrivastava. External authentication approach for virtual private network using ldap. First International Conference on Networks and Soft Computing, 2014.

[4] M.Huerta L.Caldas-Calle, J.Jara Member and P.Gallegos. Qos evaluation of vpn in a raspberry pi devices over wireless network. International Caribbean Conference on Devices, Circuits and Systems, 2017.

[5] Use PAM to Configure Authentication. https://www.digitalocean.com/community/tutorials/how-to-use-pam-to-configure-authentication-on-an-ubuntu-12-04-vps. October 3, 2013.

Auto Source Code Generator For C Code

*

Jahnvi Patil
BE Computer
Pillai College of Engineering
Navi Mumbai, India
jahnvi.p0910@gmail.com

Sohail Siddique
BE Computer
Pillai College of Engineering
Navi Mumbai, India
sohailsiddique700@gmail.com

Milind Patel
BE Computer
Pillai College of Engineering
Navi Mumbai, India
patelmilind123@gmail.com

Sanket Oswal
BE Computer
Pillai College of Engineering
Navi Mumbai, India
sanketoswal944@gmail.com

***Abstract*—Sometimes being a programmer is tough. In fact, most of the time, it's easy to get bogged down in syntactical or platform specific details and lose sight of the big picture. This is where automatic Source code generation comes in. Auto Source Code Generator (ASCG) refers to using programs to generate code that the developer would otherwise have to write. As an added bonus, using ASCG creates consistent, codes more productively.**

People may possess good logical skills along with great algorithmic solution designing capabilities but the inadequate knowledge of programming languages makes them handicapped. Effective conversion of algorithms mentioned as pseudo code to C Code will enable programmers to focus on logic building and confine them from syntactical errors. An algorithm to program converter is an interpreter that is capable of converting algorithms in pseudocode (with fixed input format) to C code whose flexibility of interpretation has been enhanced by using synonyms and by the introduction of a personalized training model.

Keywords: Pseudo Code, POS Tagging, Pseudo Code Processing, Source Code

I. INTRODUCTION

There are many existing systems that work for converting pseudo code to source code but have drawbacks as they do not work for loops and functions, and also error handling. We hereby are making a system overcoming the mentioned issues. Automatic Source Code Generator for C Language (ASCG) refers to using programs to generate code that the developer would otherwise have to write. As an added bonus, using ASCG creates consistent, codes more productively and at a higher level of abstraction than manually coding projects.

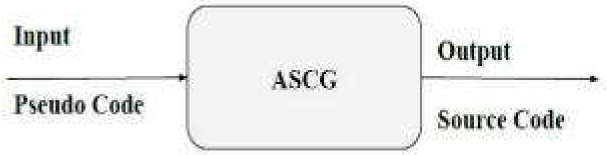


Fig. 1. Basic Working of System.

The Figure 1. shows the method of working of the system that provides a C Code as output generated. Here, the input is provided in natural language(as pseudo code) will be given to the ASCG system which comprises of various modules which are discussed further along with techniques and hence will generate a source code in C language.

II. LITERATURE SURVEY

There is a common factor between Natural Language Processing and Programming languages which is Language. Natural Language Processing and Programming languages are very important domains in computer science but very less importance has been given to the interaction in between these two fields. Previously study has been done to develop interpreter which convert algorithm in natural language to the programming language source code. But each of such is having certain limitations. Examples of such interpreter are An efficient Approach to Produce Source Code Interpreting Algorithm, CodGen and Semi Natural Language Algorithm to Programming Language Interpreter.

- **An efficient Approach to Produce Source Code Interpreting Algorithm**

Here, the first proposal was An efficient Approach to Produce Source Code Interpreting Algorithm [1], an algorithm to program converter is an interpreter that is capable of converting algorithm in English to C, C++ and Java code whose flexibility of interpretation has been enhanced by using synonyms and by the introduction of a personalized training model. Semantics of the algorithm as a whole becomes difficult to interpret and process. Use of functions, arrays, declarations and pointers. It has fixed input format.

- **CodGen**

The second proposal Design and Implementation of CodGen Using NLP [2], that discusses a software that uses NLP and text mining technique. In this algorithm conversion uses various methods such as : Splitting algorithm, Variable extraction, Assign data type to each variable, Declare the variables in c file, Attaching main() to C file and also the header file. It produces output for only C language. It has fixed input format. Semantics of algorithm becomes difficult to interpret.

- **Semi Natural Language Algorithm to Programming Language Interpreter**

The third proposal was Semi Natural Language Algorithm to Programming Language Interpreter [1]. This translator converts algorithm in natural English language to code in C and Java This interpreter has many semantic challenges such as it does not support multiple variable declaration, it also does not support printing the value of variables. Such limitations imposes constraint on user while developing fully functional program.

III. PROPOSED SYSTEM

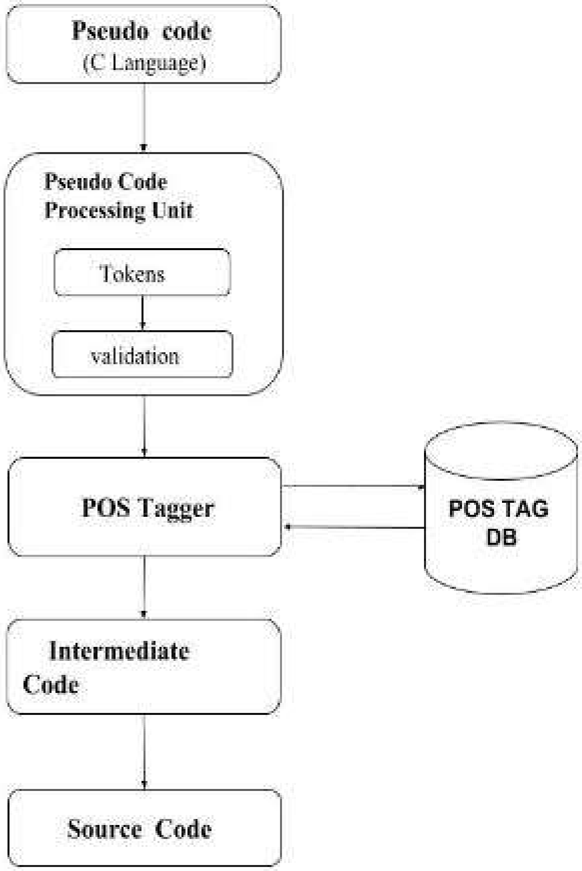


Fig. 2. Proposed System Architecture

The Proposed system architecture consists of several modules interacting with each other to accept an algorithm in natural language as pseudo code and interpret it in C Programming language.
The modules are as follows :

1) **Pseudo code [user module] :** This module indicates the end user. The Pseudo code is accepted into the system, via a desktop application which will be stored in the file. After accepting pseudo code, the file will processed by the other modules.

2) **Pseudo code Processing Unit :** After accepting the Pseudo code from the user, lexical analysis of the file will be done which will result in making the lexemes [tokens] out of the Pseudo code. While analysis if variable is extracted then it will be stored in another file with their respective data type and if data type is not mentioned in the pseudo code by default our system will assign Integer as a data type.

3) **POS Tagger :** Once tokens are generated we apply POS [Part-of-Speech] tagging technique on each token using our POS tag database. POS tag DB will contain tags, from which our system will extract an equivalent tag defining the token and assign the tag to the token. It will be repeated till all the tokens are assigned a tag.

4) **Intermediate code :** After applying POS tagging the file now contains tokens with respect to its tags. With the help of intermediate file, system will replace respective tags with targeted languages keyword and by appending the variables which we have extracted at beginning of the file inside the targeted language header will give successfully converted source code.

5) **Final Source Code :** It provides C source code attaching header file, main(), along with the intermediate file generated.

IV. CONCLUSION AND FUTURE SCOPE

The system consists of Pseudo Code, Pseudo Code Processing Unit, POS Tagger, Intermediate File Generation and Final source Code module which interact to form a formal code. ASCG is a system that is capable of converting a pseudocode (input) to C Code. Effective conversion of pseudo code to C code will enable programmers to focus on logic building and confine them from syntax worries. Although beneficial, implementation of such converter encounters numerous challenges like demarcation entailed due to semantics of the English language since only pseudo codes in a different format is used. We have opened promising results using our current model and we plan to extend it and incorporate functions, pointers and input in English Language. This part can be covered by creating further modules with associated triggers and logic for the same. Further, we aim to overcome the challenge related to use of English Language and its semantics as our future scope.

V. ACKNOWLEDGEMENT

It is a great pleasure and moment of immense satisfaction for us to express my profound gratitude to our dissertation

Project Guide, Prof. Shubhangi Chavan whose constant encouragement enable us to work enthusiastically. Her perpetual motivation, patience and excellent expertise in discussion during progress of the dissertation work has benefited us to an extent, which is beyond expression. We are highly indebted to her for her invaluable guidance and ever-ready support in the successful completion of this dissertation in time. Working under her guidance has been fruitful and unforgettable experience. Despite of her busy schedule, she was always available to give us advice, support and guidance during the entire period of my project. The completion of this project would not have been possible without her encouragement, patient guidance and constant support.

We are thankful to Dr. Madhumita Chatterjee, Head of Computer Engineering Department for her guidance, encouragement and support during our project and also to Dr. Sandeep M. Joshi, Principal, PCE, New Panvel for his encouragement and for providing an outstanding academic environment, also for providing the adequate facilities.

At last we must express our sincere heartfelt gratitude to Prof. Manjusha Deshmukh and Prof. Gaurav Sharma (Project Coordinators) whose invaluable guidance supported me in completing this project and all the staff members of Computer Engineering Department who helped me directly or indirectly during this course of work. We would like to say that it has indeed been a fulfilling experience working on this project.

VI. REFERENCE

1) Priyanka Motkari, Bhagyashree Wable, Supriya Walzade, Pooja Velhal, An efficient Approach to Produce Source Code Interpreting Algorithm, in International Research Journal of Engineering and Technology (IRJET) ,2017 on, Vol: 04 Issue:03, page 2803-2806, March-2017
2) Priyanka , Priyanka HL, Priyanka P , Ruchika, Naveen Chandra Gowda, Design and Implementation of Cod-Gen Using NLP, in Asian Journal of Engineering and Technology Innovation (AJETI), 2017, page 11-14, 2017
3) Sharvari Nadkarni , Parth Panchmatia, Tejas Karwa, Prof. Swapnali Kurhade, Semi Natural Language Algorithm to Programming Language Interpreter, in International Conference on Advances in Human Machine Interaction (HMI - 2016), March-03-05,2016
4) Amal M R, Jamsheed C V and Linda Sara Mathew, PseudoCode to Source Programming Language Translator, in International Journal of Computational Science and Information Technology (IJCSITY), Vol.4,No.2,May 2016
5) Vishal Parekh, Dwivedi Nilesh, Pseudo Code to Source Code Translation, in Journal of Emerging Technologies and Innovative Research (JETIR), Volume 3, Issue 11, November 2016
6) https://www.youtube.com/watchv
7) https://stackoverflow.com/questions/23984614/problems-importing-ttk-from-tkinter-in-python-2-7
8) Slav Petrov, Dipanjan Das, Ryan McDonald, A Universal Part-of-Speech Tagset in Google Research, New York, NY, USA, slav, ryanmcd@google.com , Carnegie Mellon University, Pittsburgh, PA, USA, dipanjan@cs.cmu.edu, 2015.
9) https://pythonprogramming.net/python-3-tkinter-basics-tutorial/
10) http://language.worldofcomputing.net/pos-tagging/parts-of-speech-tagging.html

Code Based Neighbour Discovery Protocol In Wireless Mobile Networks

Payel Thakur
Assistant Professor
Pillai College of Engineering, New Panvel
Department of Computer Engineering email -
payelthakur@mes.ac.in

Sirish Gopalan
Pillai College of Engineering, New Panvel
Department of Computer Engineering email -
gopalansirish@gmail.com

Sumesh Nambiar
Pillai College of Engineering, New
Panvel Department of Computer
Engineering email -
sumesh0395@gmail.com

Vinay Ramesh
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - vinayr1996@gmail.com

Nikhil Haridasan
Pillai College of Engineering, New Panvel
Department of Computer Engineering
email - nikhilharidasan36@gmail.com

Abstract—**Generally routing protocol is defined as a set of rules which regulates the transmission of packets from source to destination. These characteristics are maintained by different routing protocols[1]. In MANET different types of protocols are used to find the shortest path, status of the node, energy condition of the node. In mobile wireless networks, the rising closeness based applications have prompted the requirement for exceedingly compelling also, vitality proficient neighbor discovery protocols. In any case, existing works can't understand the ideal most exceedingly bad case latency in the symmetric case, and their exhibitions with asymmetric duty cycles can even now be moved forward. In this paper, we explore nonconcurrent neighbor discovery through a code- based approach, counting the symmetric and asymmetric cases. We infer the tight most pessimistic scenario latency bound on account of symmetric duty cycle. We plan a novel class of symmetric examples called Diff-Codes, which is ideal when the Diff-Code can be stretched out from a great distinction set. We additionally consider the asymmetric case and outline ADiff-Codes. To assess (A)Diff-Codes, we direct both recreations and test bed tests. Both reenactment and test comes about demonstrate that (A)Diff-Codes essentially beat existing neighbor revelation conventions in both the middle case what's more, thinking pessimistically. In particular, in the symmetric case, the most extreme most pessimistic scenario change is up to half; in both symmetric and asymmetric cases, the middle case pick up is as high as 30%.**

Keywords—ADiff-codes, Manet , Diff-codes ;

I. Introduction

Data Transfer in Mobile Ad-hoc Networks

1.1 Fundamentals

A Mobile Ad-hoc Network is an anthology of autonomous mobile nodes that can communicate with each other through radio waves. A Mobile Ad-hoc Network has many free or autonomous nodes often unruffled of mobile devices or other mobile pieces that can organize themselves in various ways and operate without strict top-down network administration. A mobile ad-hoc network (MANET) is a network of mobile routers coupled by wireless links - the union of which forms a casual topology. The routers are free to move indiscriminately and organize themselves in unsystematic manner so the network's wireless topology may perhaps change hastily and indeterminable. In MANET the concert of the network is based on nodes uniqueness like effectiveness, energy efficiency, transmission speed etc., the concert of the network is high if the nodes in the network satisfy the distinctiveness. MANET characteristics: MANET network has an autonomous behavior where each node presents in the network; act as both host and router. During the transmission of data if the destination node is out of range then it posses the multi-hop routing[2]. Operation performed in Manet network is distributed operation. Here the nodes can join or leave the network at any time. Topology used in MANET network is dynamic topology.

Central servers can be engaged, proximity-based applications, potential can be better demoralized providing the capability of discovering close by mobile devices in wireless communication locality due to some reasons like users can enjoy the ease of local neighbor discovery at any occasion, although the federal service may be occupied due to unexpected reasons, a single neighbor discovery protocol can advantage various applications by providing more litheness than the centralized approach[3].

1.2 Objectives

The objective of this work is as follows:

1. To study and design a neighbor discovery system that desires to have the minimum possibility of collisions.
2. To simulate the data transfer using the Diff codes and A-diff codes.
3. To evaluate the performance of our designs in one-to- one and group scenarios, not only conduct comprehensive simulations, but also sampling them using testbed

1.3 Scope

Albeit central servers can be utilized, proximity based applications; potential can be better abused giving the capacity of finding close-by mobile devices in a single's remote correspondence region due to four reasons. In the first place, clients can appreciate the accommodation of nearby neighbor discovery whenever, while the brought together administration might be inaccessible because of unforeseen reasons. Second, a solitary neighbor discovery protocol can profit different applications by giving more adaptability than the concentrated approach. Third, correspondences between a central server and mobile nodes may actuate issues, for example, excessive transmission overheads, congestion, and unexpected reaction delay. Last yet not slightest, hunting down adjacent mobile devices locally is completely for nothing out of pocket.

1.4 Outline

The report is organized as follows: The introduction is given in Chapter 1. It describes the fundamental terms used in this project. It motivates to study and understand the techniques used for neighbour discovery. This chapter also presents the outline of the objective of the report. The Chapter 2 describes the Literature survey of the project, it describes about all the advancements in the field of Data Transfer done so far. The Chapter 3 presents the proposed work. It describes the major approaches used in this work. it describes of how the system works in order to achieve the expected result.

II. Existing system

Existing neighbor discovery protocols generally fall into two categories, including probabilistic protocols and deterministic protocols.

One of probabilistic protocols introduced a family of "birthday protocols, "which form the foundation of most probabilistic neighbor discovery protocols. In birthday protocols, time is slotted, and each node probabilistically determines the state for each slot from transmitting, listening, and energy-saving, independently. A node makes itself known by its neighbors when it is the only transmitting node in its vicinity in a slot[6].

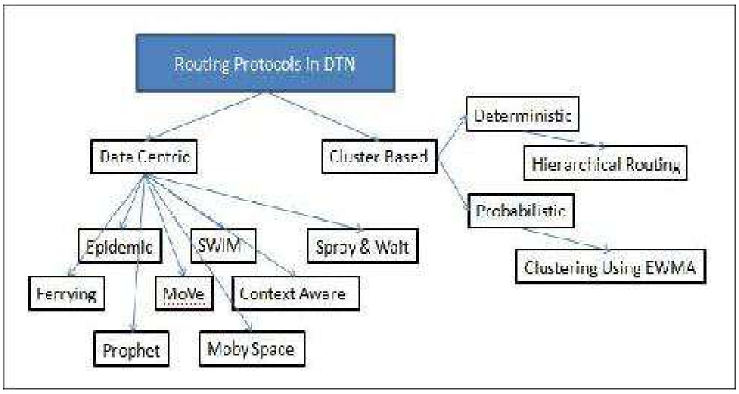
A deterministic protocol establishes a pattern scheduling the periodical operations of each node. A code-based protocol is presented utilizing constant-weight codes but it assumes synchronization among nodes. Moreover, that system applied optimal block designs in the case of symmetric duty cycle[1]. The authors concluded that their approach reduces to an NP-complete minimum vertex cover problem in the asymmetric case, whereas we prove that the bound in that can be further lowered. Besides, our designs fit for both symmetric and asymmetric cases with low complexity[5].

Disadvantages of existing system

Energy efficiency of the system is not satisfactory.

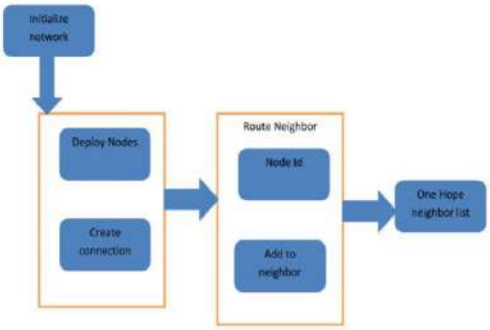
Effectiveness of the system is less.

It considers only synchronous transmission on deterministic neighbor.



III. Proposed system architecture

We adopt a code-based formulation of the neighbor discovery problem and design Diff-Codes for the symmetric case, which is optimal when the Diff-Code can be extended from a perfect difference set. Furthermore, by considering the connection between awake periods of two nodes, we extend Diff-Codes to ADiff-Codes to deal with asymmetric neighbor discovery.



We demonstrate the feasibility conditions of an asynchronous neighbor discovery protocol, from the perspective of both 0–1 code and set theory. We formulate the problem of asynchronous neighbor discovery with symmetric duty cycle mathematically[5]. By the formulation, we derive the lower bound for optimal worst-case latency and design Diff-Codes. We show that a Diff-Code is optimal when it can be extended from a perfect difference set.

We further investigate the feasibility conditions with asymmetric duty cycles and design ADiff-Codes, which can be constructed as long as two pattern codes' lengths are relatively prime. To evaluate the performance of our designs in one-to- one and clique scenarios, we not only conduct comprehensive simulations, but also prototype them using USRP-N210 testbed. Evaluation results show that (A)Diff-Codes significantly reduce the discovery latency in both the median case and worst case. Specifically, in the symmetric case, the maximum improvement is up to 50%; in both symmetric and asymmetric cases, the median case gain is as high as 30% and ADiff-Codes outperform state-of- art protocols in more than 99% of the situations[9].

Usually, there are three challenges in cunning such a neighbor discovery protocol. Neighbor discovery is nontrivial for several reasons: Neighbor discovery needs to deal with collisions. Ideally, a neighbor discovery algorithm desires to minimize the possibility of collisions and, therefore, the time to determine neighbors[4]. In many realistic settings, nodes have no awareness of the number of neighbors, which makes cope with collisions even harder. When nodes do not have right to use a global clock, they have to activate asynchronously and at rest be able to determine their neighbors competently. In asynchronous systems, nodes can potentially initiate neighbor discovery at different times and, therefore, may miss each other's transmissions Furthermore,when the number of neighbors is unknown, nodes do not recognize when or how to conclude the neighbor discovery process. To evaluate the performance of our designs in one-to- one and group scenarios, not only conduct comprehensive simulations, but also sampling them using testbed.Evaluation results show that Diff-Codes drastically decrease the discovery latency in both the median case and worst case.

IV. Equipments and proposed methodology

The entire process of botnet attack on a victim system will be done in a simulated environment . The simulation will be done using Java Netbeans IDE 7.2. The static allocation of the nodes will be done beforehand and the proposed algorithm will be applied on the simulated environment. In the next iteration network simulation will be done using NS2(Network simulator 2). This will be working along with cygwin in the background, to support the simulation in windows OS.

Modules:

1.Problem Definition

The definition of the code construction problem is as follows: For a given, construct a 0–1 code of length with as few 1-bits as possible, while ensuring that is feasible for symmetric neighbor discovery. A symmetric active-sleep pattern with a cycle length of slots should have at least active slots each cycle[9]. This lower bound is tighter than that provided by Zheng because we exploit the power of active slot nonalignment in the asynchronous case. Consequently, compared to the active-sleep patterns, which is identical with perfect difference sets, we achieve much better patterns.

2.Asymptotically Optimal Pattern via Perfect Difference Set

Referring to the set theoretic interpretation of pattern feasibility in Section IV-B, and the definition below, an -perfect difference set already corresponds to a feasible symmetric pattern code of length and weight. An -difference set contains elements. It is a subset of, and each appears exactly times as the difference of two distinct elements from it under module. Specifically, a difference set with is called a perfect difference set. However, being a perfect difference set is a stricter constraint than condition in Corollary[6]. For example, a pattern code can be verified to be feasible, whereas is not a perfect difference set since. To this end, we propose to double the length of a perfect difference set while maintaining its weight. The details can be described as follows: An active slot is extended to two consecutive slots including one active slot followed by another sleeping slot; a sleeping slot is extended to two successive sleeping slots.

3.Diff-Code Construction

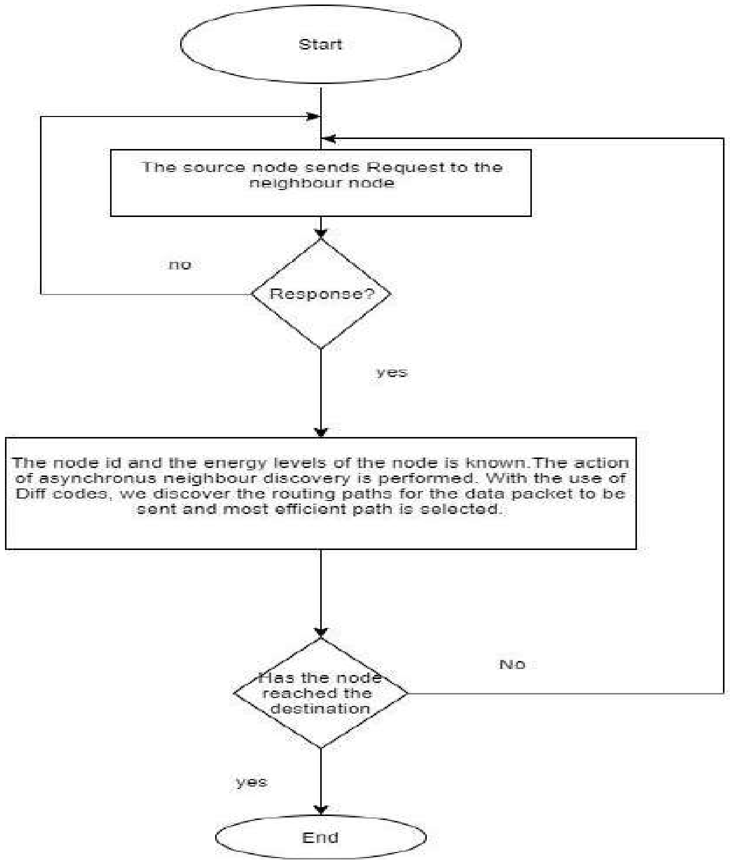
Although doubling the length of a perfect difference set can generate the optimal schedule, it is only suitable for specific code lengths. Therefore, we present the construction of Diff-Codes for any target code length in Algorithm 1. The core idea is to make use of the optimal code with similar length. The first step in the algorithm is to build an initial, but not necessarily feasible, code of the target length. The active slots in are determined by the optimal Diff-Code, whose length is the largest among all the optimal Diff-Codes shorter than intuitive method of initializing is to assign slot active as long as slot is active[9].

4.Theoretical Analysis

By fixing the code length to be, we show the theoretical bound of Diff-Codes' duty cycle. An optimal pattern code directly extended from a perfect difference set with weight will satisfy. Thus, the weight of a Diff-Code with length is at least, which is approximately the lower bound of in Theorem 2 when is fairly large. Because an active slot is overflowed by, the corresponding lower bound of duty cycle is. On the other hand, an optimal Diff-Code whose duty cycle yields that for a large. Therefore, a Diff-Code should contain at least bits to realize a duty cycle of. We compare Diff-Codes with existing protocols, e.g., Disco, U-Connect, and Searchlight, where Searchlight-S is the stripped version of Searchlight[9][10]. The table indicates that in the best cases, Diff-Codes can improve the worst-case latency bound by as high as 50% compared to Searchlight-S. As for Disco, the reduction of the worst-case latency is more than 80%. Moreover, any Diff-Code constructed by Algorithm 1, even not optimal, can outperform other protocols.

5.Diff-Code Seeking With Fixed Duty Cycle

The construction of Diff-Codes discussed until now focuses on minimizing the code weight while the code length is fixed. However, in practice, a user may prefer selecting the appropriate pattern with whatever duty cycle according to the remaining battery of his/her mobile device.Thus, it is necessary to support Diff-Codes construction that minimizes the worst-case latency with a fixed duty cycle[2]. We finish this section by a heuristic algorithm accomplishing such a task.


Start
The source node sends Request to the neighbour node
no
Response?
yes
The node id and the energy levels of the node is known.The action of asynchronus neighbour discovery is performed. With the use of Diff codes, we discover the routing paths for the data packet to be sent and most efficient path is selected.
No
Has the node reached the destination
yes
End

V. Implementation Plan

The input module of the proposed system comprises of the nodes that are going to be statically placed in the simulated environment . The nodes will be given some energy levels and also the active/sleep state of the said nodes will be defined. On the activation of beacon the active nodes will be reflecting their node ID to the surrounding active nodes to create a one-hop neighbor list. The active/sleep nodes will be defined in terms of 1-0 (with 1 being the active state and 0 being the sleep state). Using the set theory for the 1-0 patterns we generate Diff-code by deriving the lower bound for the optimal worst-case latency. The Diff-code is optimal when it can be extended from a perfect difference set. The following node-list table will be connected to the routing table via-datagram protocol and the efficient routing table will be generated, taking into account the state of the nodes, the message that is to be sent to the destination and the energy level of the nodes.

VI. Applications

1 Increased Efficiency in Mobile Data communication.

With the technological advance in today's world mobile phones are a norm. Every individual has a smartphone with them to keep themselves updated with the current trends that are going on around the world. So with the help of this code based approach we can have a better mobile internet connection which can optimize the transfer of data and thus deal with the problems of the irregular connectivity and slow internet which is frustrating in today's world. For example a college student may want to discuss a math problem with other students in the college campus using his/her mobile or tablet[4].

2 Online Multiplayer Games.

With the advent of smartphones and with better UI comes better games with better graphics that help for these genres of games.can be played with the other players in real time. New genres of games such as MMORPG (Massively Multiplayer Online Role Playing Game) , MOBA (Multiplayer Online Battle Arena) have experienced huge rise in recent years. Traditionally these genres have been played in Personal computers using broadband connection(For example-Dota). But the demand of these types of games to be played in mobile devices has increased. This approach will be of massive help.[1]

3 Proximity Based Applications

There is also a rise in proximity based applications used for data sharing and discovering people around you. In such cases this approach can be used along with the central servers and the GPS function to fully exploit the potential of the applications and thus increasing the user satisfaction and substantially decreasing the user frustration while using such application[5].

VII. Acknowledgement

We would also like to thank our Mentor Prof. Payel Thakur without whom this project would not have been possible.

We would also like to thank our Head of Department Dr. Madhumita Chatterjee for her continued support and guidance.

We would also like to extend our gratitude to our Principal Dr. R. I. K. Moorthy for providing us this opportunity. Also a special thanks to all teaching and non teaching staff for providing us with the equipments and environment necessary for this project.

VIII. References

[1] S. Vasudevan, M. Adler, D. Goeckel, and D. Towsley, “Efficient algorithms for neighbor discovery in wireless networks,” IEEE/ACM Trans. Netw., vol. 21, no. 1, pp. 69–83, Feb. 2013.

[2]X. Zhang and K. G. Shin, “E-MILI: Energy-minimizing idle listening in wireless networks, IEEE Trans. Mobile Comput” vol.11, no.9, pp. 1441–1454, Sep. 2012.

[3] W. Zeng et al., “Neighbor discovery in wireless networks with multipacket reception,” in Proc. MobiHoc, 2011, Art. No. 3.

[4] E.Magistretti, O.Gurewitz, and E.W.Knightly,“802.11ec:Collision avoidance without control messages,” in Proc. Mobi Com, 2012, pp. 65–76.

[5] P. Dutta and D. E. Culler, “Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications,” in Proc. SenSys, 2008, pp. 71–84.

[6] Jiang, J.; Tseng, Y.; Hsu, C.; Lai, T. Quorum-based asynchronous power-saving protocols for IEEE 802.11 Ad Hoc networks. In Proceedings of the 2003 International Conference on Parallel Processing, Kaohsiung, Taiwan, 6–9 October 2003; Volume 10, pp. 257–264.

[7] Tseng, Y.-C.; Hsu, C.-S.; Hsieh, T.-Y. Power-saving protocols for IEEE 802.11-based multi-hop ad hoc networks. In Proceedings of the Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, New York, NY, USA, 23–27 June 2002; pp. 200–209.

[8] McGlynn, M.J.; Borbash, S.A. Birthday protocols for low energy deployment and flexible neighbor discovery in Ad Hoc wireless networks. In Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking &amp; Computing, Long Beach, CA, USA, 4–5 October 2001; pp. 137–145.

[9] Zheng, R.; Hou, J.C.; Sha, L. Asynchronous wakeup for Ad Hoc networks. In Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking &amp; Computing, Annapolis, MD, USA, 1–3 June 2003; pp. 35–45.

[10] Anderson, I. Combinatorial Designs and Tournaments; Oxford University Press: Oxford, UK, 1988.

[11] Colbourn, C.J.; Dinitz, J.H. The CRC Handbook of Combinatorial Designs; CRC Press: Boca Raton, FL, USA, 1996.

WEB INDEXING THROUGH HYPERLINKS

Prof.Madhu.N,Aditya Honade, Niraj Pawar, Haresh Shingare, Sudesh Salunke

Abstract—As the size of the Internet is growing rapidly, it has become important to make the search for content faster and more accurate.Web indexing (or Internet indexing) refers to various methods for indexing the contents of a website or of the Internet as a whole. Crawlers have bots that fetch new and recently changed websites, and then indexes them. The objective of our project is that, the uniform resource locator (URL) will be crawled and indexing will be performed on the crawled data to display the results . The relevancy will be checked after the complete hierarchical scanning of the website. A modified version of Depth First Search Algorithm is used to crawl all the hyperlinks along with concept of APIs. These links are then accessed via source code and its meta data such as title/keywords, and description are extracted. This is called indexing of the crawled data. This content is very essential for any type of analyzer work to be carried on the Big Data obtained as a result of Web Crawling.Web indexing is mainly used by search engines.

Keywords- Web Crawling, Filtering techniques, Web Indexing algorithms, Searching techniques.

1. INTRODUCTION

USER ENTERS A QUERY AS INPUT. THIS INPUT QUERY IS THEN CRAWLED AND THE HYPERLINKS ARE INDEXED BASED ON FORWARD AND INVERTED INDEXING. AFTER ALL THE HYPERLINKS HAVE BEEN CRAWLED AND INDEXED THE URLS ARE DISPLAYED TO THE USER AS FINAL OUTPUT ALONG WITH THEIR TITLE, DESCRIPTION, CONTENT BASED SCORE, USAGE BASED SCORE AND TIME SPENT ON A PARTICULAR LINK.

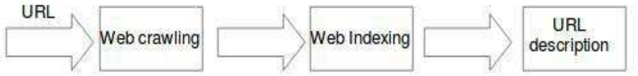


Figure 1.1 The flow

The Figure shows the process and working of the project in a simple fashion. There are 3 steps ie: Crawling, Indexing and displaying the search results.

1.2 Objectives

The objective of our project is to concentrate on crawling the links and retrieving all information associated with them to facilitate easy processing for other uses. Firstly the links are crawled from the specified uniform resource locator (URL) using a modified version of Depth First Search Algorithm along with APIs which allows for complete hierarchical scanning of corresponding web links. The links are then accessed via the source code and its metadata such as title/keywords, and description are extracted. There exists thousands of links associated with each URL linked with the Internet. As the number of pages on the internet is extremely large, even the largest crawlers fall short of making a complete index. For that reason search engines are bad at giving relevant search results. Our first focus is on identifying the best method to crawl these links from the corresponding web URLs. It then builds an efficient extraction to identify the metadata from every associated link. This would give rise to the accumulation of the documents along with the corresponding title, keywords, and description.The aim is to propose an efficient method to crawl and index the links associated with the specified URLs. Maintaining the Integrity of the Specifications.

1.3 Scope

Already a lot of research is going on in the field of web data extraction techniques. In future work can be done to improve the efficiency of algorithms. Also, the accuracy and timeliness of the search engines can also be improved. The work of the different crawling algorithms can be extended further in order to increase the speed and accuracy of web crawling [9]. A major open issue for future work about the scalability of the system and the behavior of its components. Building an effective web crawler to solve different purposes is not a difficult task, but choosing the right strategies and building an effective architecture will lead to implementation of highly intelligent web crawler application.In this domain, various challenges in the area of Hidden web data extraction and their possible solutions have been discussed. Although this system extracts, collects and integrates the data from various hidden websites successfully, this work could be extended in near future. In this work, a search engine shell has been created which was tested on a particular domain. This work could be extended for other domains by integrating this work with the unified search interface and rms do not have to be defined.

2.2 Technique

2.2.1 Crawling

Depth First Search Algorithm: R.Suganya Devi, D.Manjula and R.K. Siddharth proposed that the most effective way to crawl a web is to access the pages in a depth first manner. This allows the crawled links to be acquired in a sequential hyperlink manner. The system uses the concept of metadata tag extraction to store the URL, title, keywords and description in the database.are used to crawl links from the web which has to be further processed for future use, thereby increasing the overload of the analyser. It mainly concentrates on crawling the links and retrieving all information associated with them to facilitate easy processing for other uses. The aim was to propose an efficient method to crawl and index the links associated with the specified URLs.The links are then accessed via the source code and its metadata such as title, keywords, and description are extracted. This content is very essential for any type of analyser work to be carried on the Big Data obtained as a result of Web Crawling. This content is very essential for any type of analyser work to be carried on the Big Data obtained as a result of Web Crawling.

API

1. User queries the system. The input can be any word in users mind. The system matches the query word not only with service interface but also with its methods.
2. The request goes to Google Custom Search Engine through Google Custom Search API.
3. The engine has been scaled to the desired links to crawl. It can be scaled any time.
4. Engine crawls on all the links given and produces the results.
5. Results produced are not user understandable format. So the system parses the results produced.
6. System Extracts the Wsdl files from the set of results.
7. Results are displayed to the Client.
8. To check whether the service is available at given time. We have performed the validity check.
9. Results are displayed and sent to local database.

2.2.2 Indexing

Baseline Implementation, MapReduce was designed from the very beginning to produce the various data structures involved in web search, including inverted indexes and the web graph.Input to the mapper consists of document ids (keys) paired with the actual content (values). Individual documents are processed in parallel by the mappers. First, each document is analyzed and broken down into its component 5 terms.MapReduce was designed from the very beginning to produce the various data structures involved in web search, including inverted indexes and the web graph.A graph within a graph is an "inset," not an "insert." The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).

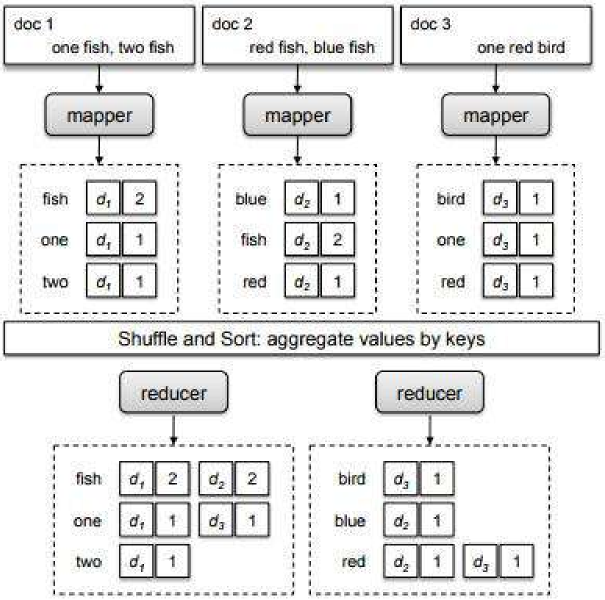
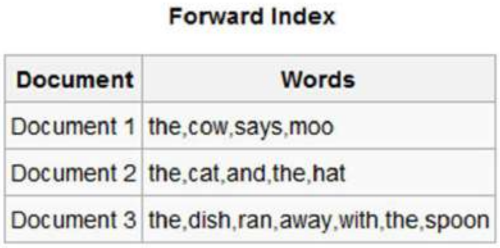


Figure 2.1

2.2.3 Inverted/Forward Indexing

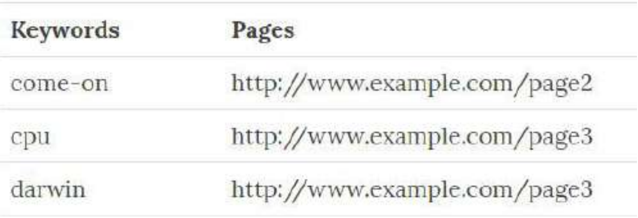
In computer science, an **inverted index** (also referred to as **postings file** or **inverted file**) is an index data structure storing a mapping from content, such as words or numbers, to its locations in a database file, or in a document or a set of documents (named in contrast to a forward index, which maps from documents to content). The purpose of an inverted index is to allow fast full text searches, at a cost of increased processing when a document is added to the database. The inverted file may be the database file itself, rather than its index. It is the most popular data structure used in document retrieval systems,used on a large scale for example in search engines.



Forward Index

Document	Words
Document 1	the,cow,says,moo
Document 2	the,cat,and,the,hat
Document 3	the,dish,ran,away,with,the,spoon

Inverted Index



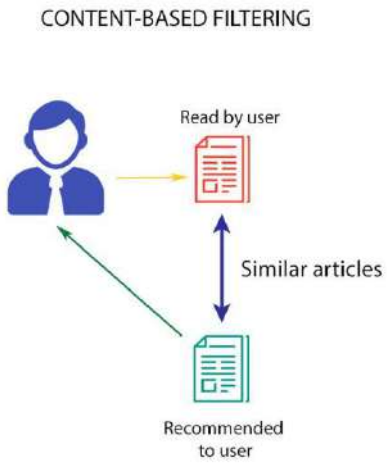
Keywords	Pages
come-on	http://www.example.com/page2
cpu	http://www.example.com/page3
darwin	http://www.example.com/page3

2.3 Filtering Techniques

2.3.1 Content Filtering

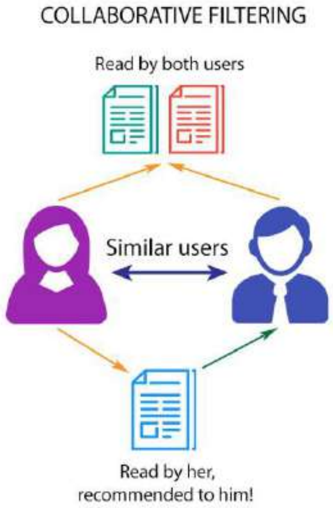
Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.

The efficiency of a learning method does play an important role in the decision of which method to choose. The most important aspect of efficiency is the computational complexity of the algorithm, although storage requirements can also become an issue as many user profiles have to be maintained.



2.3.2 Collaborative Filtering

Collaborative filtering (**CF**) is a technique used by recommender systems.Collaborative filtering has two senses, a narrow one and a more general one.In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person. For example, a collaborative filtering recommendation system for television tastes could make predictions about which television show a user should like given a partial list of that user's tastes (likes or dislikes).Note that these predictions are specific to the user, but use information gleaned from many users. This differs from the simpler approach of giving an average (non-specific) score for each item of interest, for example based on its number of votes.

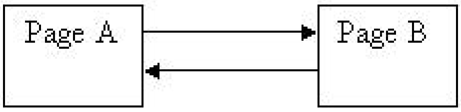


2.3.3 Page Rank

Page Rank is a topic much discussed by Search Engine Optimisation (SEO) experts. At the heart of PageRank is a mathematical formula that seems scary to look at but is actually fairly simple to understand.

PageRank is one of the methods Google uses to determine a page's relevance or importance. It is only one part of the story when it comes to the Google listing, but the other aspects are discussed elsewhere (and are ever changing) and PageRank is interesting enough to deserve a paper of its own.

$$PR(A) = (1-d) + d\,(PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$



Each page has one outgoing link (the outgoing count is 1, i.e. C(A) = 1 and C(B) = 1).

2.4 Hybrid Technique

From the first technique we get the efficient way of crawling the websites. In this technique it is said that the depth first search algorithm is the efficient approach for crawling and also the APIs (custom key) can be used. In the second technique it is well described about how to index the crawled pages or links. After crawling the websites all the hyperlinks from that site are gathered and are indexed accordingly. In the third technique the problem with dynamic pages is resolved. The dynamic pages are indexed successfully. Using all these three techniques the output ie: the links stored in a file are then displayed to the user.

3.1.1 Existing System Architecture

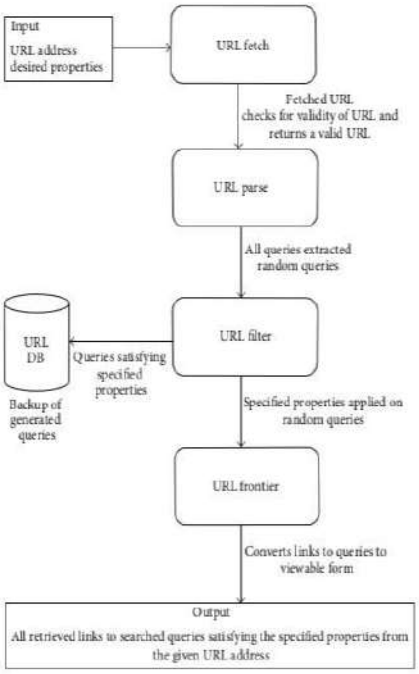

Input
URL address
desired properties
URL fetch
Fetched URL
checks for validity of URL and returns a valid URL
URL parse
All queries extracted
random queries
URL DB
Backup of generated queries
Queries satisfying specified properties
URL filter
Specified properties applied on random queries
URL frontier
Converts links to queries to viewable form
Output
All retrieved links to searched queries satisfying the specified properties from the given URL address

Figure 3.2: Existing System Architecture.

The existing system functions as a search bot to crawl the web contents from a site. The system is built by developing the front end on .NET framework on Visual Studio 2012 supported by the Microsoft SQL Server Compact 3.5 as the back-end database. It is then used to interpret the crawled contents based on a user created file named robots.txt file. The working of this system is based on ability of the system to read the web URL and then access all the other pages associated with the specified URL through hyperlinking. This allows the user to build a searchable indexer. This is facilitated by allowing the system to access the root page and all its subpages. The robots.txt file can be used to control the search engine, thereby allowing or disallowing the crawling of certain web pages from the specified URL.

3.1.2 Proposed System Architecture

Proposed System

The system overview is presented in this Section. The classification of various techniques the domain is given in Figure 3.3.

In proposed system the user enters an input url. the url is then processed and filtered from the data sets,where the validity of the url is checked.All the hyperlinks in the particular websites are stored into stack.Using depth first search algorithm each link is popped from the stack in crawled file,till the stack gets empty.Check if you have already crawled the URLs and/or if you have seen the same content before. If not add it to the index. In general, hybrid recommender are systems that combine multiple recommendation techniques together to achieve a synergy between them. The proposed architecture is shown in Figure 3.3

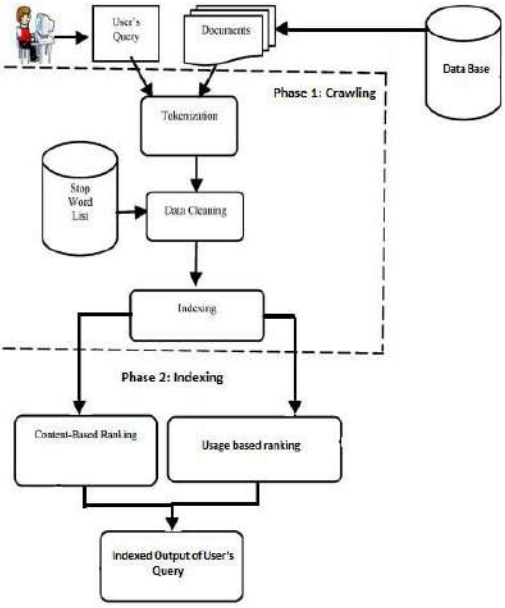


Figure 3.3: Proposed system architecture.

Hybrid models discussed two techniques that have merged in four different ways.

The System Flow Chart is as shown in the figure given below.(figure 3.4)

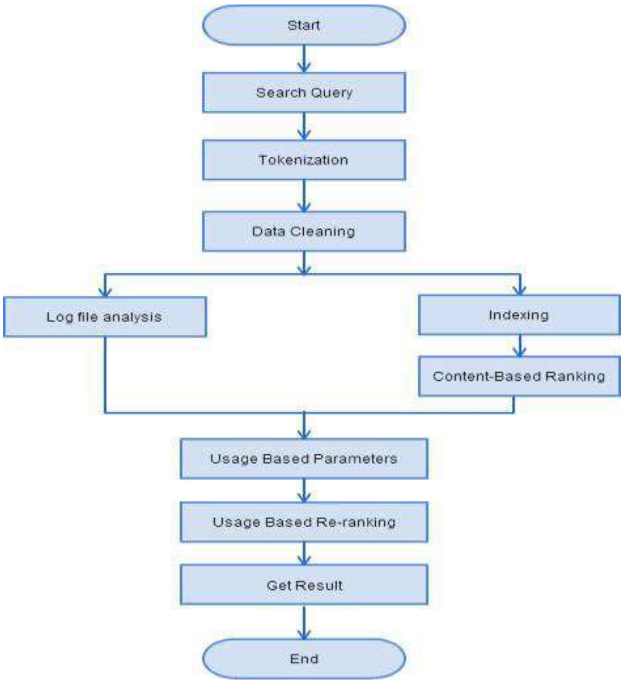


Figure 3.4 :System Flow Chart

4.1 Conclusion

Based on our study, the proposed system is mainly focused on building a database of pages and links from the World Wide Web. After an initial boost, it is found that the durable pages that are required to be crawled occur with more probability as the total number of pages increases. This shows that when applied to a real-time application which handles millions of data, the performance numbers are bound to reach the maximum efficiency, thereby presenting the most efficient Web Crawling System. The advantageous addition to this system is its information integration with the simultaneous meta tag extraction.It also focuses on re-crawling frequently changing web pages so as to keep the contents of the database current.Future works could be done on reducing the amount of bandwidth required to generate this system and make it accessible to the next level of links.

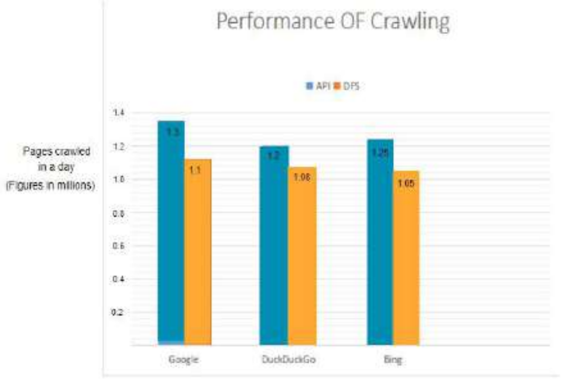


Figure 4.1 Crawling performance graph

The above figure compares the techniques used for crawling ie: Depth First Search and API's.This graph is a representation of performance achieved by making use of API's and Depth first search algorithm .This graph shows the number pages crawled by each technique in a day. API's clearly crawl more pages when compared to DFS for all three search engines.

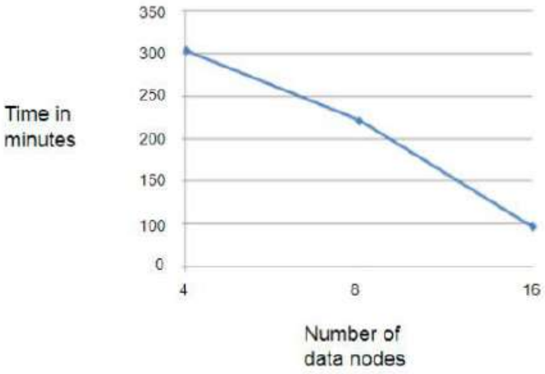


Figure 4.2 Indexing performance graph

The following graph represents the time required in minutes to index the data depending on the number of nodes. The graph is a clear indication of how the increase in number of nodes reduces the time required to index the data. This indicates that system is able to handle larger data sets by having more resources.

4.2 APPLICATIONS

There are various applications of this domain system. The application is listed here.

• Crawler can use in many legitimate sites, in particular search engines, use spidering as a means of providing up-to-date data.

• crawlers are used for automating maintenance tasks on the web site.

• A Web crawler is an Internet bot that systematically browses a metric of importance for prioritizing Web pages.

• Web crawler are use for dramatically reduced the amount of time required to find programs and documents.

• Crawler is a program that is use visits web sites and reads their pages and other information in order to create entries for a search engine index.

4.3 REFERENCES

1) An Efficient Approach for Web Indexing of Big Data through Hyperlinks in Web Crawling. R. Suganya Devi, D. Manjula, and R. K. Siddharth Department of Computer Science and Engineering, Anna University, India.Accepted 29 March 2015

2) A framework for dynamic indexing. Hasan Mahmud, Moumie Soulemane, Mohammad Rafiuzzaman Department of Computer Science and Information Technology, Islamic University of Technology, Board Bazar, Gazipur-1704,Bangladesh.

3) S. Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data, Morgan Kaufmann Publishers, 2002.

4) Algorithms, International Journal of Computer Science, vol. 8, iss. 6, no 1, Nov. 2011

5) Pavalam, S. M., SV Kashmir Raja, Felix K. Akorli, and M. Jawahar, A Survey of Web Crawler

6) Pavalam, S. M., SV Kashmir Raja, Felix K. Akorli, and M. Jawahar, A Survey of Web Crawler Algorithms, International Journal of Computer Science, vol. 8, iss. 6, no 1, Nov. 2011

7) Breadth First Search, Accessed March 16, 2013, en.wikipedia.org/wiki/Breadthfirstsearch

8) D. Singh and C. K. Reddy, A survey on platforms for big data analytics, Journal of Big Data, vol. 2, article 8, 2014. Swati Mali, Dr. B.B Meshram, Implementation of multiuser personal web crawler, In CSI 6th Int. Conf. on SE(CONSEG), IEEE Conf. Publication, 2012.

9) SIGNATURE FILE METHODS FOR INDEXING, Wangchien Lee and Dik L Lee Department of Computer and Information Science Ohio State University Columbus M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

10) Holger Lausen and Thomas Haselwanter, “Finding WebServices” 2007.

11) Mydhili K Nair, Dr. V.Gopalakrishna, “Look Before You Leap: A Survey of Web Service Discovery” International Journal of Computer Applications (0975–8887) Volume 7 No.5, September 2010

12) Xin Dong Alon Halevy Jayant Madhavan Ema Nemes Jun Zhang, “Similarity Search for Web Services” Proceedings of the 30th VLDB Conference,Toronto, Canada, 2004.

.

TV Program Recommendation System using Classification Techniques Based on Reviews

Jinesh Jain, Rejosh Rajan, Mili Taneja ,Tejas Bangera and Satishkumar Varma
Department of Computer Engineering, PCE, New Panvel

Abstract: There are large number of online reviews about television shows. These reviews are available on the Web, such as Facebook, Netflix, Twitter. Television ratings information is a key element in the entertainment business of the consumers. These reviews contain valuation information for both consumers and firms. There are good number of channels and also the competition across various channel categories is high. The program content helps channels to gain a high TRP (Television Rating Point). So it is important to know the contents of the program and expectation of the viewers from consumer's reviews. To extract information about a program and user ratings from reviews, different data mining techniques are used. The different data sets available online are used to know about the ratings of such programs. In order to do good business, different data mining techniques are implemented to extract information that can be used by the firm to select appropriate program time slot and improved content. In this project, the classification algorithm such as Naive Bayes, decision tree are used. The user stands a chance to evaluate the program based on the ratings given and also get to know the context of the program. Program's rating depends on its broadcast time. System also categorizes theme of the program on the basis of the content for e.g. drama, comedy, horror, etc. This help to get the period of high rating that can be used by the companies for customer behaviour and business purposes.

Keywords: Sentiment analysis, Classification Algorithm, Filtration, NoSql Database, Visualization, Twitter API.

1. Introduction

Entertainment is a very important part of our life. Most of the people entertain themselves by watching television. The current explosion of the number of available channels is making the choice of the program to watch and experience more and more difficult for the TV viewers. Such a huge amount obliges the users to spend a lot of time in consulting TV guides and reading synopsis, with a heavy risk of even missing what really would have interested them. With the evolution of TV programs, the need of recommended systems for TV has increased substantially. As the TV shows are going on increasing it becomes difficult for the users to choose between the best of the programs. The objective of our project makes the user comfortable with choosing the best program which most of the people opt to watch in their free time. This project will help people to differentiate the programs based on comedy, drama, action, etc.

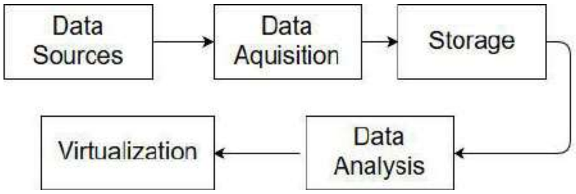


Figure 1. Flow Diagram

The objective of our project makes the user comfortable with choosing the best program which most of the people opt to watch in their free time. This project will help both the consumers who sees the TV shows and channelside to increase the value of the channel. Reduce the wastage of time, Enhanced performance with easily distinguishable interface. Easy and understandable Pie-charts are being generated for customer review. Customer requirement and their values are maintained. People will be able to get to know about the popularity of the show and also the reviews given by the different viewers. People will also be able to see comedy, drama as per the peoples requirement without actually watching the program and knowing himself. This projects help the parent to know about TV shows whose content is children based and has values associated with its relevance to the theme.

2 . Literature survey

The recommendation system of traditional television is mainly based on audience rating, which reacts the audience's attitude towards the program. Data sources are defined in two dimensions, mobility and structure of data. Streaming data refers to a data flow to be processed in real time, e.g. a Twitter stream. Second, structure of the data source is defined. Most data acquisition scenarios assume high-volume, high-velocity, high variety, but low value data, making it important to have adaptable and time efficient gathering, filtering, and cleaning algorithms that ensure that only the high value fragments of the data are actually processed by the data-warehouse analysis.

Paper	**Techniques**	**Data set**	**Parameter used**
Zhaocai Ma et al. 2014	Collaborative filtering technique, Self organizing algorithm, K means clustering	100k dataset which is collect by the GroupLens Res team of the Minnesota university. The experimental data set contains 943 users, 1682 films, and 100,000 ratings.	Users search history, User's past behavior (such as browse or purchase records, etc).
Huayu Li et al. 2015.	Gibbs Sampling and variational inference.	Three types of online review data: the hotel review data collected from TripAdvisor 2, the beer review data collected from RateBeer 3, and the app review data crawled from Applause	document distribution θ, neutral ratio t, predicted ratings Ω, words distribution φ, and two latent variables (i.e., topic z and sentiment index s) need to be estimated in the model.
Mengyi Zhang, et al. 2016	Data Mining techniques like clustering and classification techniques	Dynamic. 20000 to 40000 hits in 3-4 hrs.	program ratings, television ratings, program type, program broadcast Time.

Chengfeng Zhang, et al. 2016.	Naïve Bayes classification, Classification and Regression Tree(CART)and Random Forest (RF)	370 examples of credit applicants.	Gender (M/F), Type of Loan (New, Existing), Amount Requested, Currency, Period/Months, Purpose of Loan, Credit History, IS the borrower in Salary Project (S or N/S), Currency of income, Pledge_Gur, Pledge_Pldg, Status (Approved/Rejected/Disputed).
Wararat Songpan, et al. 2017	Classification technique like naïve bayes and decision tree is used out of which naïve bayes is proved beneficial	400 customer reviews. 36 words of positive and negative compare with.	Reviews, Positive words, negative words.

3. Existing System Architecture

The recommendation system of traditional television is mainly based on audience rating, which reflects the audience's attitude towards the program. Data sources are defined in two dimensions, mobility and structure of data. Streaming data refers to a data flow to be processed in real time, e.g. a Twitter stream. Second, structure of the data source is defined. Structured data has a strict data model. Examples of semi-structured data include XML and JSON documents. Most data acquisition scenarios assume high-volume, high-velocity, high-variety, but low-value data, making it important to have adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value fragments of the data are actually processed by the data-warehouse analysis.

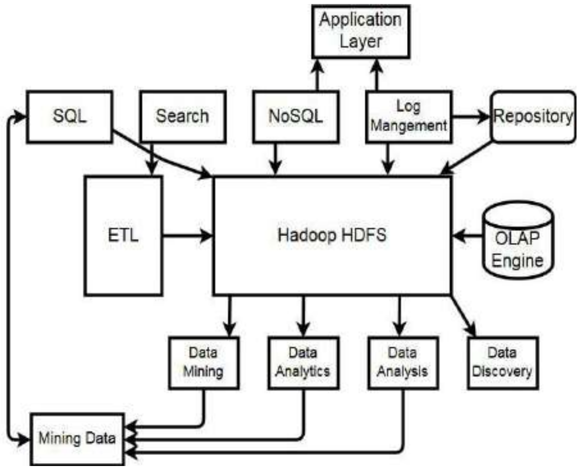


Figure 2. Existing system architecture

Instead of applying schema on write, NoSQL databases apply schema on read. With MongoDB organizations are serving more data, more users, more insight with greater ease and creating more value worldwide. MongoDB document model enables us to store and process data of any structure: events, time series data, geospatial coordinates, text and binary data, and anything else. Many algorithms are used for data analysing based on the consumers required. Clustering and Classification algorithms are used for recommendation of objects.

4. Proposed System Architecture

The proposed system architecture focuses on the following objectives which are helpful in increasing the security of data storage. The classification of various techniques used is given in Figure.

Data sources: It fetches data from the twitter official pages as our datasets. The tweets of the people are collected and analysis is done on that data.The shows of Netflix are taken People give positive and negative reviews about the TV shows on the pages.

Data Acquisition: Stopwords, Retweet tags, links, non english words are removed from the tweets. In filtration process, non english words are matched with the english words which is stored in text file acting as word database. Sentimental analysis is done on the filtered english reviews given by the people. Based on number of positive and number of negative words score is generated.

Data Analysis: For storage NoSQL database Mongodb is used. It is flexible and expandable. With the help of queries score gets stored in Mongodb. With the score, date, category, positive, negative, show name are also stored in database.

Data Analysis:

These techniques helps the system to classify the tv shows based on categories and score. This classification technique categorizes the programs based on the scores generated and shows the results to the user. Decision tree and Naive Bayes is used as classification techniques for data analysis process.

Visualization: This ratings help the channels side people to understand the popularity of the channel, viewers, revenue generated etc. Histogram and pie charts are used to represent the results to the user in the most convincing way. This helps to know about the complete popularity and growth about particular tv shows.

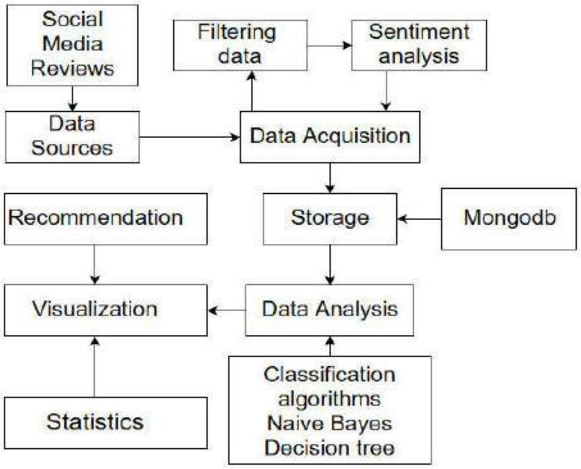


Figure 3. Proposed system architecture.

4.1 Naive Bayes Method

Author Wararat Songpan [2] applied similar techniques on different datasets. Naive Bayesian probability model help to mine the massive program text information to extract users, nature and practical learning algorithms and prior knowledge and observed data can be combined.

Advantages

It is a relatively easy algorithm to build and understand. It is faster to predict classes using this algorithm than many other classification algorithms.

Disadvantages

If a given class and a feature have 0 frequency, then the conditional probability estimate for that category will come out as 0. Another disadvantage is the very strong assumption of independence class features that it makes.

4.2 Decision tree

Decision tree builds classification or regression models in the form of a tree structure. The final result is a tree with decision nodes and leaf nodes. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

Advantages

Able to handle both numerical and data. Requires little data preparation. Other techniques often require data normalization. Since trees can handle qualitative predictors, there is no need to create dummy variables. Large amounts of data can be analysed using standard computing resources in reasonable time.

Disadvantages

Trees do not tend to be as accurate as other approaches. A small change in the training data can result in a big change in the tree. The problem of learning an optimal decision tree is known to be NP-complete under several aspects of optimality and even for simple concepts.

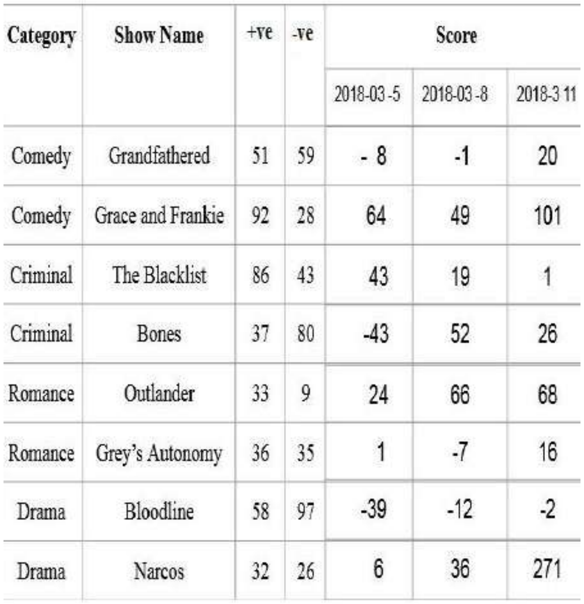
5. Result Analysis

5.1 Dataset Used

There are different datasets used in the making of this recommendation system. They are the tweets obtained from the twitter. we obtained 200 tweets from each programs from the Twitter. Based on the tweets, we classify it as how popular the show is.

Categories	No of programs	No of tweets captured
Drama	5	200
Comedy	6	200
Horror	5	200
Romance	5	200
Mystery	5	200
Criminal	5	200

Table : Sample Dataset for Experiment



Category	Show Name	+ve	-ve	Score		
				2018-03 -5	2018-03 -8	2018-3 11
Comedy	Grandfathered	51	59	- 8	-1	20
Comedy	Grace and Frankie	92	28	64	49	101
Criminal	The Blacklist	86	43	43	19	1
Criminal	Bones	37	80	-43	52	26
Romance	Outlander	33	9	24	66	68
Romance	Grey's Autonomy	36	35	1	-7	16
Drama	Bloodline	58	97	-39	-12	-2
Drama	Narcos	32	26	6	36	271

Figure 4. database of tv shows

5.2 Performance Metrics

The complexity parameter (cp) is used to control the size of the decision tree and to select the optimal tree size. If the cost of adding another variable to the decision tree from the current node is above the value of cp, then tree building does not continue. It regulates the splitting of nodes and growing of a tree by preventing splits that are deemed not important enough. In particular, those would be the splits that would not improve the fitness of the model by at least the cp value.

Kappa coefficient of agreement were calculated to compare the performance of the models. Kappa is used to compare the performance of classifiers as it provides a more robust measure of agreement than accuracy, because it takes into account the expected agreement by random chance. This would imply that the model has a capacity to generalize patterns and has not been over-fitted to the training data.

Techniques : Naive Bayes

The Naive Bayes classifier calculates class probabilities based on Bayes' rule. It assumes that each input feature is independent and that the probability distribution of each class for each feature is Gaussian and so is the only parametric technique tested here.

User kernel	Accuracy	Kappa
FALSE	0.995	0.994
TRUE	0.986	0.984

Table 1: Metrics of naive bayes model

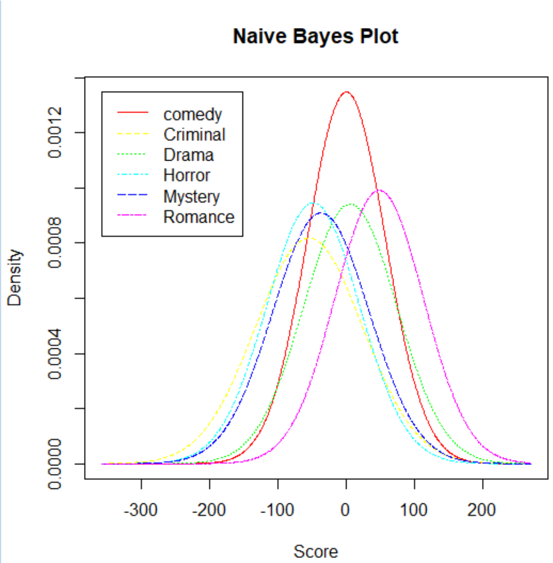


Figure 5 . Graph of the model with respect to score

Techniques : Decision Tree

Database of score is divided into training and testing dataset. Kappa coefficient of agreement were calculated to compare the performance of the models. Kappa is used to compare the performance of classifiers as it provides a more robust measure of agreement than accuracy, because it takes into account the expected agreement by random chance. To implement decision tree caret package in R is used to classify target variable tv shows and predictor variable i.e score and category. Caret package stands for classification and regression tree. Here first data is spilt with the probability of 0.85 as training data and respective testing data. Then train control is used for random resampling, repeated cross validation on the training data.

Sr. no	Cp	Accuracy	Kappa
1	0.007	0.506	0.489
2	0.009	0.494	0.477
3	0.011	0.488	0.471
4	0.012	0.470	0.453
5	0.014	0.446	0.428
6	0.016	0.425	0.407
7	0.018	0.424	0.405
8	0.029	0.315	0.295
9	0.031	0.230	0.210
10	0.033	0.082	0.065

Table 2. Metrics of decision model

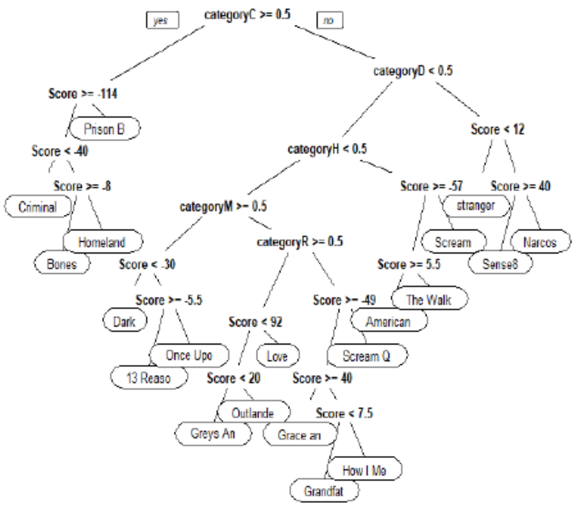


Figure 6. Decision tree of score, category as independent variable

5.3 Results

The Implementation of the code is as follows:

The implementation of the project yielded the following results. Here based on the score, program recommendation is done.

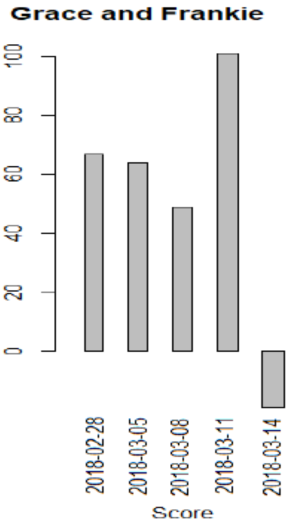


Figure 7. Score of Grace and Frankie with respect difference of 3 days

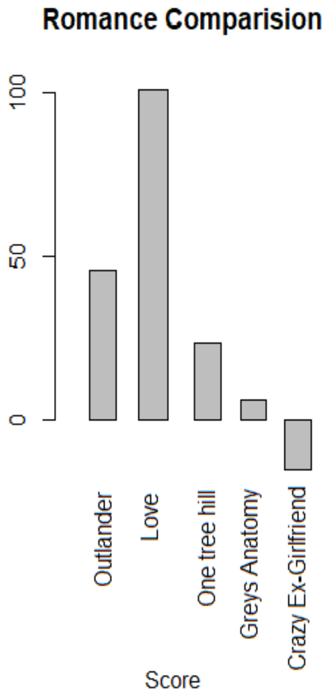


Figure 8. Comparison of different romance TVshows by averaging the score

7. Conclusion

In this report, the study of different classification techniques is presented. The different techniques such as Decision tree and Naive Bayes is explained with examples. There are errors in program ratings in traditional recommendation system, and the TV program list is affected by human emotion as well. Our TV Program Recommended system based on Data Mining reasonably gives solution to those drawbacks. According to the data collected and multi-dimensional analysis, we can find the most beneficial television broadcast playbill, and discover the hot topics. On the basis of this paper, further work on detailed data of audience's behavior can be carried on. The different standard data-sets or variable inputs are defined that may be used in experiment for this domain systems. The different data-sets identified are Score, Characters, Reviews, Likes, Comment, Category, Content, Name of the show, Language.

8. Acknowledgement

Dr. Satishkumar Varma, for acting as our project guide and for his valuable guidance, patience, keen interest and constant encouragement and for his invaluable support. Dr. Sharvari Govilkar, Head, Department of Information Technology and Dr. Madhumita Chatterjee, Head, Department of Computer Engineering Department for their invaluable support. Dr. R.I.K Moorthy and Dr. Sandeep Joshi for their valuable support. Staff members of the Department of Computer Engineering and Information Technology for their critical advice and guidance.

References

1. M. Zhang, M. Shi, Z. Hong, S. Shang and M. Yan, "A TV program recommendation system based on big data," 2016 IEEE/ACIS 15th International Conference on Computer and Information Science. Okayama, 2016.
2. S. L. Wu, R.D. Chiang and Z.H. Ji," Development of a Chinese opinion-mining system for application to Internet online forum", The Journal of Supercomputing, Springer US[Online], 2016.
3. Guoshuai Zhao, Xueming Qian, Member, IEEE, Chen Kang, "Service Rating Prediction by Exploring Social Mobile Users.
4. Wararat Songpan Department of Computer Science, Faculty of Science, Khon Kaen University Khon Kaen, Thailand "The Analysis and Prediction of Customer Review Rating Using Opinion Mining.
5. Qing Wang, "Design and implementation of recommender system based on Hadoop," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 295-299.
6. T. Qing-ji, W. Hao, W. Cong and G. Qi, "A personalized hybrid recommendation strategy based on user behaviors and its application," *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, Shenzhen, China, 2017, pp. 181-186.
7. S. Liu, Y. Dong and J. Chai, "Research on personalized recommendation system of media tags based on system dynamics," *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Shanghai, China, 2017, pp. 1-5.
8. N. Arunachalam, A. Amuthan, M. Sharmilla and K. Ushanandhini, "Survey on web

service recommendation based on user history," *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, Melmaruvathur, 2017, pp. 305-309.

9. Z. Zhao *et al.*, "Social-Aware Movie Recommendation via Multimodal Network Learning," in *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 430-440, Feb. 2018.doi: 10.1109/TMM.2017.2740022
10. S. Khater, D. Gračanin and H. G. Elmongui, "Personalized Recommendation for Online Social Networks Information: Personal Preferences and Location-Based Community Trends," in *IEEE Transactions on Computational Social Systems*, vol. 4, no. 3, pp. 104-120, Sept. 2017.doi: 10.1109/TCSS.2017.2720632
11. N. Arunachalam, S. J. Sneka and G. MadhuMathi, "A survey on text classification techniques for sentiment polarity detection," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, 2017, pp. 1-5.
12. A. D. Dave and N. P. Desai, "A comprehensive study of classification techniques for sarcasm detection on textual data," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, 2016, pp. 1985-1991.
13. J. Kumar and V. Garg, "Security analysis of unstructured data in NOSQL MongoDB database," *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, Gurgaon, 2017, pp. 300-305.
14. K. V. Isabella, L. Sampebatu and I. Albarda, "Analysis of earthquake magnitude level based on data Twitter with decision tree algorithm," *2017 International Conference on Information Technology Systems and Innovation (ICITSI)*, Bandung, 2017, pp. 73-76.
15. S. Ryu, K. h. Han, H. Jang and Y. I. Eom, "User Adaptive Recommendation Model by Using User Clustering based on Decision Tree," *2010 10th IEEE International Conference on Computer and Information Technology*, Bradford, 2010, pp. 1346-1351.

Image Geotagging Using Self-Organizing Map

Alok Jha, *Student, PCE*, Sandeep Menon, *Student* Pranav Nakhwa , *Student, PCE*, and Vaibhav Magar, *Student, PCE*, and

Dr.Satish Kumar Varma, Faculty, *PCE*

Abstract: Automated identification of the geographical coordinates based on image content is of particular importance to data mining systems, because geo-location provides a large source of context for other useful features of an image. However, successful localization of images which are not annotated requires a large collection of images that cover all possible locations. Brute-force searches over the entire databases are costly in terms of computation and storage requirements, and achieve limited results. Knowing what visual features make a particular location unique or similar to other locations can be used for choosing a better match between spatially distance locations. However, doing this at global scales is a challenging problem. In this paper we propose an on-line, unsupervised, clustering algorithm called Location Aware Self-Organizing Map (LASOM), for learning the similarity graph between different regions. The goal of LASOM is to select key features in specific locations so as to increase the accuracy in geo-tagging untagged images, while also reducing computational and storage requirements. Different from other Self-Organizing Map algorithms, LASOM provides the means to learn a conditional distribution of visual features, conditioned on geospatial coordinates. We demonstrate that the generated map not only preserves important visual information, but provides additional context in the form of visual similarity relationships between different geographical areas. We show how this information can be used to improve geo-tagging results when using large databases. However, the size and nature of these databases pose great challenges. Our method achieves promising results when used on a large dataset. We further show that the learned representation results in minimal information loss as compared to using k-Nearest Neighbor method. The noise reduction property of LASOM allows for superior performance when combining multiple features.

I. Introduction

We live in an information age touched by technology in all aspects of our existence, be it work, entertainment, travel, or communication. The extent to which information pervades our lives today is evident in the growing size of personal and community footprints on the web, ever improving modes of communication, and fast evolving internet communities (such as Flickr, Twitter, and Facebook) promoting virtual interactions. In some aspects, man has transformed from a social being into an e-social being.

Images and video constitute a huge proportion of the Web information that is being added or exchanged every second. The popularity of digital cameras and camera phones has contributed to this explosion of personal and Web multimedia data. Finally, determining where an image was taken is valuable to the intelligence. community for use in surveillance. The availability of geo-tagged images on sites such as Flickr has allowed researchers to explore the problem of automatic geo-tagging of images and videos that are missing such information.

II. Guidelines For Manuscript Preparation

When you open TRANS-JOUR.DOC, select “Page Layout” from the “View” menu in the menu bar (View | Page Layout), (these instructions assume MS 6.0. Some versions may have alternate ways to access the same functionalities noted here). Then, type over sections of TRANS-JOUR.DOC or cut and paste from another document and use markup styles. The pull-down style menu is at the left of the Formatting Toolbar at the top of your *Word* window (for example, the style at this point in the document is “Text”). Highlight a section that you want to designate with a certain style, then select the appropriate name on the style menu. The style will adjust your fonts and line spacing. Do not change the font sizes or line spacing to squeeze more text into a limited number of pages. Use italics for emphasis; do not underline.

To insert images in *Word,* position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with “float over text” unchecked).

IEEE will do the final formatting of your paper. If your paper is intended for a conference, please observe the conference page limits.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have already been defined in the abstract. Abbreviations such as IEEE, SI, ac, and dc do not have to be defined. Abbreviations that incorporate periods should not have spaces: write “C.N.R.S.,” not “C. N. R. S.” Do not use abbreviations in the title unless they are unavoidable (for example, “IEEE” in the title of this article).

B. Other Recommendations

Use one space after periods and colons. Hyphenate complex modifiers: “zero-field-cooled magnetization.” Avoid dangling participles, such as, “Using (1), the potential was calculated.” [It is not clear who or what used (1).] Write instead, “The potential was calculated by using (1),” or “Using (1), we calculated the potential.”

Use a zero before decimal points: “0.25,” not “.25.” Use “cm^3,” not “cc.” Indicate sample dimensions as “0.1 cm × 0.2 cm,” not “0.1 × 0.2 cm^2.” The abbreviation for “seconds” is “s,” not “sec.” Use “Wb/m^2” or “webers per square meter,” not “webers/m^2.” When expressing a range of values, write “7 to 9” or “7-9,” not “7~9.”

A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) In American English, periods and commas are within quotation marks, like “this period.” Other punctuation is “outside”! Avoid contractions; for example, write “do not” instead of “don’t.” The serial comma is preferred: “A, B, and C” instead of “A, B and C.”

If you wish, you may write in the first person singular or plural and use the active voice (“I observed that ...” or “We observed that ...” instead of “It was observed that ...”). Remember to check spelling. If your native language is not English, please get a native English-speaking colleague to carefully proofread your paper.

A. How to Create a PostScript File

First, download a PostScript printer driver from http://www.adobe.com/support/downloads/pdrvwin.htm (for Windows) or from http://www.adobe.com/support/downloads/pdrvmac.htm (for Macintosh) and install the “Generic PostScript Printer” definition. In *Word,* paste your figure into a new document. Print to a file using the PostScript printer driver. File names should be of the form “fig5.ps.” Use Open Type fonts when creating your figures, if possible. A listing of the acceptable fonts are as follows: Open Type Fonts: Times Roman, Helvetica, Helvetica Narrow, Courier, Symbol, Palatino, Avant Garde, Bookman, Zapf Chancery, Zapf Dingbats, and New Century Schoolbook.

III. Math

If you are using *Word,* use either the Microsoft Equation Editor or the *MathType* add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation *or* MathType Equation). “Float over text” should *not* be selected.

A. Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). First use the equation editor to create the equation. Then select the “Equation” markup style. Press the tab key and write the equation number in parentheses. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Use parentheses to avoid ambiguities in denominators. Punctuate equations when they are part of a sentence, as in

$$X = \sum_{i=0}^{5} R_i + T \qquad (1)$$

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. Italicize symbols (T might refer to temperature, but T is the unit tesla). Refer to “(1),” not “Eq. (1)” or “equation (1),” except at the beginning of a sentence: “Equation (1) is ... .”

IV. Units

Use either SI (MKS) or CGS as primary units. (SI units are strongly encouraged.) English units may be used as secondary units (in parentheses). This applies to papers in data storage**.** For example, write “15 Gb/cm^2 (100 Gb/in^2).” An exception is when English units are used as identifiers in trade, such as “3½-in disk drive.” Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation.

The SI unit for magnetic field strength H is A/m. However, if you wish to use units of T, either refer to magnetic flux density B or magnetic field strength symbolized as $\mu_0 H$. Use the center dot to separate compound units, e.g., “A·m^2.”

V. Some Common Mistakes

The word “data” is plural, not singular. The subscript for the permeability of vacuum μ_0 is zero, not a lowercase letter “o.” The term for residual magnetization is “remanence”; the adjective is “remanent”; do not write “remnance” or “remnant.” Use the word “micrometer” instead of “micron.” A graph within a graph is an “inset,” not an “insert.” The word “alternatively” is preferred to the word “alternately” (unless you really mean something that alternates). Use the word "whereas" instead of "while" (unless you are referring to simultaneous events). Do not use the word "essentially" to mean "approximately" or "effectively." Do not use the word "issue" as a euphemism for "problem." When compositions are not specified, separate chemical symbols by en-dashes; for example, "NiMn" indicates the intermetallic compound $Ni_{0.5}Mn_{0.5}$ whereas "Ni–Mn" indicates an alloy of some composition Ni_xMn_{1-x}.

Be aware of the different meanings of the homophones "affect" (usually a verb) and "effect" (usually a noun), "complement" and "compliment," "discreet" and "discrete," "principal" (e.g., "principal investigator") and "principle" (e.g., "principle of measurement"). Do not confuse "imply" and "infer."

Prefixes such as "non," "sub," "micro," "multi," and "ultra" are not independent words; they should be joined to the words they modify, usually without a hyphen. There is no period after the "et" in the Latin abbreviation "*et al.*" (it is also italicized). The abbreviation "i.e.," means "that is," and the abbreviation "e.g.," means "for example" (these abbreviations are not italicized).

A general IEEE styleguide is available at http://www.ieee.org/web/publications/authors/transjnl/index.html

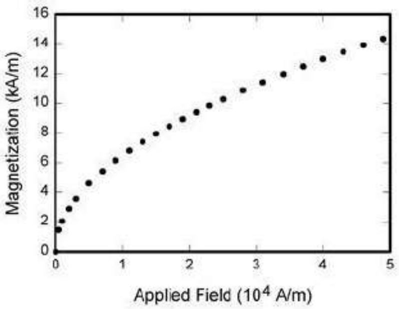


Fig. 1. Magnetization as a function of applied field.

TABLE I

Units for Magnetic Properties

Symb ol	*Quantity*	*Conversion from Gaussian and CGS EMU to SI* [a]
F	*magnetic flux*	*1 Mx ® 10^{-8} Wb = 10^{-8} V·s*
B	*magnetic flux density,* *magnetic induction*	*1 G ® 10^{-4} T = 10^{-4} Wb/m^2*
H	*magnetic field strength*	*1 Oe ® 10^3/(4p) A/m*
m	*magnetic moment*	*1 erg/G = 1 emu* *® 10^{-3} A·m^2 = 10^{-3} J/T*
M	*magnetizatio n*	*1 erg/(G·cm^3) = 1 emu/cm^3* *® 10^3 A/m*
4pM	*magnetizatio n*	*1 G ® 10^3/(4p) A/m*

VI. Guidelines for Graphics Preparation and Submission

A. Types of Graphics

The following list outlines the different types of graphics published in IEEE journals. They are categorized based on their construction, and use of color / shades of gray:

1) *Color/Grayscale figures*

Figures that are meant to appear in color, or shades of black/gray. Such figures may include photographs, illustrations, multicolor graphs, and flowcharts.

2) *Lineart figures*

Figures that are composed of only black lines and shapes. These figures should have no shades or half-tones of gray. Only black and white.

3) *Author photos*

Head and shoulders shots of authors which appear at the end of our papers.

4) *Tables*

Data charts which are typically black and white, but sometimes include color.

B. Multipart figures

Figures compiled of more than one sub-figure presented side-by-side, or stacked. If a multipart figure is made up of multiple figure types (one part is lineart, and another is grayscale or color) the figure should meet the stricter guidelines.

C. File Formats For Graphics

Format and save your graphics using a suitable graphics processing program that will allow you to create the images as PostScript (PS), Encapsulated PostScript (.EPS), Tagged Image File Format (.TIFF), Portable Document Format (.PDF), or Portable Network Graphics (.PNG) sizes them, and adjusts the resolution settings. If you created your source files in one of the following programs you will be able to submit the graphics without converting to a PS, EPS, TIFF, PDF, or PNG file: Microsoft Word, Microsoft PowerPoint, or Microsoft Excel. Though it is not required, it is recommended that these files be saved in PDF format rather than DOC, XLS, or PPT. Doing so will protect your figures from common font and arrow stroke issues that occur when working on the files across multiple platforms. When submitting your final paper, your graphics should all be submitted individually in one of these formats along with the manuscript.

D. Sizing of Graphics

Most charts, graphs, and tables are one column wide (3.5 inches / 88 millimeters / 21 picas) or page wide (7.16 inches / 181 millimeters / 43 picas). The maximum depth a graphic can be is 8.5 inches (216 millimeters / 54 picas). When choosing the depth of a graphic, please allow space for a caption. Figures can be sized between column and page widths if the author chooses, however it is recommended that figures are not sized less than column width unless when necessary.

There is currently one publication with column measurements that don't coincide with those listed above. Proceedings of the IEEE has a column measurement of 3.25 inches (82.5 millimeters / 19.5 picas).

The final printed size of author photographs is exactly
1 inch wide by 1.25 inches tall (25.4 millimeters x 31.75 millimeters / 6 picas x 7.5 picas). Author photos printed in editorials measure 1.59 inches wide by 2 inches tall (40 millimeters x 50 millimeters / 9.5 picas x 12 picas).

E. Resolution

The proper resolution of your figures will depend on the type of figure it is as defined in the "Types of Figures" section. Author photographs, color, and grayscale figures should be at least 300dpi. Lineart, including tables should be a minimum of 600dpi.

F. Vector Art

While IEEE does accept, and even recommends that authors submit artwork in vector format, it is our policy is to rasterize all figures for publication. This is done in order to preserve the figures' integrity across multiple computer platforms.

G. Color Space

The term color space refers to the entire sum of colors that can be represented within the said medium. For our purposes, the three main color spaces are Grayscale, RGB (red/green/blue) and CMYK (cyan/magenta/yellow/black). RGB is generally used with on-screen graphics, whereas CMYK is used for printing purposes.

All color figures should be generated in RGB or CMYK color space. Grayscale images should be submitted in Grayscale color space. Line art may be provided in grayscale OR bitmap colorspace. Note that "bitmap colorspace" and "bitmap file format" are not the same thing. When bitmap color space is selected, .TIF/.TIFF is the recommended file format.

H. Accepted Fonts Within Figures

When preparing your graphics IEEE suggests that you use of one of the following Open Type fonts: Times New Roman, Helvetica, Arial, Cambria, and Symbol. If you are supplying EPS, PS, or PDF files all fonts must be embedded. Some fonts may only be native to your operating system; without the fonts embedded, parts of the graphic may be distorted or missing.

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating "outline" type. This converts fonts to artwork what will appear uniformly on any screen.

I. Using Labels Within Figures

1) Figure Axis labels

Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity "Magnetization," or "Magnetization *M*," not just "*M*." Put units in parentheses. Do not label axes only with units. As in Fig. 1, for example, write "Magnetization (A/m)" or "Magnetization (A $\cdot$ m^{-1})," not just "A/m." Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)," not "Temperature/K."

Multipliers can be especially confusing. Write "Magnetization (kA/m)" or "Magnetization (10^{3} A/m)." Do not write "Magnetization (A/m) × 1000" because the reader would not know whether the top axis label in Fig. 1 meant 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 8 to 10 point type.

2) Subfigure Labels in Multipart Figures and Tables

Multipart figures should be combined and labeled before final submission. Labels should appear centered below each subfigure in 8 point Times New Roman font in the format of (a) (b) (c).

J. File Naming

Figures (line artwork or photographs) should be named starting with the first 5 letters of the author's last name. The next characters in the filename should be the number that represents the sequential location of this image in your article. For example, in author "Anderson's" paper, the first three figures would be named ander1.tif, ander2.tif, and ander3.ps.

Tables should contain only the body of the table (not the caption) and should be named similarly to figures, except that '.t' is inserted in-between the author's name and the table number. For example, author Anderson's first three tables would be named ander.t1.tif, ander.t2.ps, ander.t3.eps.

Author photographs should be named using the first five characters of the pictured author's last name. For example, four author photographs for a paper may be named: oppen.ps, moshc.tif, chen.eps, and duran.pdf.

If two authors or more have the same last name, their first initial(s) can be substituted for the fifth, fourth, third... letters of their surname until the degree where there is differentiation. For example, two authors Michael and Monica Oppenheimer's photos would be named oppmi.tif, and oppmo.eps.

K. Referencing a Figure or Table Within Your Paper

When referencing your figures and tables within your paper, use the abbreviation "Fig." even at the beginning of a sentence. Do not abbreviate "Table." Tables should be numbered with Roman Numerals.

L. Checking Your Figures: The IEEE Graphics Checker

The IEEE Graphics Checker Tool enables authors to pre-screen their graphics for compliance with IEEE Transactions and Journals standards before submission. The online tool, located at http://graphicsqc.ieee.org/, allows authors to upload their graphics in order to check that each file is the correct file format, resolution, size and color space; that no fonts are missing or corrupt; that figures are not compiled in layers or have transparency, and that they are named according to the IEEE Transactions and Journals naming convention. At the end of this automated process, authors are provided with a detailed report on each graphic within the web applet, as well as by email.

For more information on using the Graphics Checker Tool or any other graphics related topic, contact the IEEE Graphics Help Desk by e-mail at graphics@ieee.org.

M. Submitting Your Graphics

Because IEEE will do the final formatting of your paper, you do not need to position figures and tables at the top and bottom of each column. In fact, all figures, figure captions, and tables can be placed at the end of your paper. In addition to, or even in lieu of submitting figures within your final manuscript, figures should be submitted individually, separate from the manuscript in one of the file formats listed above in section VI-J. Place figure captions below the figures; place table titles above the tables. Please do not include captions as part of the figures, or put them in "text boxes" linked to the figures. Also, do not place borders around the outside of your figures.

N. Color Processing / Printing in IEEE Journals

All IEEE Transactions, Journals, and Letters allow an author to publish color figures on IEEE *Xplore*® at no charge, and automatically convert them to grayscale for print versions. In most journals, figures and tables may alternatively be printed in color if an author chooses to do so. Please note that this service comes at an extra expense to the author. If you intend to have print color graphics, include a note with your final paper indicating which figures or tables you would like to be handled that way, and stating that you are willing to pay the additional fee.

VII. Conclusion

In this report, the study of different classification techniques is presented. The different techniques such as Decision tree and Naive Bayes is explained with examples. There are errors in program ratings in traditional recommendation system, and the TV program list is affected by human emotion as well. Our TV Program Recommended system based on Data Mining reasonably gives solution to those drawbacks. According to the data collected and multi-dimensional analysis, we can find the most beneficial television broadcast playbill, and discover the hot topics. On the basis of this paper, further work on detailed data of audience's behavior can be carried on. The different standard data-sets or variable inputs are defined that may be used in experiment for this domain systems. The different data-sets identified are Score, Characters, Reviews, Likes, Comment, Category, Content, Name of the show, Language.

Acknowledgment

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

References

1. Dmitry Kit, Yu Kong, Yun Fu, "Efficient Image Geotagging Using Large Databases", IEEE Transactions on Big Data, vol. 2, no. , pp. 325-338, Dec. 2016.

2. Hatem Mousselly-Sergieh, Daniel Watzinger "World-wide scale geotagged image dataset

3. for automatic image annotation and reverse geotagging," ACM, 2014.

4. Hatem Mousselly-Sergieh, Daniel Watzinger "World-wide scale geotagged image dataset for automatic image annotation and reverse geotagging," ACM, 2014.

5. A. R. Zamir and M. Shah. Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs. TPAMI, 2014.

Design and Implementation of Mobensic Tool to aid Mobile Forensics

Shahana Shamim
Computer Engineering
Pillai College of Engineering
Mumbai, India
shahana.shamim61@gmail.com

Sumit Sharma
Computer Engineering
Pillai College of Engineering
Mumbai, India
sumithansrajsharma@gmail.com

Queeny Priyangel Srivastava
Computer Engineering
Pillai College of Engineering
Mumbai, India
queeny.priyangel97@gmail.com

Shivani Thakare
Computer Engineering
Pillai College of Engineering
Mumbai, India
thakareshivani3@gmail.com

Madhumita Chatterjee
Computer Engineering
Pillai College of Engineering
Mumbai, India
mchatterjeee@mes.ac.in

***Abstract*- Mobile phones have become an integral part of our daily lives. Today it is difficult to think of a life without a mobile phone because it is not only a phone but also a calculator, camera, computer, email, a storehouse of information, PlayStation and a music system too. But the advancement of mobile has led to a subsequent increase in the rate of cyber crimes through mobiles. Mobile forensics is used to detect and analyze any malicious activity that might have been performed using the device. Our objective is to help reduce the criminal activities by creating a toolkit to aid mobile forensics for android devices. Currently, there is no single compiled tool available to perform mobile forensics, hence we propose to design a toolkit for the same. The process of mobile forensics includes three major steps, image acquisition, data extraction and data analysis. The toolkit will help to create an image of the entire device, extract deleted and hidden files and perform analysis of video, audio and multimedia files.**

***Keywords* - Android Live Imaging, Android Debug Bridge, Kali Linux, Mobile Forensics, Rooting, Forensic Toolkit Imager, Autopsy.**

I. Introduction

Digital Forensics is the process of uncovering and interpreting electronic data. The goal of the process is to preserve any evidence in its most original form while performing a structured investigation by collecting, identifying and validating the digital information for the purpose of reconstructing past events.

The context is most often for usage of data in a court of law, though digital forensics can be used in other instances.

The term "forensics" implies that digital forensics is used to recover evidence to be used in the court of law against some offender. This is very useful to detect corporate frauds, perhaps an employee stole a valuable data or even for the analysis of mobiles recovered at a crime site. The contents of the device, like chats, images etc. can be used to provide evidence against such crimes.

Mobile forensics is a branch of digital forensics which deals with the recovery of digital evidence or data from a mobile device under forensically sound conditions. The use of mobile phones/devices in crime has widely increased for few years, but the forensic study of mobile devices is a new field, from the early 2000s. There are various challenges that are faced while recovering data from mobile due to many reasons. To remain competitive the manufacturers change the original equipment file structures, data storage etc. and hence forensics examiner has to find out alternative ways than used in computer forensics. The storage capacity of devices grows continuously. These are some of the challenges faced in mobile forensics.

Kali Linux is a Debian-derived Open Source Linux distribution designed for digital forensics and penetration testing. It is maintained and funded by Offensive Security Ltd. Kali has more than 600 penetration testing tools along with multi-language support. The Kali Linux operating system is completely customizable all the way down to the kernel and is developed in a secure environment. It is specifically tailored to the needs of penetration testing professionals, thus providing a secure environment to carry out various forensic activities.

Android is a mobile-based operating system developed and maintained by Google. It is based on modified version of the Linux operating system and other open source software. Android is available for devices such as smartphones and tablets. Google has also developed Android TV for television and Android Wear for wrist watches. There are various versions of Android available ranging from earliest Gingerbread (2.3) to the latest Oreo (8.0).

II. LITERATURE SURVEY

In paper[1] The Author gives us a tool to extract data from memory card and analysis of WhatsApp application installed on the memory card from different models of mobile phone. There are many mobile forensics tools that can retrieve information from both internal and external memory. Because of the complexity of using different forensics tools and processing time, there is a requirement of one tool that automates the process. The methods followed are File Extraction, File Recovering, File Converting and Decrypting and Reporting and GUI.

In File Extraction, the input to the tool is disk image file and OS relevant file categories will be extracted like pictures, video, audio, and documents.

In File Recovering process the deleted files are extracted and recovered files are sorted in various categories.

In File Converting and Decrypting the audio, video, thumb files containing pictures and additional information and WhatsApp databases are decrypted into a readable format.

The last method which is Reporting and GUI offer UI and final report to the investigator.

In paper[2] the author proposes a solution to the anti-forensic technique of steganography by designing and developing an application that will detect the presence of stegno data within the Android device and then perform logical data acquisition of images, audio, and videos. The application proposed by the author that is Mobile forensic Analyser is developed with the hash function and buttons like extract and report. The analysis of stegno data will be in png, mp3, mp4. The tool is also used for detecting hidden data on an image, audio, video. It maintains the integrity of data by using strong tools like hash.

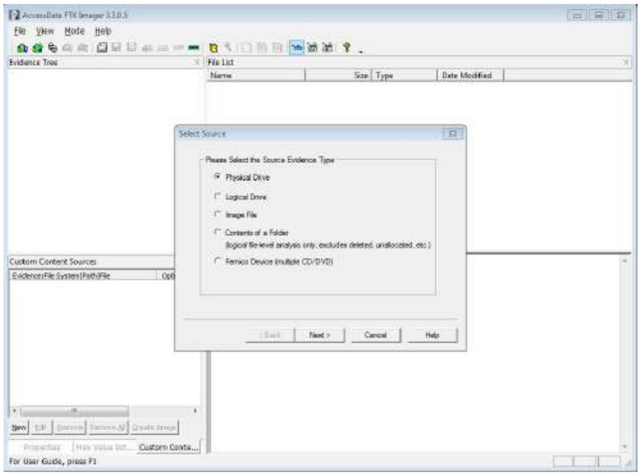
The authors of the paper[3] have proposed file signature analysis which is used to detect if the file extension has tampered or not. The two methods used by them are multimedia file signature acquisition in which they have extracted and compared multimedia file signature of different mobile phones using hex editor, whereas in second method that contents inspection there are two steps the first step is similar to the above and the second step is to compare content and metadata of original and amended multimedia files in order to detect changes. The results obtained by the authors after smartphone multimedia file signature analysis on camera images examined has a file extension .jpg. The camera videos file extension observed are .mp4 (Samsung, Blackberry, Lenovo, Nokia) and .mov. The audio file extensions examined are .wav(Samsung, Nokia), m4a(iPhone) and .amr(Blackberry and Lenovo). The results obtained after content examination for camera images/videos/audio contains metadata which has information such as a timestamp(creation time and date) and company name (manufacturer name, device name, OS).

The content examination of application video obtained multimedia files extracted from WhatsApp have different file extension such as .jpg, .mp4, .mov etc.

III. EXISTING SYSTEM

Mobile forensic is a vast field with a lot of exploration that needs to be performed. The number of mobile phones keeps on increasing day by day with newer versions of a certain phone being released biannually. This has led to an increase of data being produced in a day, this has, in turn, led to increasing of cybercrime at an alarming rate ultimately resulting in a high demand for a complete mobile forensic tool. Currently, there are some tools available for performing image creation process like FTK Imager and for analysis of the created image like Autopsy.

FTK Imager is a Forensic Toolkit Imager which is distributed by AccessData used for forensic imaging. It is a commercial software package. FTK Imager is often used for creating images of disks and portable devices**.** This image is stored as a single file or as segments that may later be reconstructed to obtain the full disk image. It offers MD5 hash calculation and hence confirms the integrity of the data. The resulting image file can be saved in several formats including the DD raw format.



Select Source

Please Select the Source Evidence Type

● Physical Drive

○ Logical Drive

○ Image File

○ Contents of a Folder
(logical file-level analysis only; excludes deleted, unallocated, etc.)

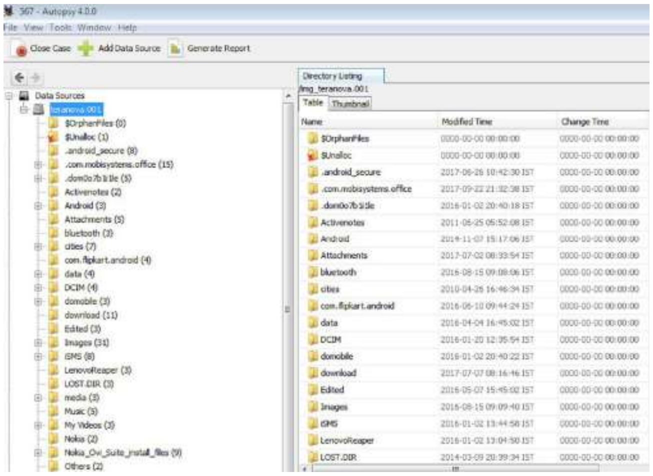
○ Fernico Device (multiple CD/DVD)

For User Guide, press F1

Fig. 1. Forensic toolkit Imager.

An autopsy is a computer software that is used for the forensic analysis process, making it easier for the investigators to carry out their analysis in a secure and efficient manner. This tool is designed with three principles in mind: extensible, framework and ease of use. Extensibility states that the user should be able to add new functionality that can analyze the underlying data source. Frameworks offer standard approaches for investigation, analysis, and reporting. Ease of use makes it easier for users to repeat their steps without reconfiguration.

To initiate the process of analysis we provide the image of the concerned device to the tool in formats such as dd, raw etc. The autopsy software then begins the analysis process, segregating the files on the image into various suitable formats such as documents, multimedia, deleted as well as emails etc. The autopsy GUI provides a simple way to access, analyze and extract the files that are required by the forensic expert.



Data Sources
teranova.001
$OrphanFiles (0)
$Unalloc (1)
.android_secure (8)
.com.mobisystems.office (15)
.dom0o7b3i3le (5)
Activenotes (2)
Android (3)
Attachments (5)
bluetooth (3)
cities (7)
com.flipkart.android (4)
data (4)
DCIM (4)
domobile (3)
download (11)
Edited (3)
Images (31)
iSMS (8)
LenovoReaper (3)
LOST.DIR (3)
media (3)
Music (3)
My Videos (3)
Nokia (2)
Nokia_Ovi_Suite_install_files (9)
Others (2)

/img_teranova.001

Name	Modified Time	Change Time
$OrphanFiles	0000-00-00 00:00:00	0000-00-00 00:00:00
$Unalloc	0000-00-00 00:00:00	0000-00-00 00:00:00
.android_secure	2017-06-26 10:42:30 IST	0000-00-00 00:00:00
.com.mobisystems.office	2017-09-22 21:32:38 IST	0000-00-00 00:00:00
.dom0o7b3i3le	2016-01-02 20:40:18 IST	0000-00-00 00:00:00
Activenotes	2011-06-25 05:52:08 IST	0000-00-00 00:00:00
Android	2014-11-07 15:17:06 IST	0000-00-00 00:00:00
Attachments	2017-07-02 08:33:54 IST	0000-00-00 00:00:00
bluetooth	2016-08-15 09:08:06 IST	0000-00-00 00:00:00
cities	2010-04-26 16:46:34 IST	0000-00-00 00:00:00
com.flipkart.android	2016-06-10 09:44:24 IST	0000-00-00 00:00:00
data	2016-04-04 16:45:02 IST	0000-00-00 00:00:00
DCIM	2016-01-20 12:35:54 IST	0000-00-00 00:00:00
domobile	2016-01-02 20:40:22 IST	0000-00-00 00:00:00
download	2017-07-07 08:16:46 IST	0000-00-00 00:00:00
Edited	2016-05-07 15:45:02 IST	0000-00-00 00:00:00
Images	2016-08-15 09:09:40 IST	0000-00-00 00:00:00
iSMS	2016-01-02 13:44:58 IST	0000-00-00 00:00:00
LenovoReaper	2016-01-02 13:04:50 IST	0000-00-00 00:00:00
LOST.DIR	2014-03-09 20:39:34 IST	0000-00-00 00:00:00

Fig. 2. Autopsy.

Thus we observe that even when we have such tools available for forensic analysis in the market, these do not provide a complete tool to carry out the process of forensic analysis. Each tool provides the functionality to perform one part of the complete task. Hence we propose a complete mobile forensic analysis toolkit called as **Mobensic**. This will help us in performing the various tasks of image creation and data analysis in one platform itself.

IV. Proposed System

As we have observed that from the existing tools available for mobile forensics the procedure to get all the usable information from the internal memory of the mobile phone is a time-consuming process. There is a need for developing a single tool that simplifies the forensic process. So we propose to design a single toolkit to aid mobile forensics and simplify the investigation of internal memory of the mobile phone. The important thing is that with the help of new toolkit digital investigators can start with the investigation without searching all kinds of tools. Proposed tool will be user-friendly, simple and time-saving. The Mobensic tool works in the Kali Linux environment. The entire process from image creation to the analysis and report generation will be provided by a single tool which will make the process of collecting evidence from the mobile phone much easier.

Figure (a), gives the architecture of our proposed **Mobensic Tool.** It includes the process of creating an image of the mobile device, extracting the required data from the created image and finally performing analysis on the data extracted. Once the data analysis is completed, a detailed report of the entire forensic process is generated for the expert to view.

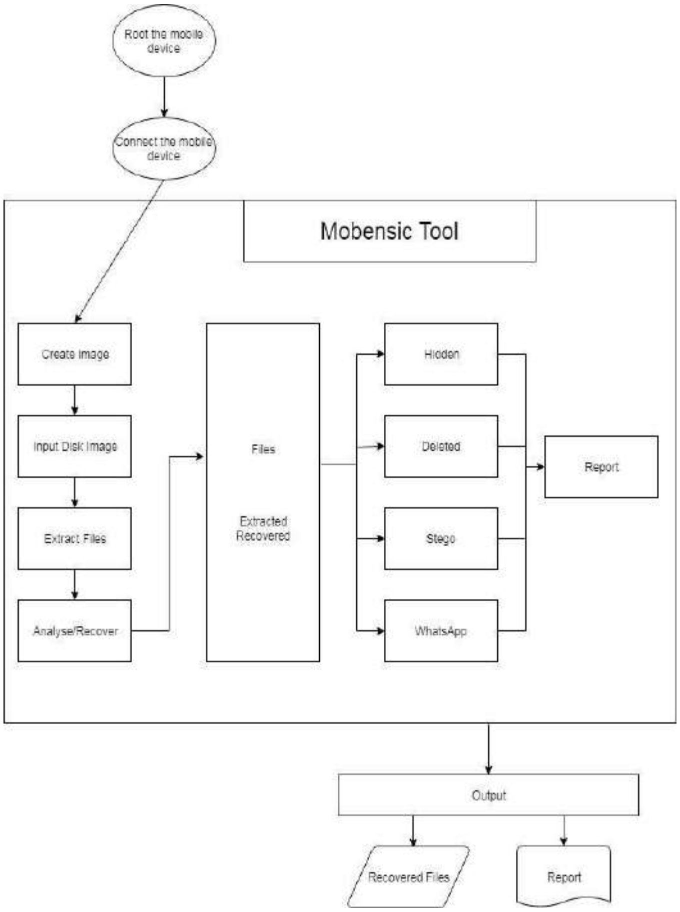


Fig. 1. The architecture of Proposed System.

1. *Rooting the device:*

The process of rooting allows the user of smartphones, tablets and other devices running on the Android operating system to gain root access to the android subsystems. The Android operating system uses the Linux kernel and hence rooting gives similar administrative permissions as on Linux or any other Unix like operating system.

For the designing of Mobensic toolkit, a Moto G 3rd Generation device running on Android OS version 6.0.1 was used. For the Moto G 3rd generation device, first, unlock the bootloader on the device(if locked) and install the necessary device drivers. Next, install ADB and Fastboot tools along with the latest version of SuperuserSu and TWRP manager.Now make use of the necessary drivers and tools to root the device and attain administrative(Superuser) access.

However, the rooting process may not be the same for each and every device. It may vary depending on the device in consideration as well as the Android OS running on the device.

2. *Image creation*

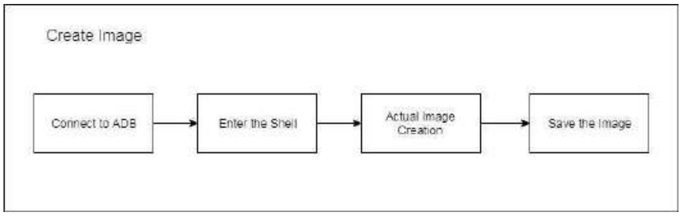


Fig. 2. Internal working of Image creation.

Figure 2, further elaborates the image creation module from the proposed architecture. To create an image of the mobile device, very first step is to activate the write blocker function. Write blocker is a function that will disable all the write access rights on the device, making sure that the device and its contents have not been tampered with. A write blocker will help the forensic expert to prove that the device and its contents have not been manipulated, which is a very important aspect in the court of law to use a mobile device as a proof.
After write blocker has been activated, the forensic expert now connects the device to the toolkit using Android Debug Bridge (ADB). The user now enters the ADB Shell. In the shell, perform the actual function of creating the image using Android live Imaging process. This process creates a complete image of the internal memory of the device. The image is then saved for further analysis.

3. *Data extraction*

Once the image of the entire device has been created, move towards extraction of data from the image. The data extracted is stored in a folder format for easy retrieval and analysis. Mobensic tool will be able to extract the hidden files from, stegno data files, deleted files and also the WhatsApp conversation details from the device.

4. *Data analysis and Reporting*

After performing the action of data extraction, the expert will need to analyze the data extracted. This will be done in the data analysis and the reporting module of the tool. The forensic expert will be able to classify and analyze the data into different formats like Whats App data, stegno data, multimedia files, and Documents.
The toolkit will further also generate a report on the data that is extracted and classified.

V. RESULTS ACHIEVED

In this section, we deploy our Mobensic tool for analysis and testing. It is difficult to build one tool that can perform all Forensic process as mentioned in section III. This Mobensic tool can simplify the process by integrating all Forensic steps in one single tool.In this section, we test Mobensic tool by analyzing internal memory of mobile devices.

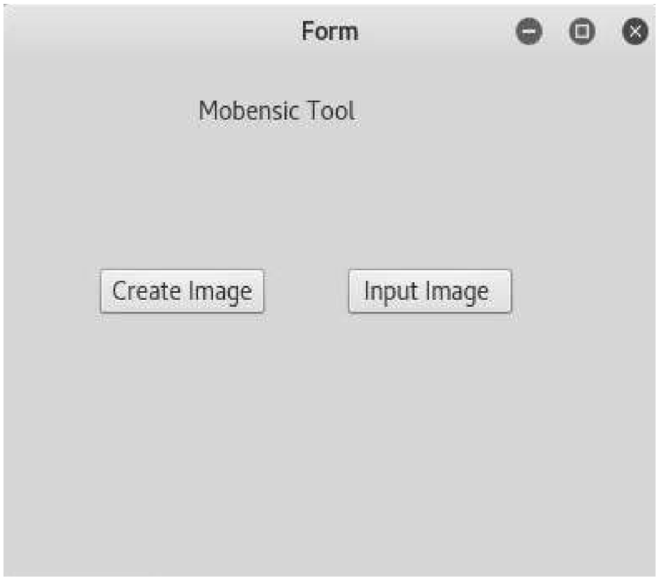
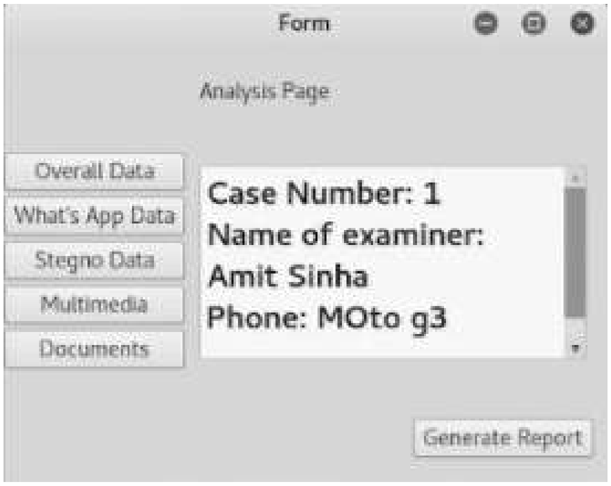


Fig. 1. GUI for Mobensic Tool.

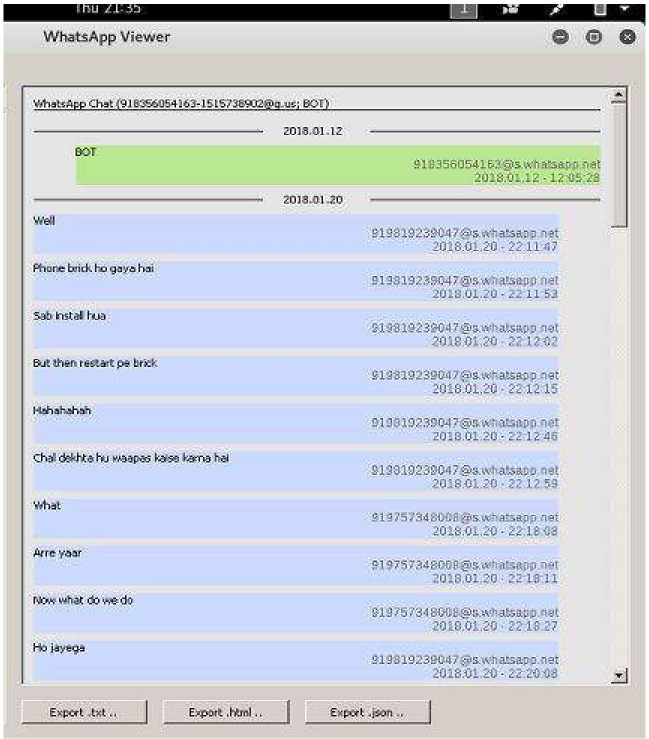
The toolkit provides two options one is to create an image of the internal memory of the mobile device or to directly input the image of the mobile device. In creating image option the image of the mobile device connected is created and stored on your machine whereas in input image option the image of the device is loaded for further analysis.



Case Number: 1
Name of examiner:
Amit Sinha
Phone: MOto g3

Fig. 2. Analysis screen for Mobensic Tool.

The figure 2 above shows the analysis screen where the input image is analyzed and the data which is recorded is classified as Whatsapp data, Stegno data, Multimedia, Documents. The toolkit also provides a report generation option for a summary of all the extracted data. Now from this, the user can click on any of the options to view and analyze the various data extracted.



WhatsApp Chat (918356054163-1515738902@g.us; BOT)

2018.01.12

BOT
918356054163@s.whatsapp.net
2018.01.12 - 12:05:28

2018.01.20

Well
919819239047@s.whatsapp.net
2018.01.20 - 22:11:47

Phone brick ho gaya hai
919819239047@s.whatsapp.net
2018.01.20 - 22:11:53

Sab install hua
919819239047@s.whatsapp.net
2018.01.20 - 22:12:02

But then restart pe brick
919819239047@s.whatsapp.net
2018.01.20 - 22:12:15

Hahahahah
919819239047@s.whatsapp.net
2018.01.20 - 22:12:46

Chal dekhta hu waapas kaise karna hai
919819239047@s.whatsapp.net
2018.01.20 - 22:12:59

What
919757348008@s.whatsapp.net
2018.01.20 - 22:18:08

Arre yaar
919757348008@s.whatsapp.net
2018.01.20 - 22:18:11

Now what do we do
919757348008@s.whatsapp.net
2018.01.20 - 22:18:27

Ho jayega
919819239047@s.whatsapp.net
2018.01.20 - 22:20:08

Fig. 3.Whatsapp Viewer.

The above figure 3 shows Whatsapp Viewer which display the Whatsapp chats which were recovered during the analysis phase. When the user clicks on "WhatsApp data" option the conversations stored in the mobile device are displayed to the user.

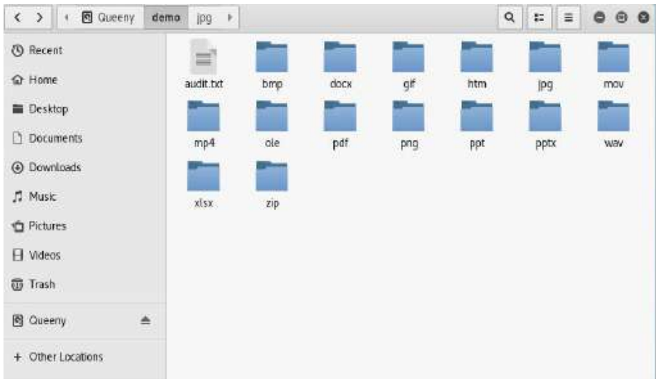
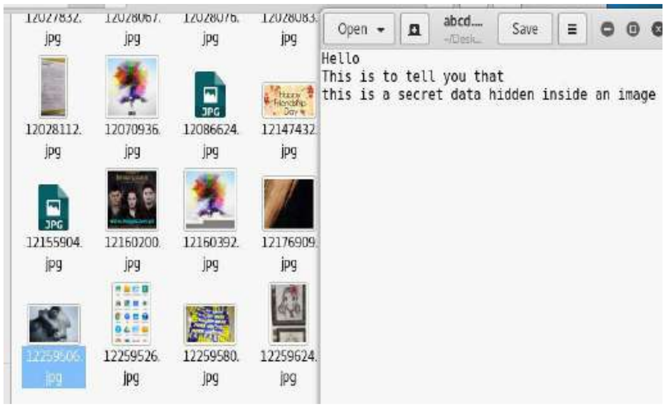


Fig. 4.Data Recovered

Figure 4 above shows the classification output in the extraction of the data from the image of the device. Once the user clicks on the "Overall data" option, the tool gives a complete view of the various sub-folders containing data like jpg files, png files, pdf files, text files etc. which have been recovered in the extraction module.



```
Hello
This is to tell you that
this is a secret data hidden inside an image
```

Fig. 5.Stegno Image.

The figure 5 above depicts an example of the stegno image that has been extracted using the Mobensic tool. When the user clicks on “Stegno data” option, the stegno image stored in the mobile device along with its hidden text is recovered by the tool and displayed to the expert.

VI. Conclusion

In the past decade, advancement in technology has made us more and more dependent on our mobile devices for day to day activities. This in turn has led to an increase in the number of frauds and malicious activities being performed with the help of the mobile phones. A tool like ours can help in analyzing the matter and further reach conclusions.

Mobensic tool can be used in a vast frame of applications like,

- Military intelligence
- Corporate investigations
- Private investigations

- Criminal and civil defense
- Electronic discovery

In future, the Mobensic tool can be further be enhanced to extract and analyze Call Logs, Contact Information, text messages, and Email. Further, the toolkit can also be available for other operating systems like iOS. The rooting process can also be incorporated into the toolkit, making the process even easier for the forensic expert.

VII. Acknowledgment

We would like to take this opportunity to express our profound gratitude and deep regard to Prof. Dr. Madhumita Chatterjee for her guidance and constant encouragement throughout the course of this project. We are immensely obliged for her cordial support, supervision and providing necessary information.

We remain immensely obliged to Dr. Madhumita Chatterjee for introducing this topic, and for her invaluable support in garnering resources for us either by way of information or computers also her guidance and supervision which made this project happen. We are thankful to our college, Pillai College of Engineering for providing us healthy competitive environment and outstanding educational facilities that played an important role in keeping us highly motivated to achieve our goals.

References

[1] Rob Witteman, Arjen Meijer, Toward a new Tool to Extract the Evidence from a Memory Card of Mobile Phones, 2016, School of Computer Science, University of Dublin, Ireland

[2] Walter T. Mambodza, Nagoor Meeran A.R, Android Mobile Forensic Analyzer for Stegno Data, 2015, Department of Information Technology, SRM University

[3] T. Baker, B. Shah, Multimedia File Signature Analysis for Smartphone Forensics, 2016, Department of Computer Science, Liverpool John Moores University, UK

[4] Neha S Thakur, Forensic Analysis of WhatsApp on Android Smartphones, 2013, Master of Science in Computer Science Information Assurance University of Pune

[5] Mark Lohrum, Live imaging an Android device, 2014, http://freeandroidforensics.blogspot.in/2014/08/live-imaging-android-device.html

[6] Qt Designer, Qt Designer Manual (Documentation Archives)http://doc.qt.io/archives/qt-4.8/designer-manual.html

[7] Ajinkya, How to install TWRP and root Motorola Moto G 3rd Gen(2015) https://devsjournal.com/how- to- install-twrp -root-motorola-moto-g-3rd-gen.html

[8] Ajinkya, How to easily unlock bootloader in Moto G 3rd Gen(2015)https://devsjournal.com/how-to-easily-unlock-bootloader-in-moto-g-3rd-gen-2015.html

[9] Satish Bommisetty, Rohit Tamma, Heather Mahalik, Practical Mobile Forensics, 1st ed, 2014, Livery Place, 35 Livery Street, Birmingham B3 2PB, UK

[10] Kevin Mandia, Chris Prosise, Matt Pepe, Incident Response and Computer Forensics, 2nd ed, 2014, McGraw-Hill, Inc. New York

[11] Andrew Hogg, Android Forensics Investigation, Analysis and Mobile Security for Google Android, 1st ed, 2011,Oak Park Illinois,USA