VIGYAAN'20

NATIONAL CONFERENCE

ON

INTERNET OF THINGS

Date : Thursday, 20 February 2020

Venue : NIMIT, Pongam, Thrissur

PROCEEDINGS

Naipunnya
To reach the unreachable

Organised by :

Department of Computer Science
NAIPUNNYA INSTITUTE OF MANAGEMENT AND INFORMATION TECHNOLOGY (NIMIT)

VIGYAAN-2020

Vigyaan-2020

The Conference Proceedings-*"Internet of Things"*

Manager

Fr.(Dr).Paul Kaithottungal
Executive Director & Principal
Naipunnya Institute of Management and Information Technology

Editor

Ms.Emily Ittiachan
Vice Principal
Naipunnya Institute of Management and Information Technology

Editorial Advisory Council

Mr. Jayakrishnan S
Head of the Department
Department of Computer Science
Naipunnya Institute of Management and Information Technology

Editorial Board

Dr. Sarika S
Assistant Professor
Department of Computer Science
Naipunnya Institute of Management and Information Technology
Mr. Fredy Varghese
Assistant Professor
Department of Computer Science
Naipunnya Institute of Management and Information Technology
Ms. Siji Jose
Assistant Professor
Department of Computer Science
Naipunnya Institute of Management and Information Technology

Editorial and Administrative Office
Naipunnya Institute of Management and Information Technology
Pongam,Korraty East,Trissur,Kerala-680 308,Ph:0480 2730340,2730341
Web: www.naipunnya.ac.in,Email:mail@ naipunnya.ac.in

ISBN-978-93-5406-047-2

CONTENTS

Messaging Protocols for IoT Systems: MQTT, CoAP and HTTP — A Comparative Study

Gopikrishna P.B.
MSc. Computer Science
St. Thomas College (Autonomous), Thrissur
Email: gopikrishnapb8@gmail.com

Dr. Jiju A. Mathew
Assistant Professor
St. Thomas College (Autonomous), Thrissur
Email: jijuamathew@gmail.com

***Abstract*—Internet of Things (IoT) refers to a self-configuring wireless network which enables the Internet to reach out into the real world of physical objects. It is the need of the hour to communicate remotely with the day-to-day interacting appliance using a portable device like a smartphone with internet connectivity and was made possible through IoT. Many messaging protocols — for exchanging messages between the connected devices; are used in IoT devices at various layers of the OSI model. These are set of rules, formats and functions which work on the application layer of the OSI model to exchange data in a structured and meaningful way and its usage is based on the type of application and its functionality that the system demands. MQTT, XMPP, DDS, AMQP and CoAP are a few of the widely used messaging protocols for the IoT application layer. In this paper, we present an evaluation of three established messaging protocols *viz.* MQTT, CoAP and HTTP for IoT systems. The characteristics and working of each protocol is compared; based on this evaluation, the user can decide their appropriate usage in various IoT systems according to their requirements and suitability.**

I. Introduction

The Internet of Things is an intelligent connectivity of physical devices embedded with Internet connectivity, sensors and other hardware that allow communication and real time working of devices which can be controlled through the web. Nowadays, there is a range of objects around us that are capable of collecting, sending and processing data to the other servers and other applications. IoT protocols enable to exchange data in a structured and meaningful way [1]. These messaging protocols are modes of communication that protect and ensure optimum security to the data being exchanged between connected devices online. This is one of the reasons why the IoT needs standardized protocols and it will transfer data only when the communication network between the two connected devices is safe. The important characteristics of messaging protocols are; Speed — Amount of data that can be transferred/second, Latency — amount of time a message takes to be transferred, Power consumption, Security and Availability of software stacks. Messaging protocols are implemented in the application layer of the OSI model in order to transmit messages between the devices which were founded on TCP and UDP to solve the communication challenges faced in an IoT project. The TCP protocol enables the XMPP, MQTT and REST/HTTP messaging protocols and UDP protocol enables DDSI, CoAP *etc.*

MQTT is one of the most commonly used protocols in IoT projects and stands for Message Queuing Telemetry Transport. It is designed as a lightweight messaging protocol which uses publish/subscribe operations to exchange data between clients and the server [2]. Constrained Application Protocol (CoAP) is a specialized web transfer protocol which uses constrained nodes and constrained networks in the IoT and is designed to enable simple, constrained devices to join the IoT even through constrained networks with low bandwidth and low availability [3]. HTTP is a world wide web protocol which works on the basis of client/server system. The future of the IoT lies on these several messaging protocols and a single protocol cannot deal with all the possible IoT use cases. Consequently, it is necessary to discuss the pros and cons of the widely accepted and emerging messaging protocols for IoT systems to find their best-fit scenarios. Therefore, this paper presents an evaluation of the three messaging protocols: MQTT, CoAP and HTTP. These three are the most promising IoT protocols to establish efficient communication between devices. The characteristics of each protocol is compared and its evaluation helps the user to choose appropriate messaging protocol according to their IoT application.

MQTT — Message Queuing Telemetry Transport

MQTT is a lightweight client-server publish/subscribe messaging transport protocol which runs over TCP. It is open and simple designed so as to be easy to implement and efficient to use in environments for communication in Machine to Machine(M2M) and IoT. MQTT was introduced in 1999 by AndyStanford-Clark of IBM and Arlen Nipper of Arcom Control Systems Ltd (Eurotech) [2]. MQTT is a binary protocol with low bandwidth with 2 bytes of fixed header. MQTT connection typically involves two kinds of agents: MQTT clients and MQTT server. MQTT server is known as MQTT broker. Data that is being transported by MQTT is referred to as an application message. A MQTT client is a device or program that is connected to the network to exchange application messages through MQTT. It can either be publisher: who can publish application messages or subscriber
: who requests for the application messages. A device or program that interconnects the MQTT clients is called MQTT server/broker and it accepts and transmits the application messages among multiple clients connected to it. Devices such as sensors, mobiles *etc.* which are considered as MQTT clients

ISBN -978-93-5406-047-02

have certain information to broadcast and publish the data to the MQTT server [2]. MQTT broker is responsible for data collection and organization. The application messages that are published by MQTT clients are forwarded to other MQTT clients that subscribe to it. Another great feature of MQTT is its three levels of Quality of Service (QoS) for reliable delivery of messages and control packets are exchanged based on this QoS associated with them before the transmission of the application messages [7]. An MQTT control packet consists of a fixed header, a variable header and small payloads upto maximum size of 256 MB. CONNECT, CONNACK, PUBLISH, PUBACK, PUBREC, PUBREL, SUBSCRIBE, SUBACK, *etc.* are some of the MQTT control packets exchanged between MQTT clients and the broker [4]. Routing information is provided by" Topic" in MQTT. Each topic has a topic name and topic levels associated with it. MQTT server buffers all the messages if the client is offline and delivers them to the client when the session is enabled.

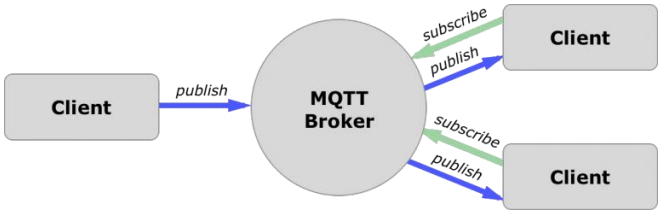


Fig. 1. MQTT

CoAP — Constrained Application Protocol

The Constrained Application Protocol (CoAP) is a specialized web transfer protocol with constrained nodes and constrained networks in the IoT systems. The protocol is designed for machine-to-machine (M2M) applications such as smart energy and home automation with a header size of 4 bytes. CoAP is one of the latest application layer protocol; created based on the wildly successful REST model [3]. It was developed by IETF in 2014. CoAP is a protocol that runs over the UDP in IoT and can be used with other protocols such as DTLS protocol to secure it. Several methods have been used to secure the CoAP protocol; following is a list of these methods, Firstly DTLS protocol with CoAP protocol, and in this case, it is responsible for the key management, for the confidentiality of the data and for the data authentication. A second choice is using IPsec with CoAP protocol when using it without DTLS (no security) it could be used with IPsec with appropriate configuration, the restricted devices may use the built-in hardware encryption in the link-layer [8]. CoAP protocol can be improved by adding a hash function to it, in order to choose the appropriate hash function to CoAP protocol, a number of hash functions are used. SHA-1, SHA-224, SHA-256 are some of the hash functions used. Each CoAP message has a unique ID; this is useful to detect message duplicates. A CoAP message is built by a binary header, a compact option and a Payload. When exchanging messages between two endpoints, these messages can be reliable which is known as Confirmable message (CON) and this message tells the client that the message will arrive at the server. A Confirmable message is sent again and again until the other party sends an acknowledge message (ACK) and it contains the same ID of the Confirmable message (CON). If the server has troubles managing the incoming request, it can send back a Rest message (RST) instead of the Acknowledge message (ACK) [8]. The other message category is the Non-confirmable (NON) messages and these are messages that don't require an Acknowledge by the server. They are unreliable messages or in other words messages that do not contain critical information that must be delivered to the server. The messages that contain values read from sensors are Non-confirmable messages. Even if these messages are unreliable, they have a unique ID. The CoAP Request/Response is the second layer in the CoAP abstraction layer. The request is sent using a Confirmable (CON) or Non-Confirmable (NON) message [8]. There are several scenarios depending on if the server can answer immediately to the client request or the answer if not available. If the server can answer immediately to the client request, then if the request is carried using a Confirmable message (CON), the server sends back to the client an Acknowledge message (ACK) containing the response or the error code. CoAP offers more functionality than MQTT such as it supports content negotiation to express a preferred representation of a resource; this allows client and server to evolve independently, adding new representations without affecting each other.

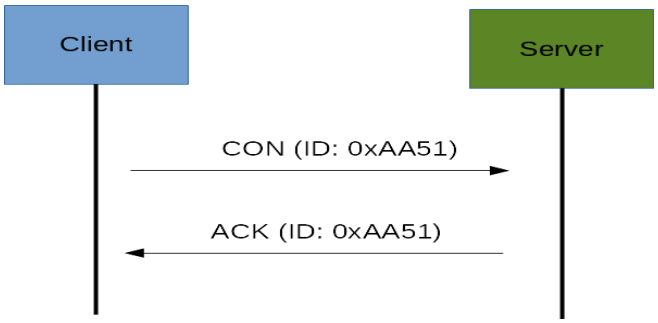


Fig.2.CoAP

HTTP- HyperText Transfer Protocol

HTTP is a messaging protocol used by the World Wide Web which defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands [7]. It is a stateless protocol because each command is executed independently, without any knowledge of the commands that came before it. Tim Berners-Lee and his team at CERN are credited with inventing the original HTTP, along with HTML and the associated technology for a web server and a text-based web browser [7]. Berners-Lee first proposed the "WorldWideWeb"; project in 1989. Later, it was developed by IETF with W3C and first published as a standard protocol in 1997. The request/response Web architecture and Web browser of HTTP initiates a request to a server, typically by opening a TCP/IP connection [8]. Each request consists of a request line, a set of request headers, and an entity while the server sends a response that consists of a status line, a set of

ISBN -978-93-5406-047-02

response headers, and an entity. The entity in the request or response can be thought of simply as the payload, which may be binary data or ASCII characters. The browser or the server may terminate the TCP/IP connection, when the response has been completed. Either and the browser can send another request and it denotes the use of HTTP with SSL (Secure Socket Layer) protocol or its successor protocol Transport Layer Security (TLS), a transport-layer protocol. A secure connection between two machines can be established using encryption. HTTP is a globally accepted web messaging protocol, which offers several features such as persistent connections, request, and chunked transfer encoding. The H TTP request uses complex header format s of TCP with 9 packets which are not required in most cases of IoT system and this creates an unnecessary waste of resources. HTTP is used commonly where the data is triggered by the client for example weather reporting, pollution status, *etc.* on time basis.

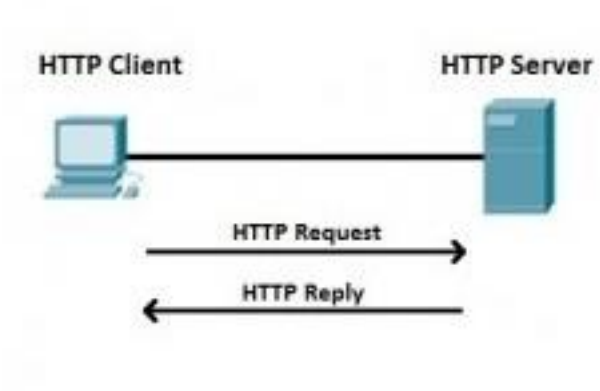


Fig3.HTTP

Analysis of MQTT, CoAP and HTTP

Message protocols are used to enhance the communication in IoT systems. MQTT is a Publish/Subscribe system while HTTP is a Request/Response System while CoAP can establish both Publish/Subscribe and Request/Response. CoA P is developed recently so that it can rectify the drawbacks of MQTT protocol. HTTP can send large amounts of data compared to MQTT and CoAP. CoAP provides its own reliability mechanism than MQTT with the use of "messages" and "Non- messages". MQTT is data centric whereas HTTP is document-centric. HTTP is request-response protocol for client-server computing but not always optimized for mobile devices. Many benefits of MQTT in these terms are transfers data as a byte array and publish/subscribe model, which makes it perfect for resource-constrained devices and help to save battery whereas CoAP support for content negotiation and discovery allowing devices to probe each other to find ways of exchanging data. Due to the efficient architecture CoAP has a higher performance ratio than MQTT and HTTP.

Comparison

Table1.Comparison Table

Criteria	MQTT	CoAP	HTTP
Year	1999	2014	1990
Architecture	ient/Broker	Client/Broker Client/Server	Client/Server
Abstraction	Publish/Subsc ribe	Request/Response Publish/Subscribe	Request/Response
Header Size	2 Byte	4 Byte	Undefined
Message Size	Small and Undefined	Small and Undefined	Large and Undefined
RESTful	No	Yes	Yes
Transport Protocol	TCP	UDP, SCTP	TCP
Security	TSL, SSL	DTLS, IPSec	TLS/SSL
Encoding Format	Binary	Binary	Binary

Conclusion

This paper presented an evaluation of three messaging protocols in IoT. MQTT, CoAP and HTTP. A detailed study of all these protocols are done and their characteristics such as Year, Architecture, Abstraction, Message size, Header size, Security etc. are compared. MQTT Protocol is easy to use and it is used when response time, throughput, lower battery and bandwidth usage of an IoT system are in the first place. It's also perfect in case of intermittent connectivity. Even though HTTP is worthy and extendable, MQTT is more suitable when it is referred to IoT development and does its best as a communication bus for live data. MQTT messages can be used for any purpose, but all clients must know the message formats upfront, in order to allow communication. CoAP, conversely, provides inbuilt support for content negotiation and discovery allowing devices to probe each other to find ways of exchanging data. It is best suited to a state transfer model and it is not purely event based. MQTT allows for persistent connections which can save significant resources over HTTP which is most relevant if you are using SSL. Analogous to CoAP, HTTP uses Universal Resource Identifier (URI) instead of topics. When small metrics are transmitted, MQTT will generally be more bandwidth efficient than HTTP. Messaging protocols are implemented in IoT to ensure security and efficient data transmission. Even thou choosing appropriate messaging protocol totally depends upon the type of application; MQTT can be used efficiently in IoT systems with low

ISBN -978-93-5406-047-02

bandwidth and supports many to many communications while CoAP and HTTP is request-response protocol for communication and not always optimized for IoT systems.

REFERENCES

[1]Performance evaluation of IoT protocols under a constrained wireless access network
https://ieeexplore.ieee.org/abstract/document/7496622
[2]A survey on MQTT: A protocol of Internet of Things(IoT)
https://www.researchgate.net/profile/Dipa_Soni/publication/316018571_A_SURVEY_ON_MQTT_A_PROTOCOL_OF_INTERNET_OF_THINGSIOT/links/58edafd4aca2724f0a26e0
[3]CoAP: An Application Protocol for Billions of Tiny Internet Nodes
https://ieeexplore.ieee.org/document/6159216
[4]A Token-based Protocol for Securing MQTT Communications
https://ieeexplore.ieee.org/document/8555834
[5]oneM2M Architecture Based Secure MQTT Binding in OS
https://ieeexplore.ieee.org/document/8802473
[6]Enhance the Security in Smart Home Applications based on IOT-CoAP protocol
https://ieeexplore.ieee.org/document/8357000
[7]Comparison with HTTP and MQTT In Internet of Things (IoT)
https://ieeexplore.ieee.org/document/8597401
[8] Internet of Things (IoT) with CoAP and HTTP Protocol: A Study on Which Protocol Suits IoT in Terms of Performance

ISBN -978-93-5406-047-02

Role of Artificial Intelligence and Machine Learning in the Current era

Jamshiya Jamaludheen

Computer science department,Ansar women's college ,perumpilavu,kerala,*india*

jamshiyak8@gmail.com

Abstract-**Artificial Intelligence is one of the most influential and powerful technologies in today's world. We are far from seeing its full potential. The future of our globe will be far different from the current state with the progress in the research field of AI. This article is based on the study conducted to know the basic concepts relating to artificial intelligence and also to know the working principle behind various devices that use AI. The applications of AI is vast and a study of them creates a chance to explore the full potential and develop the field. The different types of AI , the differences governing its classification are important to be considered to know its development rate and the standard of current inventions. The object detection technique, natural language processing, face recognition, tracking , Google Map, Sophia-the human like Robot, Tesla autopilot, Amazon Alexa are just a few examples where AI is applied and are briefly discussed in this article. Each of them works on the basic principle of mimicking human behavior , but the huge set of algorithms for each of these differ greatly. All of them required years to model and come in to existence and practice as we see today**

Keywords: Artificial intelligence, potential research, Sophia,Tesla autopilot, Amazon Alexa, natural language processing, mimicking human behaviour

1. INTRODUCTION

Today, artificial intelligence is one of the fastest-growing emerging technologies and describes machines that can perform tasks that previously required human intelligence. The term artificial intelligence (AI) was coined by Stanford professor John McCarthy in 1956.Artificial intelligence that's used in the engineering sector uses software and hardware components. As machines become more sophisticated, they will be able to support not only smart production lines and complex manufacturing tasks, but will also be able to design and improve tasks over time—with little or no human intervention—through machine learning.

AI is currently increasing the popularity and gaining support due to its wide applications. Researches are continuously held to develop the immense potential and make maximum use of AI. AI is being tested and used in the health care industry.
AI is also gaining scope ion the case of machines with artificial intelligence include that play chess . Self driving cars are the future of our transportation system. Artificial intelligence also has applications in the financial industry, where it is used to detect in banking and finance such as unusual debit card usage and large deposit of cash in to accounts all of which help a bank's fraud department.

Even Google, a company that once said that mobile was its priority, has shifted its focus toward AI. Nearly every technology company is heavily investing in AI research and development, which demonstrates the importance that AI, holds for businesses in general . New technologies are introduced at a very fast rate and it is difficult to cope up with it..

2. SCOPE OF THE STUDY

Artificial Intelligence is currently increasing the popularity and gaining support due to its wide applications. Researches are continuously held to develop the immense potential and make maximum use of AI. AI is being used in the health care industry for testing drugs and different treatment in patients, and for surgical procedures . AI is also gaining scope ion the case of machines with artificial intelligence including the one that can categorize raw data into specific fields. Self driving cars are the future of our transportation system. Artificial intelligence also has applications in the financial industry, where it is used to detect in banking and finance such as unusual debit card usage and large account deposits—all of which help a bank's fraud department.

3. OBJECTIVE OF STUDY

1. To have a basic concept about artificial intelligence and machine learning.

ISBN -978-93-5406-047-02

2. To explore the role of Artificial Intelligence and Machine Learning in a few technologies and inventions recently launched.

4. RESEARCH METHODOLOGY

Meta Analysis is the method used for the study. It was done by collecting ample information from relevant sources that includes the websites of the developer of the discussed systems, the experiences of the users of the system, the researchers' scholarly article describing the system .

5. ARTIFICIAL INTELLIGENCE

Artificial Intelligence or AI refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

The prior characteristic of artificial intelligence is the capacity to rationalize and take actions that have the best chance of achieving specific goals.
Artificial intelligence is based on the principle that human intelligence can be defined in a way that a machine can easily copy it and execute task, from simple to those that are even more complex. The goals of artificial intelligence include studying, reasoning, and perception.

TYPES OF AI

A. Artificial Narrow Intelligence (ANI):

Narrow Artificial Intelligence is one that is programmed to perform a single task only like checking the weather, being able to play chess, or analyzing raw data to write journalistic reports, or any other task already programmed.

ANI systems can attend to a task in real-time, but they capture information from *a specific data-set*. As a result, these systems don't perform anything out of the single task that they are designed to perform.

Every sort of machine intelligence that surrounds us today is Narrow Artificial intelligence . Google Assistant, Siri, Google Translator, and other natural language processing tools are examples of Narrow AI. People may think that that these tools aren't "weak" because of their ability to interact with us and process human language, but the reason that we call it "weak" AI is because these machines are not at all close having human-like intelligence. They lack the self-awareness, feelings, and genuine intelligence to match human intelligence. That is they can't think for themselves .This explains why when we ask abstract questions about things like the meaning of life or how to approach a personal problem to Siri or Google Assistant, we get vague responses that don't make sense, or we get links to existing articles from the Internet that address the same questions. On the other hand, when we ask Siri what the weather outside is, we get a very accurate response. That's because answering basic questions about the whether outside is within the range of intelligence that Siri is designed to perform.

B.. Artificial general intelligence:

Artificial General intelligence or "Strong" Artificial intelligence refers to machines that exhibit human intelligence. In other words, AGI can successfully perform *almost all* intellectual task that a human being can. This is the sort of AI that we see in films in which humans interact with machines and operating systems that are conscious, sentient, and driven by emotion and self-awareness.

Now, machines are able to process data faster than we can. But as human beings, we have the ability to think reasonably, and get in to into our thoughts and memories to make informed decisions or come up with creative and new ideas. This type of intelligence makes us superior to machines, but it's little bit tough to define because it's primarily driven by our ability to be sentient creatures. Therefore, it's something that is very difficult to replicate in machines.

AGI is expected to be able to reason, solve problems, make decisions under uncertainty, plan, learn, integrate prior knowledge in decision-making, imaginative and creative.

But for machines to achieve *true* human-like intelligence, they need to be able of experiencing consciousness.

C. Artificial super intelligence:

Artificial Super Intelligence (ASI) will surpass human intelligence in all respect — from creativity, to general wisdom, to problem-solving and so on. Machines will be capable of exhibiting intelligence that we haven't seen in the most brilliant amongst us. This is the type of AI that humans are worried about, and the type of AI that people like Elon Musk predict will lead to the extinction of the human race.

6. MACHINE LEARNING

Machine learning is an subset of artificial intelligence (AI) that provides systems the ability to automatically study and show better performance from experience without getting programmed already. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves like we do learn from our past experiences.

There are mainly four forms of machine learning; supervised, unsupervised , semi-supervised and reinforcement learning. Each form of Machine Learning has differing approaches, but they all follow the same concept

ISBN -978-93-5406-047-02

of process and theory. This explanation covers the concept of general Machine Leaning.

TYPES OF MACHINE LEARNING ALGORITHM

A. Supervised machine learning algorithms

It can apply what has been learned in the past to new data using examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces a function to make forecasts about the result values without any human intervention. The learning algorithm can also compare its output with the correct, occured output and find errors to modify the design accordingly.

B. Unsupervised machine learning algorithms

They are used when the information used to learn is neither classified or labeled. Unsupervised learning studies how systems can infer a function to describe a concealed structure from unlabeled data. The system doesn't find out the right output, but it explores the data and can make inferences from datasets to describe concealed structures from unlabeled data.

C. Semi-supervised machine learning algorithms

This type of AI use both labeled and unlabeled data for learning– usually a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to relatively improve learning accuracy. Usually, semi-supervised learning is chosen when the given data requires talented and accurate resources in order to make the system familiar with it or learn from that.

D. Reinforcement machine learning algorithms

It is a learning method that interacts with its environment by actions and discovers errors and mistakes. Trial and error search and late reward are the two most relevant characteristics of reinforcement learning. This is a method that allows machines and software to automatically understand the best behavior in order to increase its performance. Simple reward feedback is required for the agent to learn which action is better; this is known as the reinforcement signal.

8. APPLICATIONS & INVENTIONS

A. OBJECT DETECTION USING AI

Object detection is incorporated with Computer Vision and a system that can identify the presence and accurate location of an object or body within an image. There can be singule or multiple occurrences of the same object to be detected.

The output of an object detection process is an image with bounding boxes around the objects in focus

- **Face Detection**: It is the term given to the task of implementing systems that has the ability to automatically recognize and identify location of human faces in images and videos. Face detection is present in applications incorporated with facial recognition, photography, and motion capture.

- **Pose Estimation**: It is the process of deducing the location of the main joints of a body from provided digital assets such as images, videos, or a sequence of images already provided. This application is present in systems that has the function of Action recognition, Human interactions, creation of assets for virtual reality and 3D graphics games, robotics and more

- **Object Recognition:** The process of identifying the classification a target object is associated with. Object recognition and detection are techniques with similar end outputs and implementation methods, Although the recognition process comes before the detection steps in some systems and algorithms.

- **Tracking:** It is the method of identifying, detecting, and following an object said within a sequence of images over some time. Applications of tracking are found in many systems like surveillance cameras and traffic monitoring devices.

B. AMAZON ALEXA

Alexa is introduced as a virtual digital assistant. It is developed by Amazon for its Amazon Echo and Echodot line of computing devices**.** Alexa responds to voice control by the user by returning information on products , music, news, weather, sports and many others. The back-end engine for Amazon's Alexa runs on Amazon Web Services in the cloud, which makes Alexa learn an person's or family's preferences and expand its functionality over time. In addition to returning information, Alexa also enables Echo devices to function as smart home hubs that can be

ISBN -978-93-5406-047-02

used to function internet connected devices like smart lights, thermostats and electronics.

Machine learning is the basic of Alexa's power, and it's only getting stronger as its popularity and the amount of data it gathers increasing day by day. Every time Alexa makes a mistake in interpreting the request of the user, that data is used to make the system smarter the next time it functions. Machine learning is the reason for the sudden development in the capabilities of voice-activated user interface.

As a subset of artificial intelligence , natural language generation (NLG) is the ability to get natural sounding written and verbal responses back based on data that is fed into a computer system. Human language is very complex, but today's natural language generation capabilities are becoming more easy . NLG can be understood as a writer that turns data into language that can be easily understood and communicated. Natural Language Processing (NLP) refers to model of communicating with an intelligent device using a natural language that people use such as English.

Processing of Natural Language is necessary when a person wants an intelligent system like robot to perform as per the instructions, when a person want to understand the decision from a dialogue based clinical expert system, that person will need this system .The field of NLP consists of making computers to perform important tasks with the natural languages humans use. The input and output of an NLP system are speech and text.

So, when one asks Alexa, any question, the appliance records the voice. Then that recording is sent through the Internet to Amazon's Alexa Voice Services which converts the recording into commands it understands. Then, the system sends the relevant output back to the appliance. When one asks about the weather for example, an audio file is sent back and Alexa tells you the weather forecast. That person may not have any idea there was any back and forth between systems. Alexa works as long as there is an internet connection .When people talk they interrupt themselves, change topics or repeat the same dialogue, use body language to emphasis and use a wide variety of words that have more than one meaning depending on the context.

C. GOOGLE MAP

Google map has become so common among us. It can give accurate data and traffic in any densely populated or remote areas. Machine Learning; a subset of AI uses statistical models and algorithms to perform tasks that are not previously programmed. To perform a tasks, Machine Learning Algorithms develop a mathematical model based on sample data. Google Maps provides accurate directions and real-time traffic information to millions of users all round the globe. This information is updated constantly, continuously to mirror the changes that has occurred. It is impossible to manually analyze more than 70 billion images to find latest, or updated, information for Google Maps. One of the goals of Machine Learning is to allow the automatic extraction of information from geo located imagery to improve Google Maps.

In case of accuracy, precision, and update frequency of maps, people always relay on it. Machine learning and large-scale image collection are the primary keys to providing location data in an easy way.

Google is using a very basic form of Artificial intelligence in its current Maps, given the fact that the algorithm automatically picks the best route for people to take. . This has been used by many industry analysts, who stated the fact that Google will massively use futuristic AI in the newly arriving versions of their flagship application.

Google Maps recently launched live bus delay prediction where foundation is machine learning in hundreds of major cities around the world. Google collects real-time data on bus locations from almost all transit agencies today. Real-time bus forecasts, which is available in India too, can help forecasts in places where transit agencies fail to provide real-time bus location data today.

Using a combination of time, distance travelled, and some other factors as data , the AI makes it possible for Google to provide forecasts without relying on bus schedules provided by public transportation agencies.

D. SOPHIA; THE HUMAN LIKE ROBOT

Hanson Robotics' most advanced human-like robot, Sophia, gives a glimpse of the future of AI. As a unique combination of science, engineering, and artistry, Sophia has attracted the attention of al the technologists

It's a robot that that is programmed and developed using artificial intelligence to see people, understand conversation, and form relationships. On the surface, Sophia is not at all similar to the AI-powered robots in movies. It can crack jokes, make facial expressions, and seemingly understand what's going on around it, which is

ISBN -978-93-5406-047-02

that unique feature developed by Hanson Robotics. It can learn from one experience and make the knowledge practical in new situations, as humans do. Cameras fixed inside Sophia's eyes collaborated with computer algorithms allow her to see. She can follow faces, sustain eye contact, and recognize individuals and remember them once known. She is able to process speech and have conversations with people using a natural language subsystem. Around January 2018 Sophia was upgraded with functional legs and the ability to walk, that was a turning point in Hanson Robotic's history.

According to a publications on Sophia's software, it was said that deep neural networks let the robot understand someone's emotions from their tone of voice and facial expression and react in kind. Sophia also can mirror people's postures,which is absolutely accurate and her code generates realistic facial movements of humans. Hanson has then patented the flexible rubber skin that covers Sophia's face which is the key for her human perception by people.
Goertzel says Sophia is more of a user-interface than a human being which means that it can be programmed to run different code according to the situations. He adds that Sophia can be pre-loaded with text that it'll speak, and then use machine learning to match facial expressions and go to the text. Sophia is also capable of running a dialogue system, where it can look at people, listen to what they say, and choose a pre-written response based on what the person said. Typically, Sophia's software can be classified into three configurations:

- A research platform for the team's AI research.
- A speech-reciting robot.
- A robotic chatbot. and other factors gathered from the internet like crypto currency price.

Di Sturco was the first photographer to step into Hong Kong-based Hanson Robotics. He said that"In the beginning, it was a bit difficult, She didn't recognize the camera ,but after three days, she kind of learned,I don't know if the engineer put something in the software, or if she went online and did some research, but she started to pose; It was actually really strange ---at one point, I realized I was even speaking with her," he adds"I had to step back and realize that she was a robot, not a human being."

E. IBM WATSON SUPERCOMPUTER

Watson is an IBM that joins together artificial intelligence and sophisticated analytical software for optimal performance as a "question answering" machine. The supercomputer is named after IBM's founder, Thomas J. Watson, the foundation of it.
The Watson supercomputer has the unique capacity to processes at a rate of 80 teraflops (trillion floating point operations per second). To replicate a high-functioning human's ability to answer questions, Watson accesses 90 servers with a joined data store of over 200 million pages of information, which it processes with the use of t six million logic rules.
Health care remains a primary and important focus for IBM as it tries to prove Watson technology, and the company continues to have partnerships with health care organizations. In May 2018, for example, India's largest speciality health care systems, Apollo, came forward to adopt Watson for Oncology and Watson for Genomics, which was a revolution ion the health care sector. The two IBM cognitive computing platforms will support doctors make decisions for personalized cancer care. IBM's use of Watson to find a solution to some of the biggest problems around patient care and also using data-driven insights to suggest treatment options would prove the value of Watson technologies very soon.
The first commercial implementation of Watson came in 2013 when the Memorial Sloan Kettering Cancer Center began using the system to recommend treatment options for lung cancer patients to ensure they received the right treatment while reducing price.
Watson Analytics is the important and primary implementations of Watson technology. It is a platform for exploring, visualizing and presenting data that utilizes Watson's cognitive capabilities to automatically surface data-driven insights and find a way that suits to presentation of the data.
IBM has published a range of application program interfaces (APIs) on its cloud that allow users to make their own AI applications that utilize Watson's core technology on the back end. There are APIs that support best development frameworks like Java, Python and others.

F. J.P. MORGAN CHASE'S CONTRACT INTELLIGENCE(COiN) PLATFORM

JP Morgan Chase came out as a pioneer in this regard, with this unique technology. In 2017, the company came out with a new software known as COiN. This software has been developed in order to automatically perform the document review process. This automation, once proven error proof, can produce accurate results, consuming only a

ISBN -978-93-5406-047-02

few seconds which is a very short time span. It's an alternative to the conventional documentation done by lawyers, which usually takes days to weeks time and may also contain errors. One of the important components of corporate law is to review documents. This process is long and complicated. It requires a lot of time and concentration to be put in. This made him come up with the idea to develop the software.

The primary technique that is used by the software is known as image recognition. By using image recognition, the software is able to compare and note the difference between different agreements. The company has also indicated that COiN is an unsupervised learning software, which means that there is a very little human involvement once it has been applied. The company has also stated that COiN can recognize over a hundred attributes of the contracts and then classify them into different groups.

So far, the company has stated that the software has proved to be even more precise than the lawyers who have been previously working on the documents. In addition to this, the software has also saved a lot of man hours. The bank's new software is already proving to be cost-effective, even though it has not been employed on a very large scale yet. The company has also indicated that it wants to expand the capability of the software, and cover other forms of contracts in addition to credit agreements. Further, it also focus to understand and classify new regulations that are formulated by the authorities.

According to Saxena, AI will help financial services companies expand banking worldwide, launch new products and strengthen customer engagements. AI has helped technology companies and other companies outside of traditional banking enter financial services, such as with mobile banking and digital money offerings. However, only firms that can earn customer trust, meet regulatory compliance requirements and enhance customer service will succeed in implementing it.

G. TESLA AUTOPILOT

It is an advanced driver assistance feature offered by the company called Tesla that has the ability to automatically change lanes, navigate autonomously on limited access freeways, and the ability to control the car to and from a garage or parking spot. In all of these features the driver is responsible to have constant and continuous supervision.

The data is used to generate highly data-dense maps showing all the important information from the average increase in traffic speed over a stretch of road, to the location of hazards which cause drivers to take action. Machine learning in the cloud takes care of the software system as a whole, while at an individual car level, edge computing decides what action the car needs to take right now. A third level of decision-making also exists,which helps the system.

Autopilot has the ability to maintain a safe distance from the vehicle in front of it as accelerating and braking according to the need as that vehicle speeds up and slows down. It also slows on narrow curves, on interstate ramps, and when a car enters or exits the road in front of it. It can be enabled at any velocity between 18 mph and 90mph. By default, it sets the limit at the current speed limit plus/minus a driver-specified offset, then adjusts its target velocity according to changes in speed limits. If road conditions demands, autosteer and cruise control disengage and an audio and visual demand that the driver has to take full control is displayed, which is a safety measure.

Autopilot includes a video display of what it sees around it, from a driver's point of view. It displays driving lanes and vehicles in front, behind the car and on either side of it (in other lanes). It also sees lane markings and speed limits, like a driver needs to note. It displays stop signs and traffic signals, it can also distinguish between pedestrians, bicyclists/motorcyclists, small cars, and larger trucks. Another notable feature is autopark which parks the car in perpendicular or parallel parking spaces. Autopilot can detect an accident as front or side collision with another vehicle, bicycle or pedestrian within a distance160 m, if any such event is found it sounds a warning. Autopilot has automatic emergency breaking that can understand objects that may hit the car and applies the brakes. The car may also automatically change the way to prevent a collision.

9. CONCLUSION

AI is one of the fastest growing sector in the field of research and development. The possibilities and advantages of it is abundant. Still, we need to consider that they all are machines and however hard we program them, they can't exhibit human like emotions. They may be programmed to show facial expressions and also to recognize them, but they lack feelings like sympathy, empathy, kindness e.t.c. hence human being is always superficial to human made robots. There should not be a situation where an invention of human is harmful to himself.

REFERENCES

1. Fact article on artificial intelligence https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

ISBN -978-93-5406-047-02

2. fact article on machine learning https://expertsystem.com/machine-learning-definition/
3. Fact article on artificial intelligence https://en.wikipedia.org/wiki/Artificial_intelligence
4. Fact article on" What is artificial intelligence? How does it work?" https://builtin.com/artificial-intelligence
5. Fact article on Benefits and Risk of artificial intelligence https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/
6. fact article on "Machine learning, an introduction, towardsDatascience" https://towardsdatascience.com/introduction-to-machine-learning-f41aabc55264
7. Fact article on " What is machine learning?" -MIT Technologyreview https://www.technologyreview.com/s/612437/what-is-machine-learning-we-drew-you-another-flowchart/
8. Fact article on "Amazon Alexa"https://en.wikipedia.org/wiki/Amazon_Alexa
9. Fact article on " Appliance science- How does Alexa work" https://www.cnet.com/news/appliance-science-alexa-how-does-alexa-work-the-science-of-amazons-echo/
10. Fact article on " Machine Learning In Practice: How Does Amazon'sAlexaWork?" https://www.forbes.com/sites/bernardmarr/2018/10/05/how-does-amazons-alexa-really-work/#732f726b1937
11. Fact article on " Tesla Autopilot" https://en.wikipedia.org/wiki/Tesla_Autopilot
12. Fact article on " Artificial Intelligence- Face Recognition" https://www.everteam.com/en/artificial-intelligence-face-recognition/
13. Fact article on " J. P. Morgan Chase's Contract Intelligence" https://mint2save.com/jp-morgan-chases-contract-intelligence-coin/
14. Fact article on "Google map and AI; What's the future of google's flagship" https://www.marktechpost.com/2019/04/02/google-maps-and-ai-whats-the-future-of-googles-flagship-application/
15. Information on Sophia, the Robot, designed by Hanson Robotics https://www.hansonrobotics.com/
16. Video clip on " Artificial Intelligence in 10 minutes" https://youtu.be/oV74Najm6Nc

ISBN -978-93-5406-047-02

Bigdata Applications And Challenges in Health Care

Anas A[#],Binju Saju[*]

[#]*computer science department*

[1]anas1405alikhan@gmail.com

[2]binju@naipunnya.ac.in

[#]*PG Scholar*

Naipunnya Institute of Management and Information Technology, Koratty, Thrissur, Kerala, India

*Assistant Professor

Naipunnya Institute of Management and Information Technology, Koratty, Thrissur, Kerala, India

***Abstract*— Big data is the catchword today. It is heard all over, especially in the healthcare industry. The wide variety of big data and the pace at which it is managed makes it overpowering. With the help of big data, the vast amount of data can be stored scientifically. Now doctors and other healthcare consultants can make informed decisions as they have access to a wide range of data. Of course, the data generated will developed by leaps and bounds, and newer systems will be able to process it quickly and cost effectively. This paper makes a study on big data application and challenges on healthcare**

***Keywords*— Big data, Health care, applications of big data, Challenges of big data**

I. INTRODUCTION

Bigdata in Healthcare refers to the large health data collected from various sources including electronic health records, medical imaging, major records, medicinal research etc...

Healthcare Industry is one of the world's biggest and widest developing industries. During, the recent years the healthcare management around the world is changing from disease-centered to a patient-centered model [1] and volume based to a value-based healthcare delivery model [2]. Educating the superiority of health care and decreasing the cost is a principle behind the developing movement toward value based healthcare delivery model and patient-centered care. The capacity and demand for big data in healthcare organizations are growing gradually [3]. To provide active patient-centered care, it is crucial to manage and analyze vast health data. The outdated data management implements are not sufficient enough to analyze big data as variety and volume of data sources have increased in the past two decades. There is a need for new advanced big data tools and technologies that can meet and increasing the ability of managing healthcare data [4].

Nowadays, there is an increasing demand for more information by the patients about their healthcare options or choices, and want participation in their health decision-making [5]. The big data will help to provide patients with up-to-date information to assist them to make the best decision and to comply with the medical treatment.

The big data are used to predict the diseases before they arise based on the medical records. Many countries' public health systems are now providing electronic patient records with advanced medical imaging media [6]. The practice of big data takes the prospective to encounter the upcoming market needs and trends in healthcare establishments [7]. Big data provides a great opportunity for epidemiologists, physicians, and health policy experts to make data-driven judgments that will eventually develops the patient care [8].

In spite of the integral complexities of healthcare data, there is potential and benefit in developing and implementing big data solutions within this kingdom. A report by McKinsey Global Institute suggests that if US healthcare were to use big data creatively and effectively, the sector could create more than $300 billion in value every year. Two-thirds of the value would be in the form of reducing US healthcare expenditure [9].

In this paper we discussed about Various applications of big data in health care like Product Development, Telemedicine, Real-Time Alerting, Personal health record (PHR)Predictive Analytics in Healthcare, Clinical trials, Prevention of Unnecessary ER visits etc…. Also big data have lot of challenges like data management, miscommunication gaps, data modeling, data accessibility etc…

This paper is organized as follows; Section 2 describes the re**l**ated works, Section 3gives an overview of different applications of big data in healthcare, Section 4 contain challenges of big **d**ata in healthcare. Finally we conclude in Section 5

ISBN -978-93-5406-047-02

II. RELATED WORK

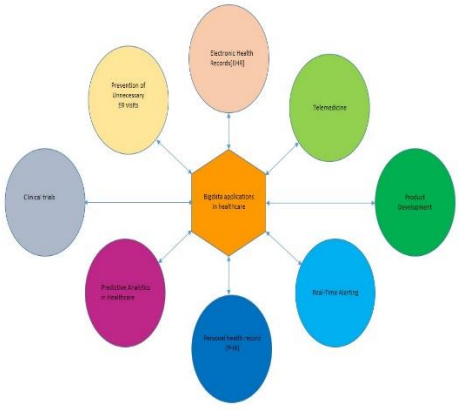
The Big Data revolution has begun for many industries. The healthcare industry has been playing catch up and has finally reached a consensus on the value of Big Data as a transformative tool. According to McKinsey & Company study entitled The Big Data Revolution in Healthcare, big data revolution is under way in health. The combined power of information from real-time devices, people, clinical systems, and historical population data makes Big Data a very helpful tool in improving the healthcare system. [10]

As the healthcare industry witnesses large volumes of data, the first step will involve governance and linking accurate and actionable data in real time. In this age of connectivity, integrating health systems with large amounts of clinical, financial, genomic, social and environmental data will be crucial for real-time analytics and patient care. In addition, the new wave of digitizing medical records has seen a paradigm shift in the healthcare industry. [11]

The quality for Big Data in human services today is to a great extent restricted to examine in light of the fact that utilizing huge information requires an extremely specific aptitude set. [12][13]

Big Data refers to large, complex datasets that are beyond the capabilities of traditional data management systems to store, manage, and process in a timely and economical manner. Often in petabytes, structured, semi structured, and unstructured, Big Data creates challenges in data capture, transfer, encryption, storage, analysis, and visualization [14]. Big Data is a disruptive phenomenon, still in the early stages of adoption for many sectors, but it is very clear that harnessing its capabilities can provide compelling benefits.

III. APPLICATIONS OF BIG DATA IN HEALTHCARE



Fig(1):applications of big data

i. *Electronic Health Records(EHR)*

An EHR (Electronic Health Record) is a digital version of a patient's paper chart. EHR hold the medical and diagnoses, medications, treatment plans, allergies, laboratory and test results etc. Its allow accessto evidence-based tools that workers can use to make decisions about a patient's care. Health information can be created and managed by authorized providers in a digital format capable of being shared with other providers across more than one health care organization by using EHR such as laboratories, specialists, medical imaging facilities, pharmacies, emergency facilities, etc..., so they contain information from all clinicians involved in a patient's healthcare.treatment histories of patients. EHR hold a patient's medical history,

Fig(2): Pictorical representation of HER

The advantages of EHR over normal record system

- Providing exact, latest, and complete information about patients at the point of care
- Permitting quick access to patient records for more synchronized, efficient care
- Fast sharing electronic information with patients and other clinicians
- Helping providers more effectively diagnose patients, moderate medical errors, and provide harmless care
- Refining patient and provider interaction and communication, as well as health care convenience
- Permitting safer, more reliable prescribing
- Helping promote legible, complete documentation and exact, updated coding and billing
- Improving privacy and security of patient data
- Helping providers progress productivity and work-life balance
- Enabling providers to expand efficiency and meet their business goals
- Reducing costs through reduced paperwork, improved safety, cheap duplication of testing, and improved health.

ISBN -978-93-5406-047-02

ii. Product Development

Discovering and developing new medicines and other health-related products takes an incredible amount of time and money. Big data help us to reduce the time involved in amount of different ways. This, of course, serves to moderate the costs involved. For example, during the usual research and development phase of product development, it is not always clear how to use total data exposed. With big data methods, R&D teams are capable to find valuable data much faster and more efficiently, therefore decreasing the time needed to develop the product and get it to market. Product development involves extensive trial and error experimenting, which requires ample time. Big data removes the guesswork permitting research and development companies to get results more quickly and thus develop more accurate products faster. Real-time data analytics help healthcare groups upgrade their products based on huge data sets.

iii. *Telemedicine*

The term telemedicine refers to delivery of remote clinical services using technology. It is used for primary consultations and initial diagnosis, remote patient monitoring, and medical education for health professionals. Some more specific uses include telesurgery – doctors can perform operations with the use of robots and high-speed real-time data delivery without physically being in the same location with a patient. Clinicians use telemedicine to provide personalized treatment plans and prevent hospitalization or re-admission. It allows clinicians to predict acute medical events in advance and prevent decline of patient's conditions. By keeping patients away from hospitals, telemedicine helps to reduce costs and improve the quality of service. Patients can avoid waiting lines and doctors don't waste time for unnecessary consultations and paperwork. Telemedicine also improves the availability of care as patients' state can be monitored and consulted anywhere and anytime.

Fig(3):Pictorical representation of telemedicine

iv. Real-Time Alerting

In hospitals, Clinical Decision Support (CDS) software analyzes medical data on the spot, providing health practitioners with advice as they make prescriptive decisions. However, doctors want patients to stay away from hospitals to avoid costly in-house treatments. Analytics, already trending as one of the business intelligence buzzwords in 2019, has the potential to become part of a new strategy. Wearable's will collect patients' health data continuously and send this data to the cloud. Additionally, this information will be accessed to the database on the state of health of the general public, which will allow doctors to compare this data in socioeconomic context and modify the delivery strategies accordingly. Institutions and care managers will use sophisticated tools to monitor this massive data stream and react every time the results will be disturbing.

For example, if a patient's blood pressure increases alarmingly, the system will send an alert in real time to the doctor who will then take action to reach the patient and administer measures to lower the pressure.Another example is that of Asthma polis, which has started to use inhalers with GPS-enabled trackers in order to identify asthma trends both on an individual level and looking at larger populations. This data is being used in conjunction with data from the CDC in order to develop better treatment plans for asthmatics.

v. Personal health record (PHR)

As its name suggests, it is the health-related data and information of patients and about public's ultimate health information. It is available for additional use. This stands in difference to the commonly used electronic medical record, which is operated by organizations (such as hospitals) and contains data entered by clinicians (such as billing data) to support insurance claims. The goal of a PHR is to deliver a complete and exact summary of an individual's medical history which is available online. The health data on a PHR might include patient-reported consequence data, lab results, and data from devices such as wireless electronic weighing scales or (collected passively) from a smartphone. PHRs award patients access to a extensive range of health information sources, best medical practices, and health knowledge. All of a person's medical records are stored in one place as a replacement for paper-based files in various doctors' offices. Upon encountering a medical condition, a patient can well access test results, communicate with their doctors, and share information with others suffering in the same way [15][16].

vi. Predictive Analytics in Healthcare

Predictive analysis point to patient's safety and quality care. It preserves doctors about the patient's updated medical histories

ISBN -978-93-5406-047-02

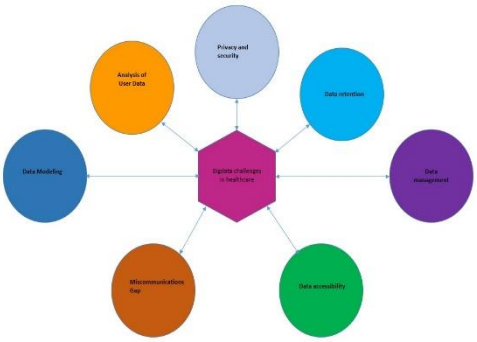
and helps predict results for future. For example, the analytics tools would be capable to predict which patient is at risk of what disease, so to make decisions as a result to improve patient's health. Predictive algorithms using different programming languages can be produced to predict the health of a patient over time. Predictive analytics is the practice of mining information from existing data sets in order to decide patterns and predict future outcomes and trends. Predictive analytics does not tell you what will happen in the upcoming. Instead, it forecasts what might happen in the future with a suitable level of reliability, and includes what-if states and risk assessment.

vii. Clinical trials

Clinical trials are research studies done in people that are aimed at assessing a medical, surgical, or behavioral intervention. They are the key way that researchers find out if a new treatment, like a new medicine or diet or medical device (for example, a pacemaker) is safe and in effect on people. Often a clinical trial is used to study if a new treatment is more effective and/or has less harmful side effects than the normal treatment. Other scientific trials test ways to find a disease initial, sometimes before there are symptoms. So others test ways to prevent a health problem. A clinical test may also look at how to make life better for people living with a life-threatening disease or a long-lasting health problem. Clinical trials sometimes study the role of caregivers or support groups.

viii. *Prevention of Unnecessary ER visits*

Hospitals want to condense the number of ER visits or Emergency visits of patients. They consider that it increases healthcare costs and sometimes does not lead to improved results for patients. For example, a man suffering with severe abdominal pain comes to an emergency room. The doctor will try to find out the cause of the problem such as kidney stone or appendicitis or something else. Now if he has a technique of knowing the patient's previous medical results, he could begin the treatment as soon as possible. The inspection would take less time and would also reduce the amount of money.For this, Almeda county hospitals in California, USA planned to create a program which called PreManage ED. According to this program, the records of the patients are shared with the emergency sectors such as, if the patient has already done some tests at other hospitals or earlier what advice were given to the patient. This decreases the time of patient to get the details of previous tests to them and do needless formalities. This is indeed a great application of big data analytics in healthcare area which saves both time and money.



Fig(4):Graphical representation of challenges

IV. *CHALLENGES OF BIG DATA IN HEALTH CARE*

i. *Privacy and security*

Privacy and security have particular importance for health care businesses. Successful attacks on health care data can be extremely profitable for criminals and tremendously damaging for organizations. And the financial costs of data breaches may be just the beginning – reputational charges are harder to measure but may remain for long periods. And individuals whose data is stolen may hurt most of all, since health records contain personal data ranging from credit card numbers to details about diagnoses and lab tests, raising threats of identity theft and even blackmail.

ii. Data retention

Health data must halt accessible for at least five years. That means businesses need to take a long-term attitude to data stewardship and keep track of when the data gets retrieved, by whom, and for what purpose. Medical data management software allows users to launch access rights and processes, such as those that give temporary data viewing abilities to representatives in different departments in a hospital. These products can index data and notes and trace the path when data entered the system. Organizations must put processes in place to occasionally sort through the data to delete it when appropriate, or modify and analyze it to use it in new ways, such as to gauge trends across several years.

iii. *Data management*

Health organizations face big-data-related challenges that can impact patient safety. All data that health organizations collect needs to be described, formatted, deduced and checked for accuracy, and made accessible for various uses like medical, billing, administrative etc. The volume and velocity of big data makes this task more difficult.Some hospitals now employ

ISBN -978-93-5406-047-02

patient safety experts, but in addition to having medical expertise, people in these roles must understand how data management practices can improve or hinder patient safety.

iv. Data accessibility

All data management strategies fall short if they don't result in content that's accessible and in the correct format for reporting. Data analysts must be empowered to access the data they need and share what the data exposes.

v. Analysis of User Data

The major focus of analysis of user data is in determining users determined. This is certainly the focus of a lot of the predictive analytics used in online advertising, and the reason that search advertising is far more effective than display advertising.

vi. Data Modeling

Although big data is excellent for modelling and simulation, there is a need to identify, structure and pool the proper relevant data so that it can be used to model the problems, which later can be used for intervention. Without the proper structured data, it is challenging to analyze and visualize the output and to extract specific information or data.

vii. Miscommunications Gap

The miscommunications or the gap between the users and data scientists is one of the biggest problems in relations to big data. The understanding of the users on data generated by data scientists' maybe low and this may affect the effective usage of big data. The health data from all clinics and hospitals need to be pooled together as stored at one-stop center (big data). At the moment, all the information are kept separately. As such, it is difficult to get a clearer picture of the patients due to the incomplete information gathered. Thus, this waste a lot of time as the doctor will need to start all over from the beginning taking the patients history. Since big data has the ability to predict future medical issues which is a positive thing, big data can also pose risk and undermine doctors. The patients too will rely on the technology rather consulting the healthcare practitioners

V. *CONCLUSION*

Big Data refers to large, complex datasets that are beyond the capabilities of traditional data management systems to store, manage, and process in a timely and economical manner. The EHR is about quality, safety, and efficiency. It is a great tool for physicians, but cannot ensure these virtues in isolation. Achieving the true benefits of EHR systems requires the transformation of practices, based on quality improvement methodologies, system and team based care, and evidence-based medicine. Not only the EHR big data has various applications in health care like Product Development**,** Telemedicine, Real-Time Alerting, Personal health record (PHR)Predictive Analytics in Healthcare**,** Clinical trials**,** Prevention of Unnecessary ER visits etc…. Also big data have lot of challenges like data management, miscommunication gaps, data modeling, data accessibility etc…

REFERENCES

[1] J. W. Cortada, D. Gordon, B. Lenihan, The value of analytics in healthcare: From insights to outcomes, IBM Global Business Services, Executive Report, 2012.

[2] T. Huang, L. Lan, X. Fang, P. An, J. Min, F. Wang, Promises and Challenges of Big Data Computing in Health Sciences, Big Data Res. 2 (2015) 2–11. doi:10.1016/j. bdr.2015.02.002.

[3] M. W. Stanton, Expanding patient-centered care to empower patients and assist providers, Research in Action. 5 (2002) 1-12.

[4] K. Feldman, D. Davis, N. V. Chawla, Scaling and contextualizing personalized healthcare: A case study of disease prediction algorithm integration, J. Biomed. Inform. (2015) 1–9.

[5] https://www.cognizant.com/industries-resources/healthcare/BigData is-the-Future-of- Healthcare.pdf

[6] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, K. Taha, Efficient Machine Learning for Big Data: A Review, Big Data Res. 2 (2015) 87–93. doi:10.1016/j. bdr.2015.04.001.

[7] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Heal. Inf. Sci. Syst. 2 (2014) 1–10. doi:10.1186/2047-2501-2-3.

[8] D. I. Sessler, Big Data and its contributions to peri-operative medicine, Anaesthesia. 69 (2014) 100–105. [9] S. V. Nuti, B. Wayda, I. Ranasinghe, S. Wang,

[9]J. Manyika, M. Chui, B. Brown et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.

[10] 2013 IEEE International Conference on Big Data- A Look at Challenges and Opportunities of Big Data Analytics in Healthcare, Raghunath Nambiar Cisco Systems, Inc. San Jose, CA 95134, USA

[11] Big data security and privacy issues in healthcare Nanthealth, Harsh Kupwade Patil and Ravi Seshadri, 2014 IEEE International Congress on Big Data

[12] BIG DATA APPROACH IN HEALTHCARE USED FOR INTELLIGENT DESIGN - Software As A Service, Weider D. Yu, Jaspal Singh Gill, Maulin Dalal, Piyush Jha, Sajan Shah, 2016 IEEE International Conference on Big Data (Big Data)

[13] "How Big Data Impacts Healthcare", Harvard Business Review Journal,Aug2014https://hbr.org/resources/pdfs/comm/sap/18826_HBR_SAP_Healthcare_ Aug_2014.pdf

[14] http://attunelive.com/big-data-applications-healthcare/

[15] *Archer, N.; Fevrier-Thomas, U.; Lokker, C.; et al. (2011). "Personal health records: A scoping review". Journal of the American Medical Informatics Association.* ***18*** *(4): 515–22. doi:10.1136/amiajnl-2011-000105. PMC 3128401. PMID 21672914.*

[16] Assadi, V.; Hassanein, K. (2017). "Consumer Adoption of Personal Health Record Systems: A Self-Determination Theory Perspective". *Journal of Medical Internet Research*. **19** (7): e270. doi:10.2196/jmir.7721. PMC 5553007. PMID 28751301

ISBN -978-93-5406-047-02

Passphrase Based Authentication to Prevent Shoulder Surfing Attacks

Fathima Beevi v s[#], Dr.Sarika S[*]

[#]*computer science department*

[1]fathimabeevivs1998@gmail.com

[2]sarika@naipunnya.ac.in

[#]*PG Scholar*

Naipunnya Institute of Management and Information Technology, Koratty, Thrissur, Kerala, India

*Assistant Professor

Naipunnya Institute of Management and Information Technology, Koratty, Thrissur, Kerala, India

***Abstract*— The estimated number of mobile devices is around 5.8 billion, which is believed to have grown exponentially in every year. This makes us fully dependent on mobile devices with our sensitive data being transported all over. As a result mobile security is one of the most important concepts to take in consideration. Screen lock is the first step for smart phone protection. But smart phone locks are not eliminating the chance of unauthorized access from attacks like shoulder surfing. Shoulder surfing is direct observation techniques, such as looking over some ones shoulder to get information. Shoulder surfing is the effective way to get information in crowded places because it's relatively easy to stand next to someone and watch as they fill out a form, enter a pin number at an ATM machine or use a calling card at a public pay phone. Shoulder surfing can also be done long distance with the aid of binocular or other vision enhancing devices. This paper makes a comparative study of TicToc pin entry method, flywheel pin entry method and color pass method used to resist shoulder surfing attacks.**

***Keywords*— Shoulder Surfing attack, Smart phone security, graphical password, Color Pin entry,passphrase authentication**

VI. INTRODUCTION

Security has gained a lot of importance recently as attacks are increasing and security techniques become incompetent to resist them. A number of techniques has been proposed, to detect recent types of social engineering attacks[5] like the methods discussed in [1][2][3][4]. Truly speaking, there is no single silver-bullet solution existing to resist all the attacks effectively, so multiple techniques are required to mitigate specific attacks. Whenever a kind of attack is identified and taken care of, a newer, more intricate version surfaces. That's why, attack detection is a complicated and continuous process which should be handled technically using appropriate security measures.

Now a days more users and businesses uses smartphones to communicate. These technologies make profound changes in the hierarchy of information systems and therefore they have become the source of new risks. Moreover, smartphones collect and compile an increasing amount of sensitive information to which access must be controlled to protect the privacy of the user.

There are 4.39 billion internet users in 2019, an increase of 366 million (9 percent) versus year 2018. There are 3.48 billion social media users in 2019, with the worldwide total growing by 288 million (9 percent) since this time last year.3.**26 billion** people use social media on mobile devices in June 2019. For login to the social Medias and for mobile banking we need passwords and pin numbers. But these passwords and pin entry is disposed to shoulder surfing attacks.

Shoulder Surfing is a type of social engineering techniques refers to a direct observation to others for obtain confidential information like passwords, pin numbers etc. The entry of a Pin can easily be observed in crowded place by standing next to user or with the help of mirrors or concealed miniature cameras.

In this paper, we discuss about various methods to prevent shoulder surfing attack such as color pass method, TicToc pin entry method and flywheel pin entry method and a comparative study is performed in terms of some parameters like security, usability, time taken for login, complexity, user friendliness and error during login.

The paper is organized as follows: Section 2 describes The Related works, section 3 gives an overview of the different methods to resist shoulder surfing attacks and section 4 composed of the comparative analysis .Finally, we conclude in section 5.

VII. RELATED WORK

In this section, we review the related works and discuss several issues closely related to our work. Shoulder-surfing attacks are directed at common people and make their prevention through various techniques quite infeasible. A number of studies are conducted to discuss about security like MatsumotoandImai [6] and the work by Wangetal [7], but they are within the limitations of humans. The main focus was to incorporate an indirect method for secret transfer. It means separating the visible key entry procedure from the secret itself.

ISBN -978-93-5406-047-02

Research works are conducted based on textual passwords [8], [9] graphical passwords [10], [11] and PINs [12], [13], [14], [15]. Some another works focused on designing leakage resilient password entry on touchscreen mobile devices [16], [17]. The existence of these diverse schemes testifies as to how challenging it is to design an authentication scheme that is both secure and usable [6]. The systems with more security are likely to result in highly complex, error-prone, and tedious user procedures, while putting more stress on usability can lead to insecure systems [18] [19] .

In 2002, Sobrado and Birget [20] proposed three shoulder surfing resistant graphical password schemes, the Movable Frame scheme, the Intersection scheme, and the Triangle scheme. But, the Movable Frame scheme and the Intersection scheme fails to achieve Authentication. In the Triangle scheme, the user has been asked to select and memorize several pass icons as his password. In order for login into the system, the user should be able to succeed the predetermined number of challenges and in every challenge, the user is supposed to find three pass-icons from a set of randomly chosen icons displayed on the login screen, and click inside the invisible triangle created by those three passicons. In 2009, Gaoetal [21] proposed a graphical password scheme which uses color login and provide resistant to the shoulder surfing attack. In this scheme, the login time can be reduced by using background color. In this scheme, the probability for accidental login is more and the password space provided is too less.

In 2012, Rao et al [22] proposed a text based shoulder surfing resistant graphical password scheme, PPC. For login, the user has to mix his textual password to produce several pass-pairs, followed by four predefined rules to get his session password on the login screen. But, this technique is more complex and non-user friendly.

G.T.Wilfong [23] proposed a technique which insists the user to perform a simple mathematical operation. The user should remember a four digit PIN number and a value will be received to his protected media. The received value is combined with the PIN number and a modulo 10 operation is performed. The user enters the obtained digits using a public keyboard. The method is secure and easy for math oriented people but difficult for the non-math oriented people to work with.

In 2014 N. Chakraborty and S. Mondal [24] proposed a system to prevent shoulder surfing attacks using colors named color pass method. In 2015 Taekyoung Kwon, and Jin Hong, [25] present a pin entry method to avoid shoulder surfing attack named tic toc pin entry method which uses to sounds tic for right input and toc for wrong input.in 2018 Ms. Ojaswi K. Kasat, and Dr. Umesh S. Bhadade,[26] proposed a Revolving Flywheel PIN Entry Method to Prevent Shoulder Surfing Attacks. In this paper, we are trying to make an analysis of the recent methods color pass, tictoc pin entry, and flywheel pin entry method.

VIII. METHODS TO PREVENT SHOULDER SURFFING ATTACKS

I. *Color Pass Method*

Color Pass Method [24] is a color based pin entry method. In this method user can choose four color from 10 different colors and the colors are represented by {C_0, C1……..C_9}.User can choose one color more than one time. For example user can choose {Red, Yellow, Red, and White}. Fig 1 shows that the user interface on the screen and Fig 2 shows the colors for implementing feature table.

User will receive a secret pin number while login process. The pin number varies from one login to another login. After listening to each pin number, user selects a Feature Table according to the pin number. Corresponding to the chosen color, he locates the color cell in the feature table. The user then finds the digit corresponds to the color cell and enters the digit as response to the challenge. Similarly user will respond to the other three pin number and will complete the login process. Valid response to the pin number will validate the user. Pin number is generated according to each color he selected.

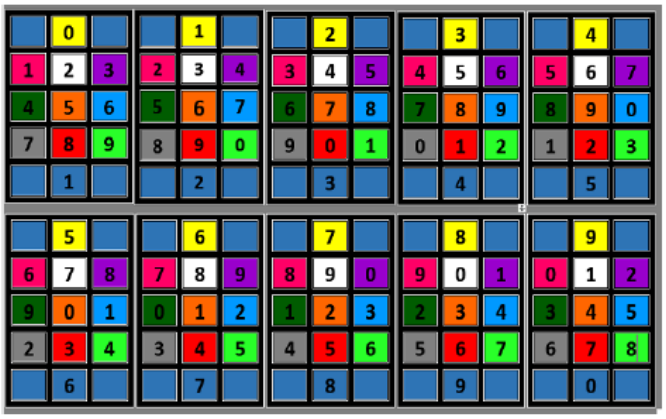


Fig 1. user interface on screen

0	Yellow
1	Pink
2	White
3	Violet
4	Dark Green
5	Orange
6	Sky Blue
7	Grey
8	Red
9	Light Green

Fig 2. Used colors for implementing feature tables

For example user selected colour is {red,white,red,yellow} and the secret pin genereted by the system is 2641.user should select feature table2 and selecte red colour that is number 9. After this user should select feature table 6 and select the white color that is number 7. After this user should select feature table 4 and select the red color that is number 1. After this user should select feature table1 and select the yellow color that is number 0. So the user's response is 9710 corresponding to the color red, white, red, yellow and the pin generated by the system 2641

ALGORITHM

Step 1 : Start

Step 2 : User can select 4 colors to register into the system

ISBN -978-93-5406-047-02

Step 3 : After login
Step 4 : 4 digit secret pin is generated
Step 5 : Display the feature tables on screen
Step 6 : Select the feature table according to the secret pin
Step 7 : Select the color chosen from the selected feature table
Step 8 : Repeat steps 6 and 7 for all the four colors
Step 9 : If no errors, login is successful
Step 10: If error occurred, login is not successful, go to step 4
Step 11: Stop

II. *TicToc Pin entry method*

TicToc Pin entry method [25] is a colored pin entry system proposed by Taekyoung kwon. TicToc PIN entry method uses four colors black, white, red, and blue. In order to enter one digit pin, it takes two rounds. Similarly, it needs eight rounds for entering four digits. First round carries a color pad with all the four colors under the numeric keypad, in this round user can select the appropriate color from the color pad. Second round completion phase carries color pad which holds three color under the numeric keypad, this round finishes with user pressing suitable color from the color pad. This method uses a vibrotactile channel for identifying the actual vibration (Tic) and a fake vibration (Toc).

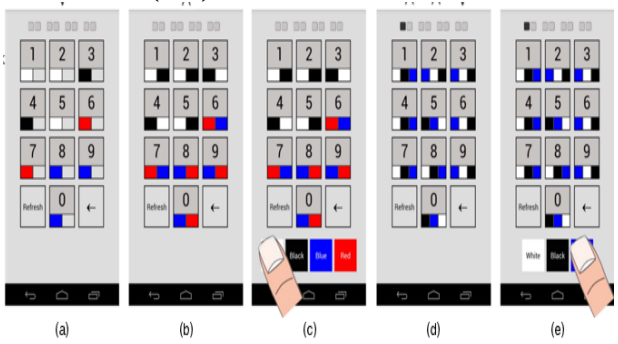


Fig 3. User interface of TicToc method

The first round contain a digit space Q = {1,2,3,4,5,6,7,8,9,0} is divided into L1 = {1,2,5}, R1 ={6,7}, L2 = {3,4}, and R2 = {8,9,0} .. Let C1 = black color C2 = white color C3 =blue color and C4 = red color. This round contain two challenging phases. first challenging phase is, left boxes of L1, L2, R1, R2 are filled with colors C1, C2, C3, C4, respectively. After 500 millisecond delay second phase begins with the right boxes of L1, L2, R1, R2 are filled with colors C2, C1, C4, C3, respectively. After 500 millisecond delay, a keypad consisting of the four colors, in random order, is displayed to receive user input. One of the two challenge phases, randomly chosen at the time of execution, is accompanied by a short 30 millisecond vibrotactile signal. The display of the accumulated challenges is maintained until the user supplies a color input.

The second round basically consists of three 500 millisecond challenge phases, with one of them, chosen at random, accompanied by a 30 millisecond vibrotactile signal. The second round contain a digit space Q = {1,2,3,4,5,6,7,8,9,0} is divided into Q1={1,4,7,8} Q2= { 2,3,6,9} and Q3={5,0}.let C5=black color, C6= white color, andC7=blue color. Each key of the numeric keypad contains three small boxes colored from left to right. The first phase of the second round assigns colors C5, C6, and C7 to the left boxes of Q1, Q2, Q3, respectively. The second phase further fills the center boxes of Q1, Q2, and Q3 with the colors C6, C7, and C5. The third phase adds colors C7, C5, C6 to the right boxes. A keypad consisting of the three colors, in random order, appears after the three 500 millisecond challenge phases.

III. *Fly Wheel Pin entry method*

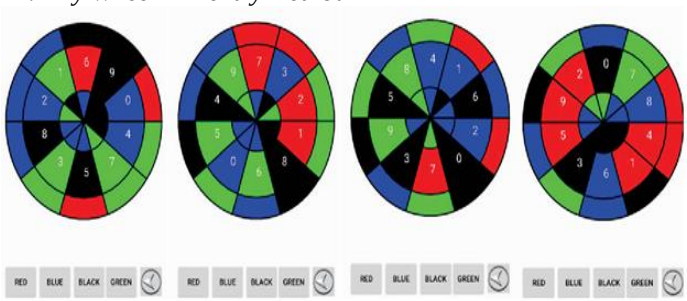


Fig 4. User interface of fly wheel pin entry method

Fly wheel pin entry method [26] is proposed by Ojaswi and Umesh s. In this method user want to select 2 colors for entering a single digit number. This method contain a fly wheel which contain three layers and 10 sectors and 30 sections outer layer, middle layer and inner layer. The outer layer and the inner layer contain two colors and the middle layer contain numbers from 0 to 9. In registration phase user want to select a secret pin number. For entering a digit user want to notice sectors at which sector his pin number belongs. User want to select outer color+ inner color button from the button given below the fly wheel.

For example user selected pin is 5846 .As user will open the screen for unlock the terminal fig 4 (a) screen will appear and to inputting first digit user has to look in which sector 5 is present and has to press upper & inner section color button to enter 5 here it is Red+ Blue. After this wheel will rotate and get refreshed. Fig 4 (b) will appear now for entering second digit of password which is 8 user has to press Black+ Black, again wheel will get refresh and rotate Fig. 4 (c) Will appear for entering 4 user has to press Green+ Blue, again it will refresh and rotate .Fig. 4 (d) will appear for entering 6 user has to press Blue + Black. So for4 digit PIN (5846) 8 clicks will be there (Red + Blue + Black +Black+ Green+ Blue+ Blue + Black). Maximum time limit for authentication is 20 sec, after 20 sec session will get refreshed and user has to start from new.

ALGORITHM

Step 1 : start
Step 2 : user register into the system by selecting 4 digit pin number
Step 3 : display the flywheel for login process
Step 4 : select the pin number by selecting outer color +inner color, by selecting the color button given below the Flywheel
Step 5 : continue step 4 for all the 4 digits
Step 6 : after selecting 4 digits validate the pin number and color

ISBN -978-93-5406-047-02

combination selected
Step 7 : if pin number and colors are matched, login successful
Step 8 : if pin number and colors are not matched, login not successful .go to step 4
Step 9 : stop

IV. *COMPARATIVE ANALYSIS*

In this section we have made an analysis of color pass method, TicToc method and fly wheel method in terms of the parameters security, usability and time taken for login.

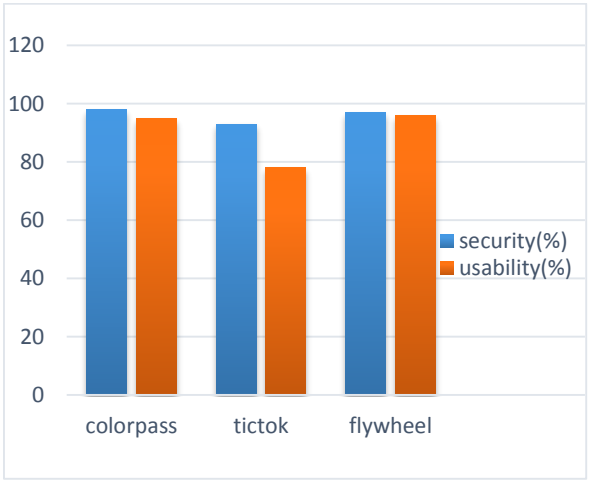


Fig 5. Comparing security and usability

Fig 5 shows the comparative analysis of three methods in terms of and usability. It has been seen that security is more for color pass method than that of TicToc and flywheel method but fly wheel method has more usability than other two.

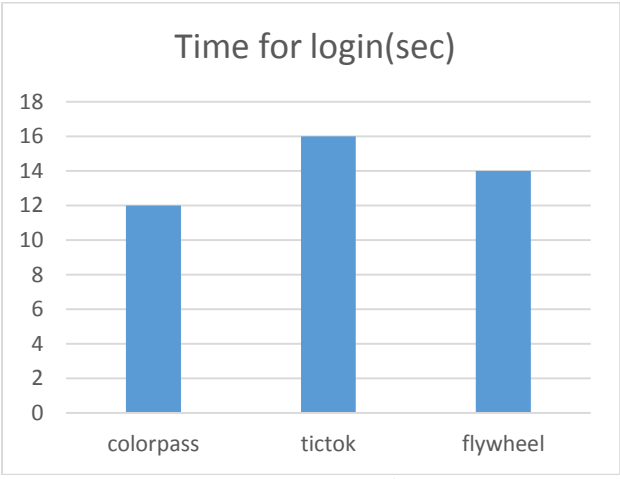


Fig 6. Comparison on Time for login (sec)

Fig 6 shows the comparative analysis of the three methods with respect to the time for login in seconds. A quick glance at the results show that color pass method uses less time for login than TicToc and fly wheel method as it is less complex, and secure than the other methods. Moreover, Color Pass method claims that it has only 2% error during login. But usability is more flywheel method.

V. *PROPOSED SYSTEM*

The Proposed system proceeds in 2 main phases. Registration phase and login phase.

i. Registration phase

In registration phase, the user is asked to enter username and password along with 5 security questions from various categories and their answers. According to the username and password, a passphrase is generated which is a combination of both. For the passphrase, first 4 letters of username and password is selected. The passphrase is framed in such a way that it is easy for the user to remember and difficult for others to guess. This passphrase is used by the user to login to the system but not in the original form. The generated passphrase is divided into different parts according to trigraph substitution method (each part contain 3 letters of passphrase).Each trigraph is substituted with the numbers from 0 to 9. Passphrase is arranged in jumbled order and the order is changed according to the security questions displayed. Regarding security questions, each one is assigned an image according to the category of question so that the user is able to identify the question by simply viewing the image instead of displaying it in a text format. The answer is also mapped into a corresponding image.

ii. Login phase

There are two steps in login process. In the first step, an image is displayed as a clue for the security question and an area for selecting the image as answer. Security questions are displayed randomly so that anyone from outside can't understand the actual question. The user is directed to second step only by completing the first step successfully. In the Second step, the user is instructed to enter the password. The trigraph substitution of passphrase with numerical values in the first phase is set as the password. The correct order of passphrase is accepted and the user is allowed to login into the system. The method is robust as It does not use *text based password* and it provides better security against shoulder surfing attack.

This is an ongoing research work and is focusing on improving security of the system and time for login. In order to evaluate the efficiency of the method in shoulder surfing, a usability study is conducted with 70 computer literate people. The current result shows that the system performs well with good margins.

VI. *CONCLUSION*

Shoulder surfing refers to a direct observation of PINs by looking over a person's shoulder or camera-based recording, to obtain information. The entry of a password can easily be

ISBN -978-93-5406-047-02

observed in crowded place by standing next to someone. In this paper, a novel approach to resist shoulder surfing attacks using passphrase based authentication with trigraph substitution is proposed. We have also made a comparative study on different methods used to prevent shoulder surfing attacks like color pass method, TicToc pin entry method and flywheel pin entry method using some parameters like security, usability and time taken for login. From the analysis, it is found that color pass is more secure and less complex among other methods. In terms of user friendliness, flywheel performs well than other methods. From the current studies it is seen that the proposed method outperforms state of the art techniques to prevent shoulder surfing attacks .The method is expected to give better results by performing few more experimental evaluations.

REFERENCES

[1] Sarika, S., & Paul, V. Intelligent Agents in Securing Internet. Journal of Internet Technology, 19(3), (2018, May) pp. 753-763.

[2]Reddy, V. P., Radha, V., & Jindal, M. .Client Side Protection from Phishing Attack. *Journal of Advanced Engineering Sciences and Technologies (IJAEST), 3*(1), (2011, January)., pp. 39-45.

[3]Liu, W., Deng, X., Huang, G., & Fu, A. Y. An antiphishing strategy based on visual similarity assessment. *IEEE Internet Computing, 10*(2), pp. (2006, March). 58-65.

[4] Sarika, S., & Paul, VParallel Phishing Attack Recognition using Software Agents. Journal of Intelligent & Fuzzy Systems, 32(5), (2017, April). pp. 3273-3284.

[5] Krombholz, K., Hobel, H., Huber, M., & Weippl, E. Advanced social engineering attacks. *Journal of Information Security and applications, 22*, pp. (2015, June). 113-122.

[6] T. Matsumoto and H. Imai, "Human identification through insecure channel," in Advances in Cryptology—EUROCRYPT. Berlin, Germany: Springer-Verlag, 1991, pp. 409–421

[7] C.-H. Wang, T. Hwang, and J.-J. Tsai, "On the Matsumoto and Imai's human identification scheme," in Advances in Cryptology— EUROCRYPT. Berlin, Germany: Springer-Verlag, 1995, pp. 382–392.

[8] X. Bai, W. Gu, S. Chellappan, X. Wang, D. Xuan, and B. Ma, "PAS: Predicate-based authentication services against powerful passive adversaries," in Proc. IEEE Annu. Comput. Secur. Appl. Conf., Dec. 2008, pp. 433–442

[9] H. Zhao and X. Li, "S3PAS: A scalable shoulder-surfing resistant textual-graphical password authentication scheme," in Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl. Workshops, vol. 2. May 2007, pp. 467–472.

[10] D. Weinshall, "Cognitive authentication schemes safe against spyware," in Proc. IEEE Symp. Secur. Privacy, May 2006, pp. 294–300.

[11] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," in Proc. ACM Int. Working Conf. Adv. Vis. Inter., 2006, pp. 177–184.

[12] A. De Luca, K. Hertzschuch, and H. Hussmann, "ColorPIN—Securing PIN entry through indirect input," in Proc. ACM CHI Conf. Human Factors Comput. Syst., 2010, pp. 1103–1106.

[13] T. Perkovic, M. Cagalj, and N. Rakic, "SSSL: Shoulder surfing safe login," in Proc. Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM), Sep. 2009, pp. 270–275.

[14] V. Roth, K. Richter, and R. Freidinger, "A PIN-entry method resilient against shoulder surfing," in Proc. 11th ACM Conf. Comput. Commun. Secur. (CCS), 2004, pp. 236–245

[15] G. T. Wilfong, "Method and apparatus for secure PIN entry," U.S. Patent 5940511, Aug. 17, 1999.

[16] D. Kim et al., "Multi-touch authentication on tabletops," in Proc. ACM SIGCHI Conf. Human Factors Comput. Syst. (CHI), 2010, pp. 1093–1102.

[17] Q. Yan, J. Han, Y. Li, J. Zhou, and R. H. Deng, "Designing leakage resilient password entry on touchscreen mobile devices," in Proc. ASIA CCS, 2013.

[18] H. J. Asghar, S. Li, R. Steinfeld, and J. Pieprzyk, "Does counting still count? Revisiting the security of counting based user authentication protocols against statistical attacks," in Proc. 20th Symp. Internet Soc. Netw. Distrib. Syst. Secur. (NDSS), Apr. 2013, pp. 1–18.

[19] Q. Yan, J. Han, Y. Li, and R. H. Deng, "On limitations of designing leakage-resilient password systems: Attacks, principals and usability," in Proc. 19th Symp. Internet Soc. Netw. Distrib. Syst. Secur. (NDSS), Feb. 2012.

[20]. L. Sobrado "Graphical passwords," The Rutgers Scholar, An Electronic Bulletin for Undergraduate Research, vol. 4,2002

[21]. H. Gao, X. Liu and R. Dai, "Design and analysis of a graphical password scheme," Proc. of 4th Int. Conf. on\ Innovative Computing, Information and Control, Dec. 2009, pp. 675-678.

[22]. Schemes using text-graphical passwords," International Journal of Information & Network Security, vol. 1, no. 3, pp. 163170, Aug. 2012.

[23]. G. Wilfong, "Method and appartus for secure pin entry." US Patent No.5,940,511, In Lucent Technologies, Inc., Murray Hill, NJ, U. S. Patent, Ed. United States, 1997.

[24] : N. Chakraborty, S. Mondal, Color pass: An intelligent user interface to resist shoulder surfing attack, 2014 IEEE, IEEE, 2014

[25] : Taekyoung Kwon, Member, IEEE, and Jin Hong, Analysis and Improvement of a PIN-Entry Method Resilient to Shoulder-Surfing and Recording attacks , IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 2, FEBRUARY 2015

[26]: Ms. Ojaswi K. Kasat, Dr. Umesh S. Bhadade, Revolving Flywheel PIN Entry Method to Prevent Shoulder Surfing Attacks, 2018 3rd International Conference for Convergence in Technology (I2CT) ,IEEE 2018,Apr 06-08, 201

ISBN -978-93-5406-047-02

A Survey on IoT Operating Systems

Amal Antony

Department of Computer Science, Naipunnya Institute of Management and Information Technology, Pongam, Thrissur

contactamalantony@gmail.com

Abstract— **An IoT development board is a small-form-factor system, complete with microprocessor(s), memory, input/output functions providing the user with all the features of a functional computer. The MCU based smaller variants house limited hardware resources and do not demand an operating system. But, the more powerful single board computers require an operating system to efficiently manage its resources and control the hardware. The choice of operating system depends on the microcontroller architecture, on-board memory, software stack used, real-time computing requirements, implementation environment and cost of the system. Operating systems for IoT applications require additional functionalities like network support, power usage monitoring, secondary storage management, multithreading and so on. This paper intends to survey the different IoT operating systems available in the market and studies the various considerations on the selection of OS for IoT development boards.**

Keywords— **IoT, operating system, embedded system, smart devices, embedded Linux, open source, development board**

I. INTRODUCTION

The "Internet of Things" abbreviated as IoT is a comprehensive model including all kinds of computing devices, that are connected to the Internet. These devices are otherwise called “things” or “smart objects” [1]. In theory, a device can be attached to almost any real world object like vehicles, home appliances, industrial, mechanical or electrical machines and even a person to let the object to communicate to the Internet. IoT finds applications in buildings and home automation, smart cities, smart industry or manufacturing, wearables, healthcare devices and automotive. There is also great interest in extending IoT to edge computing and on-device AI capabilities [2]. In order to comply with all of these requirements and applications, an operating system is essential for every IoT device. Having an operating system simplifies the developers’ job and contributes to standardization. Continuous development by industry practitioners and researchers is very essential in this domain so as to provide support for changing hardware configurations and communication standards. An ideal IoT OS has to support different hardware architectures, boards and devices.

There are a number of IoT OSs like Contiki-OS, RIOT and Zephyr to name a few. IoT devices run on low capacity microcontrollers. So, applications running on this platform has to be lean and energy-efficient. It is a prerequisite for an IoT OS to have the essential Transmission Control Protocol/Internet Protocol (TCP/IP) capabilities for seamless integration with the global internet. Apart from providing support for TCP and UDP, modern IoT operating systems are trying to accommodate new standards like Internet Protocol version 6 (IPv6) over Low-Power Wireless Personal Area Networks (6LoWPAN), Routing over Low Power and Lossy Networks (ROLL), Bluetooth Low Energy (BLE) and Bluetooth Meshes.

This paper aims to establish the need for an IoT operating system and provide a framework for choosing an OS. The study is done from a hobbyists or builder’s perspective, that the complexities of academic research is overlooked at some parts. Only those IoT OSs that are available in the market or maintained as open source projects have been considered for the purpose of the study. Proprietary and special purpose software products do not come under the purview of this paper.

The rest of the paper is organized as follows. Section II introduces the common IoT hardware platforms. Section III delves into IoT operating systems in detail, with subsections discussing the functions of an IoT OS, parameters in the selection of an OS and popular IoT operating systems. Section IV discusses the usage and adoption of IoT OSs. Section V includes the conclusion and insights.

II. IOT HARDWARE PLATFORMS

IoT hardware encompasses all devices capable of connecting to the Internet. IoT devices can otherwise mean ‘smart objects. These things or objects are responsible for providing useful information during their transactions on a network. On one hand, we have specialized wearable gadgets like the Google Glass or Fitbit, which are very compact IoT devices. The other category of IoT hardware includes general purpose development boards or Single Board Computers (SBCs) [3]. These development boards allow engineers to create prototypes of IoT solutions and test them. The peculiarity about these development boards is that they are very flexible and can be used to create applications for any domain. This allows open source building and collaboration between engineers. IoT boards can range from an 8-bit MCU to a 32-bit or 64-bit fully functional computer. These boards can have various features like USB interfacing, video output, audio jacks, networking, GPIO pins and wireless communication chips. For the purpose of this study, we will be making a survey of the popular platforms and boards used by IoT developers for initial prototyping, so that we can better understand the synergy between IoT hardware platforms and IoT operating systems. This section details the IoT development boards in common use and their features. A summary of IoT hardware platforms has been included as a table (Table 2.1).

ISBN -978-93-5406-047-02

A. Arduino

Arduino is undoubtedly the favourite among developers and hobbyists. There is an active community working on this platform. Arduino has a broad range of boards, from simple 8-b microcontroller boards to products for wearables, IoT items, three-dimensional (3-D) printing, and much more. The most popular variant of Arduino, the Arduino Uno (Figure 2.1) is based on ATmega328P microprocessor. An Arduino board is programmed with the help of an IDE, using a type B USB cable. The company also offers boards like Arduino Yun, with on-board Wi-Fi (IEEE 802.11 b/g/n) and Ethernet (IEEE 802.3 10/100Mb/s) [3]. But, Arduino is known around the world for its low-cost 8-bit and 16-bit models. Low end boards like these does not require an OS. Code for Arduino follows the structure and conventions of C language. Support for peripherals, sensors and actuators is provided through header files. Once the code is saved on the board, it is executed in an endless loop [4].

Fig. 2.1 Arduino Uno

B. Raspberry Pi

Raspberry Pi (Figure 2.2) is a family of single-board computers that come with great computing power in small package and provides the functionalities of a desktop computer. Raspberry Pi's development began in 2006 it was finally released on 19 February 2012 as two models: Model A and Model B. After the sale of 3 million units by May 2014, Model B+ was announced in July 2014. Raspberry Pi or RPi, as it is known in the hobbyist circles, can support an operating system and is seen as a low-cost replacement for desktop PCs. The operating system is installed on an SD card and provides you the familiar interface that you expect from a Linux or Windows computer [5]. Raspberry Pi supports a number of operating systems, including Raspbian Linux, Ubuntu Mate, and Windows 10 IoT Core. As Raspberry Pi can maintain an operating system and has support for languages like C/C++, Python, and JavaScript. The latest iteration of the board, the Raspberry Pi 4 features a Quad core 64-bit ARM-Cortex A72 running at 1.5GHz.

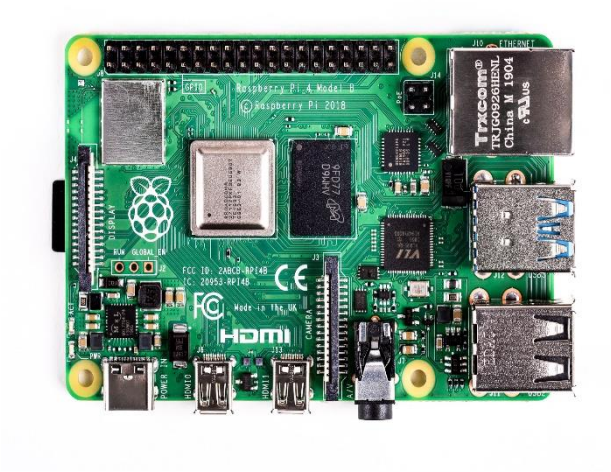


Fig. 2.2 Raspberry Pi Model B

TABLE 2.1
COMPARISON OF IOT HARDWARE PLATFORMS

Brand	Model	CPU	RAM
Arduino	17+ models	ATMega	16KB – 64MB
Raspberry Pi	A,A+,B,B+,3,4,Zero	ARM Cortex-A72, ARM Cortex A-52	1 GB - 4 GB
Nvidia	Jetson Nano	ARM Cortex-A57	4 GB
BeagleBoards	BeagleBone Black	Freescale i.MX233 (ARM926EJ-S core)	512 MB

ISBN -978-93-5406-047-02

C. BeagleBone Black

BeagleBone Black is one member of the community-supported BeagleBoard platforms (Figure 2.3). It is powered by a TI Sitara AM3358 ARM Cortex-A8 processor running at 1 GHz, with 4 GB of on-board flash memory, 512 MB of DDR3L DRAM, and a 3-D graphics accelerator. It has 46-pin headers, an Ethernet port, and several other means to establish communication. It supports the Debian, Android, and Ubuntu operating systems. BeagleBone has been developed as an open-source platform and all of its datasheets are available in BeagleBone community's website [3]. The BeagleBoard, especially the BeagleBone Black version, is easy and inexpensive to set up and use. It consumes low power and thus needs no additional cooling or heat sinks.

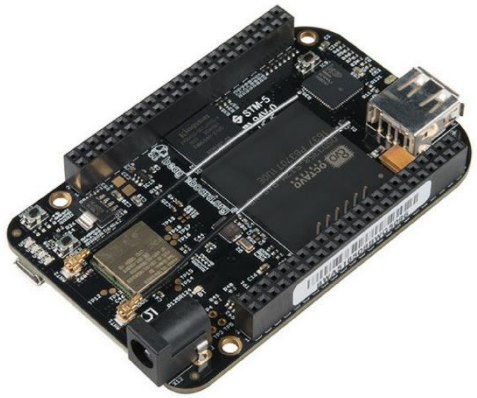


Fig. 2.3 BeagleBone Black

D. NVIDIA Jetson

Jetson is designed as a powerful computer that lets the user run artificial intelligence and machine learning applications (Figure 2.4), while taking up only very little power and space [3]. Jetson comes installed with software libraries for deep learning, computer vision, GPU computing, multimedia processing, and much more. Jetson relies on the Linux environment and provides better performance than other development boards. The availability of proprietary AI and ML tools from NVIDIA reduces the work for developers. One of the biggest advantage that Jetson offers is the GPU-accelerated parallel processing. This makes Jetson an ideal choice for developers looking to implement machine learning projects. Jetson is commonly deployed as Network Video Recorders (NVRs), smart home robots, and intelligent gateways [6].

Fig. 2.4 NVIDIA Jetson Nano

III. OPERATING SYSTEMS FOR IOT DEVICES

A. Definition

An IoT Operating System is a relatively new term and a loosely defined one. There is a lack of literature on the taxonomy and development of IoT operating systems. An IoT Operating System is a piece of software which provides for a channel of interaction between the user and IoT device and manages all hardware and software resources. It is no different from a regular operating system like Unix or Windows. The difference lies in the hardware architecture of the host system and the hardware constraints. An operating system specifically designed for IoT applications which can run under the minimal resources available in an IoT device, can be termed as an IoT OS. It can be said that IoT OS is an extension of embedded system OS. An IoT OS ensures connectivity between IoT applications and embedded systems [7]. Even though IoT OS is an evolution of embedded OS, IoT brings in additional set of constraints and requirements that need to be addressed. Embedded operating systems have been upgraded or augmented to incorporate IoT-specific features. Popular IoT OSs include TinyOS, Contiki, mbedOS, Ubuntu Core, Yocto, Windows 10 for IoT and so on [8], [9].

B. Functions of an IoT Operating System

Technical systems require an operating system as it is the major interface with which the user can interact with the computer and manage how programs functions within the computer system [10]. It takes care of the important processes behind the scenes – managing the hardware resources, providing a user interface and executing and rendering services to application programs [11]. Unlike desktop operating systems or general purpose OS, IoT OS is designed to work with the limited resources and provide capabilities like rapid development tools, standardization, easy maintenance, support

ISBN -978-93-5406-047-02

for various hardware platforms, portability of application programs and seamless integration with global internet [12], [13]. The functions performed by an IoT OS are outlined below:

1) Provides a Hardware Abstraction Layer(HAL): A HAL can be defined as all the software that is directly dependent on the underlying hardware [14]. On more elaborate terms, hardware abstraction layer (HAL) can be defined as a layer of programming or code to allow general communication between the software and hardware components of a system. This reduces the work of application developer. HAL bridges the gap between hardware and software. Otherwise, developers will have to hard-code drivers, kernels, or APIs for each hardware device. This would be a tedious task considering the diversity of IoT hardware platforms. Hardware abstraction provided by IoT operating systems gives developers access to all OS controlled devices like Bluetooth, camera, video output, audio, sensors and storage directly [15].

2) Power Management: Integrated power management techniques have been implemented on SoCs by several manufacturers. The microcontroller has the ability to enter a sleep state or standby mode to save power. The microcontroller (MCU) stops performing computations in sleep state; But, it will retain any current data, and all peripheral functions will be halted [16]. An OS can efficiently manage these sleep states and monitor the power usage of attached peripherals or sensors. IoT devices which run continuously for infinitely long periods of time can benefit from this power saving technique.

3) Concurrency Management: IoT devices today support multi-core or multi-processor setups. This makes concurrent computing an important element of embedded computing as well. The traditional methods used for parallel programming used in regular, 'mature' operating systems are not suitable for IoT systems with time and resource constraints and raises the possibility of dead-locks. So, algorithms which prioritize these requirements have to be implemented for IoT OS [17]. In the case of peripheral control or reading data from sensors, concurrent execution is not as important. The usage of event-driven asynchronous execution or collective IO is a viable solution. But when it comes to data and signal processing systems, the OS has to manage the parallelization of processes.

4) User Interface: As mentioned before, OS provides an interface between the user and the device. Some Linux based distributions available for IoT initially boots into the terminal, from where you can invoke a GUI. As in the NOOBS operating system for Raspberry Pi, the GUI is started by typing 'startx' in the terminal [5]. Most IoT OS vendors go with a graphical interface considering the large number of hobbyists or amateurs as users.

The input and output operations can also be discussed under this head. Inputs to an IoT device maybe through USB interface or GPIO pins. Most IoT development boards provides a full-fledged video output. With additional hardware, IoT boards can also support touch screen input, the example of which can be seen in a POS system or ATM.

5) Memory Management: An IoT OS has the responsibility of managing the primary memory of the device. The memory management function keeps track of the current status in every memory location, whether it's allocated or free. It measures how memory is allocated over processes, deciding which gets memory, when they receive it, and how much they are free [18].

6) Process Scheduling: Regarding an IoT OS, processes Scheduling simply refers to managing the processes currently in the system memory. In very simple terms, it decides the order of executing things [19]. Scheduling prioritizes processes, loads them into the ready queue of the system and then send for the execution. Short, medium, and long-term scheduling can be implemented by the operating system [18].

7) Shared Libraries: The operating system provides several utility libraries for the developer, which is available to all supported development platforms and programming languages. These dynamically linked libraries can be used by any number of application programs without making copies in the main memory. A lot of memory can be saved this way and is appreciable, especially in the resource constrained environment of IoT. Developers can make use of these shared libraries and reduce the computing overhead. The functionalities like logging utility, TCP/IP, cryptography, time synchronization are provided by shared libraries [19]. Another example would be the SELinux library providing security in Linux systems [20].

8) Hardware Virtualization: The primary motive behind IoT device virtualization is to different IoT devices and service functions with several applications. Virtualization enables the limited hardware resources to be shared among multiple users [19]. It ensures efficient usage of the resources available and restricts access to particular group of users. In the wider sense, virtualization also includes dockers and containers, creating secure application environments which can be deployed over cloud or local network [21].

C. Parameters for the Choice of an IoT Operating System

The following parameters must be considered before making the selection of an IoT operating system:

1) Reliability and Stability: An OS installed on IoT devices must not crash unexpectedly. IoT devices are supposed to run without powering down, for a very long time. If the OS crashes or shows glitches, debugging the system software after deployment is a difficult job to carry out. An OS which has support from vendor or an active user community must be selected.

ISBN -978-93-5406-047-02

2) Footprint: IoT devices are bound to operate with limited resources. The OS is expected to have low memory, power and processing requirements [12]. The scheduling and process management algorithms must also suit the IoT paradigm.

3) Scalability: Scalability here refers to the ability to extend the OS to operate on the two types of IoT devices – nodes and gateways [8]. This way, system architects and administrators will only have to deal with a single OS, working on all kinds of devices. Scalability can also mean the ability of the OS to improve over time through updates [10]. A scalable OS which is able to run on a variety of 8-bit,16-bit or 32-bit microcontrollers will use less resources.

4) Portability: Portability means that a system developed in one environment it should execute in another environment without the need for rewriting the code [22]. This quality allows developers to switch between IoT hardware platforms without the need to alter the OS. Portability and adaptability is an important feature of embedded system. It is also recommended to have some standard interface (e.g., POSIX), for good portability of applications, for minimizing maintenance, as well as to provide the ability to easily connect to other devices on the Internet [23].

5) Modularity: Apart from the kernel, every functionality can be designed as an add-on to the operating system so that a minimal version of the OS can be run if the situations demand it [8]. A modular architecture allows to replace or add kernel components dynamically at run time. In a modular kernel, components providing similar functionality will be stored in files called modules and can be loaded when the system needs that functionality [24]. An IoT device will require a modular operating system that separates the core kernel from middleware, protocols, and applications.

6) Hardware agnostic operation – There are several IoT hardware platforms available in the market. Keeping this in mind, it is important that the OS supports the leading platforms – leading to standardization and ease in deployment [12].

7) Network Connectivity & Protocol Support: The Internet of Things is a model of "connected" devices. So, the OS must support different connectivity and networking protocols, such as Ethernet, Wi-Fi, BLE, IEEE 802.15.4, TCP/IP, etc. [1]. An IoT OS will allow us to select the specific protocol stacks we need, saving memory on the device, and reducing our costs. It can help upgrade existing devices to new connectivity options without altering the core code of the OS [23].

8) Security: The operating system must provide the first layer of security for the IoT system. As the device is continuously communicating with the Internet, it is susceptible to threats and attacks. OS can have add-ons that bring security to the device by way of encryption SSL/TLS certification management, user authentication routines, VPN and firewall [12].

9) Eco-system & Application Development: The suitability of the OS for application development and debugging can make a big impact on the speed of development and time-to-market. If the OS is developer friendly, deployment of IoT applications can be done more efficiently. In addition to the basic C environment, the use of other programming languages and libraries is highly desirable, for example python, C ++ and STL, but that highly depends on the development toolchain adopted by the developer or organization [23].

D. Popular IoT Operating Systems

A partial list of IoT operating systems is produced below.

1) Mbed OS: MbedOS is the Real Time Operating System (RTOS) developed by ARM delivered as an open source product with an Apache 2.0 license. Mbed OS is designed for Cortex-M microcontrollers, and incorporates a tiny RTOS based on CMSIS-RTOS RTX, TLS protocol, networking standards and common device drivers in a modular architecture [1]. It is specifically designed for 32 bit ARM architecture. MbedOS supports features such as multithreading, 6LoWPAN, BLE, Wi-Fi, sub-GHz, Near Field Communication (NFC), Radio-Frequency Identification (RFID) and Long Range Low-Power Wide Area Network(LoRaLPWAN) [13]. Minimal system requirements and support for different development boards make it a highly preferred IoT OS.

2) Contiki: Among the IoT research community, Contiki has greater acceptance. The low memory requirements make Contiki well suited for low power devices. It is written in C language. Contiki offers multithreading through protothread and uses the cooperative or preemptive scheduling for the processes [13]. Contiki provides support for multiple network stacks with a comprehensive set of features like IPv6, 6LoWPAN, RPL and CoAP. It can run on IoT platforms like wismote, sky and z1. Contiki and its code simulator Cooja has been used as a development tool in several wireless sensor projects [1].

3) RIOT OS: RIOT is an open source IoT operating system. RIOT was initially developed as part of a research project by FU Berlin, INRIA, and HAW Hamburg. It is based on the microkernel named FireKernel, that was targeting wireless sensor networks. RIOT is designed to be energy efficient and modular with very low memory requirements [1]. RIOT implements modular design and uniform API access for independent hardware abstraction. RIOT supports C and C++ programming languages. It also provides multithreading with tickless, pre-emptive and priority based scheduler. RIOT also offers an emulator called Native which acts as a hardware virtualizer, helping in application development without actually

ISBN -978-93-5406-047-02

having a development board [13]. RIOT has support for hardware architectures such as AVR, ARM7, Cortex-M0 -M0+ -M3 -M4 -M7, Cortex-M23, ESP8266, ESP32, MIPS32, MSP430, PIC32, RISC-V and x86 [25]

4) Apache Mynewt: Mynewt is an open source OS with Apache License 2.0, developed by the Apache Software Foundation [1]. The OS features a flexible and powerful Bluetooth Low Energy stack (BLE 5) implementation called NimBLE which provides the option to choose HOST only or CONTROLLER only or FULL stack. Mynewt facilitates cross-platform migration as it supports a great number of hardware platforms. It is designed to be hardware agnostic and can work with Cortex M0-M4 micro controllers, MIPS and RISC-V. It is designed as a pre-emptive, multi-tasking real time operating system kernel. Mynewt's Hardware Abstraction Layer (HAL) abstracts the MCU's peripheral functions, allowing developers to easily write cross-platform code [26]. All these features make Mynewt suitable for IoT boards with low computing power and memory resouces.

5) Zephyr: Zephyr is an open source OS maintained by the Linux Foundation. Zephyr works on two OS design philosophies- a microkernel for less constrained IoT devices and a nanokernel for constrained devices. It supports multithreading with cooperative, priority-based, Earliest Deadline First (EDF), non-preemptive and preemptive scheduling [13]. The Zephyr OS is based on a small-footprint kernel designed for use on resource-constrained and embedded systems ranging from a simple embedded environmental sensors and LED wearables to sophisticated embedded controllers, smart watches, and IoT grids. The Zephyr kernel supports multiple architectures, including ARM Cortex-M, Intel x86, ARC, NIOS II, Tensilica Xtensa and RISC-V 32 that makes the OS compatible with over 200 development boards [27]. C and C++ are the languages used to develop applications in Zephyr. Zephyr provides a complete network stack for communication and includes multiple protocols. The applications can be developed, built and tested using the native posix port.

6) Android Things: Android Things is an IoT operating system developed and maintained by developed by Google. Google announced its IoT OS Brillo at Google I/O 2015, which was rebranded to Android things. As the name indicates, it is based on android which can be considered a simplified and trimmed down android version to run on low-power IoT devices. It supports development in both C and C++ programming language. It is built on top of monolithic kernel and provides completely fair scheduler [13]. The Peripheral I/O APIs allow apps to communicate with sensors and actuators using industry standard protocols and interfaces. The interfaces supported by Android Things are GPIO, PWM, I^2C, SPI, UART. Apps for IoT devices can be built using existing Android development tools, APIs, and resources along with new APIs that provide low level I/O and libraries for common components like temperature sensors, display controllers, and more [28]. Android Things provides a GUI interface, but its high memory requirements make it unsuitable for low-end constrained IoT devices rather it is designed for high-end IoT devices. Currently, Android Things OS supports two development boards - Raspberry Pi 3 Model B and NXP i.MX7D.

7) Windows 10 IoT: Microsoft ventured into embedded systems domain with Windows CE. Microsoft has withdrawn support for Windows CE and has moved to Windows 10 IoT. The operating system includes the stability and user-friendliness of Windows family of products. Windows 10 IoT comes in three flavours: IoT Enterprise, IoT Core and Server IoT. Licensing of the OS is done through OEM channels. Where reliability and safety are important, Windows OS is preferred over free or open source ones. Windows 10 IoT finds applications in aerospace, automotive, healthcare and industrial systems. Microsoft has built several other products around IoT. Windows 10 IoT bring Artificial Intelligence (AI) and Machine Learning (ML) to smart devices with Windows ML and support from Azure IoT Edge [29]. Windows 10 IoT supports boards like AAEON Up Squared, DragonBoard 410c, NXP i.MX 7, NXP i.MX 8M/8M Mini and Raspberry Pi 2/3B.

8) Embedded Linux: Embedded Linux does not refer to an individual OS; it is a categorization. Rather than discussing all Linux based operating systems separately, all of those products have been discussed under this head. Linux is a very versatile environment suitable for IoT development and is very adaptable in nature. Some examples of embedded linux distributions are Raspbian, Yocto, Ubuntu Core, RTLinux, OSMC, Arch Linux ARM, Gentoo and openSUSE. Embedded Linux's popularity can be attributed to its core characteristics: reliability, configurability and low system requirements. Linux works only on embedded systems with at least a 32-bit address space [30]. Contributions from developers all over the world is making Linux stronger by time. Commercially licenced and "closed" operating systems are not recommended if modifications are to be made to match the requirements of the deployment environment. The flexibility of Linux, combined with consistency in performance, architectural tiers, virtualization and cloud deployment environments makes Linux distros a popular choice among IoT developers [12].

IV. DISCUSSION

Section III of this paper has detailed the popular options available for IoT OS. It can be observed that most of the system software products mentioned in Section III are open source projects. If we choose open source solutions over commercial software, then the factors like technical specifications (such as support for hardware, functionality, reliability, performance, network connectivity and standards), software license of the product, project governance, founding members, commercial and community supports, size of users should be considered.

ISBN -978-93-5406-047-02

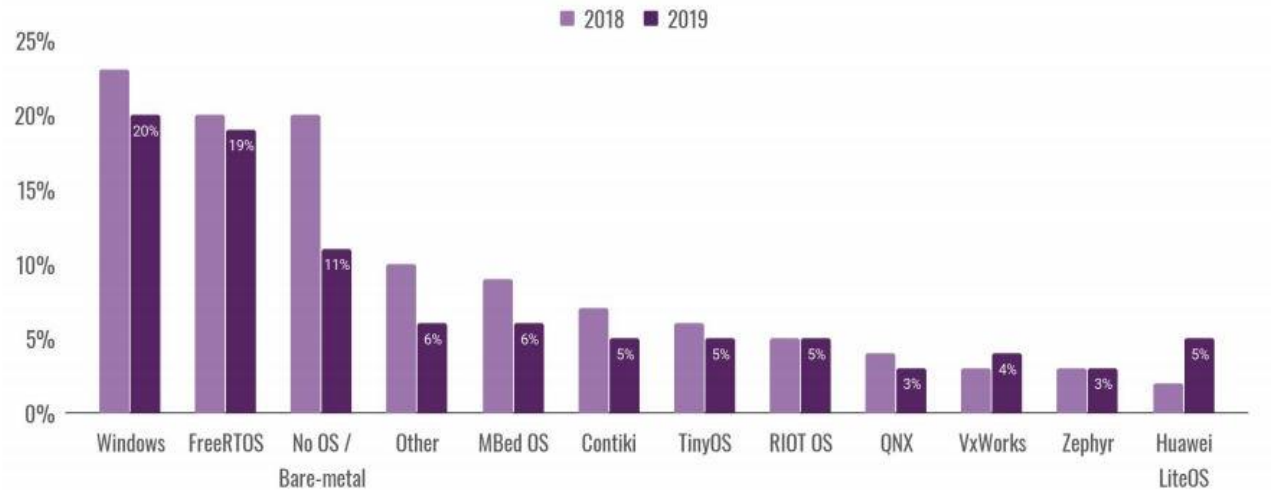


Fig 4.1 Usage of non-Linux IoT OSs (2018-19)

Contiki, Zephyr, Mynewt, and RIOT introduced in this paper are open source system software without integrated commercial cloud services [1]. Contiki and RIOT are both mature system software projects with an active developer community.

Google's Android Things in combination with Android operating system, Google cloud infrastructure, and other cloud services can be used in high-performance devices with x86 architecture. The only downside is that the OS will exclusively support Google product portfolio and will not let the developers to test software products or services from other vendors.

The other leading environment is ARM mbed OS. It has the backing of ARM and targets low-performance MCU devices. Mbed has greater importance consider the share of ARM devices in the market.

Embedded Linux is also a favourite among developers. Some Linux variants are delivered as free operating systems, while others are maintained as open source projects.

Windows 10 IoT is a commercial solution and widely adopted. 52% of Edge nodes or Gateways use Windows operating system [31]. So, developers adopt Windows OS on IoT devices to ensure better communication between all devices on the network. Considering wide range of customizable services and software offered by Microsoft, Windows 10 IoT is expected to improve its position in the coming years.

To provide perspective on the usage and adoption of IoT OSs, the IoT Developer Survey 2019, undertaken by the Eclipse Foundation can be referred here. Linux held on to its top position among operating systems. The survey does not disclose the share of Linux in total IoT usage, but shows Windows holding second position with 20 percent. Other than Linux, the other popular choices include Windows, mbed OS, RIOT OS, Contiki and others. The share of non-Linux operating systems is illustrated through a graph (Fig 4.1) [31].

V. CONCLUSION

IoT operating systems are used by hobbyists, developers and researchers. There are large number of options available in the market, both free and commercial distributions. Open source development and cross platform applications are experiencing great growth in the field of IoT. The availability of low cost development boards makes IoT more pervasive, open and community driven. This has encouraged industry leaders like Google, Microsoft, Canonical, Intel, ARM, MATLAB and Apache to venture into developing system software or a complete software suite for IoT solutions. Innovations like NVIDIA Jetson, ThingSpeak [32] and TensorFlow Lite [2] indicates the IoT industry is working towards a better coupling with evolving technologies like cloud computing, data analytics, machine learning and GPU computing.

Single Board Computers are not just tools for prototype development; they are adopted in several parts of the world as low-cost alternatives to desktop computers. The introduction of simple and intuitive operating systems like Raspbian, KODI distributions and Android has a great role to play in that. So, there is a need to create more general purpose, customer oriented operating systems targeting SBCs. IoT solutions can be developed for any domain. But in general, there is a convergence towards creating a connected world as proposed by concepts like Internet of Everything and Campus of Things. In the years to come IoT will grow into a mature technology and there will be standardization in IoT operating systems and development environments.

ACKNOWLEDGMENT

The author would like to acknowledge the support provided by the academic institution Naipunnya Institute of Management and Information Technology in completing this paper. The information and insights derived from community discussion forums and developer groups is also thankfully acknowledged.

ISBN -978-93-5406-047-02

REFERENCES

[1] Amiri-Kordestani, Mahdi, and Hadj Bourdoucen. "A survey on embedded open source system software for the internet of things." Free and Open Source Software Conference. Vol. 2017. 2017.

[2] "TensorFlow Lite," TensorFlow. [Online]. Available: https://www.tensorflow.org/lite

[3] K. J. Singh and D. S. Kapoor, "Create Your Own Internet of Things: A survey of IoT platforms.," in IEEE Consumer Electronics Magazine, vol. 6, no. 2, pp. 57-68, April 2017

[4] "loop()," Arduino Reference. [Online]. Available: https://www.arduino.cc/reference/en/language/structure/sketch/loop/

[5] Pritish Sachdeva and Shrutik Katchii, "A Review Paper on Raspberry Pi", International Journal of Current Engineering and Technology, Vol.4, No.6 ,pp. 3818-3819, 2014.

[6] "Bringing the Power of AI to Millions of Devices," *NVIDIA*. [Online]. Available: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/

[7] "Top 15 Best IoT Operating System For Your IoT Devices in 2020", UbuntuPIT, 2020. [Online]. Available: https://www.ubuntupit.com/best-iot-operating-system-for-your-iot-devices/

[8] a. vikasG, "IoT Operating Systems", Devopedia, 2020. [Online]. Available: https://devopedia.org/iot-operating-systems

[9] D. Guinard, "Operating Systems for IoT Embedded Systems – Web of Things", Webofthings.org, 2020. [Online]. Available: https://webofthings.org/2016/12/12/iot-os-embedded/

[10] A. Prabhu. S, G. Prabhu and P. R, "A STUDY OF OPERATING SYSTEM FOR EMBEDDED SYSTEMS", International Journal of Latest Trends in Engineering and Technology, no., pp. 54-58, 2016. Available: https://www.ijltet.org/journal/148299172610.pdf

[11] Operating Systems. [Online]. Available: https://homepage.cs.uri.edu/faculty/wolfe/book/Readings/Reading07.htm

[12] "IoT Operating Systems," Arrow.com, 10-Sep-2018. [Online]. Available: https://www.arrow.com/en/research-and-events/articles/iot-operating-systems

[13] Y. B. Zikria, S. W. Kim, O. Hahm, M. K. Afzal, and M. Y. Aalsalem, "Internet of Things (IoT) Operating Systems Management: Opportunities, Challenges, and Solution," Sensors, vol. 19, no. 8, p. 1793, 2019

[14] S. Sungjoo and A. Jerraya, "Introduction to Hardware Abstraction Layers for SoC," in Embedded Software for SoC, Boston, MA: Springer, 2003, pp. 179–186.

[15] "Hardware Abstraction: Definition & Purpose", Study.com [Online]. Available: https://study.com/academy/lesson/hardware-abstraction-definition-purpose.html

[16] B Kumar, "The Role of Sleep Mode in Embedded Systems", eeweb, 2020. [Online]. Available: https://www.eeweb.com/profile/kumarb/articles/the-role-of-sleep-mode-in-embedded-systems

[17] Schramm, Norbert, and Anita Sabo. "Concurrent programming method for embedded systems." 9th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics. Vol. 41. 2008.

[18] Wael Alabdulaly," Memory Management techniques and Processes Scheduling", International Journal of Scientific & Engineering Research", Volume 7, Issue 4, pp. 1182-1184, 2016

[19] "Embedded Operating Systems for the IoT," cs.virginia.edu. [Online]. Available: https://www.cs.virginia.edu/~bjc8c/class/cs6501-f18/

[20] "Main Page," SELinux Wiki. [Online]. Available: http://www.selinuxproject.org/page/Main_Page

[21] Ogawa, Keigo, et al. "IoT Device Virtualization for Efficient Resource Utilization in Smart City IoT Platform." 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2019.

[22] Jabeen, Qamar, et al. "A survey: Embedded systems supporting by different operating systems", International Journal of Scientific Research in Science, Engineering and Technology ,Vol.2, Issue 2,pp. 664-673, 2016.

[23] Milinković, Aleksandar, Stevan Milinković, and Ljubomir Lazić. "Choosing the right RTOS for IoT platform." Proceedings of the international scientific professional symposium Infoteh, Jahorina. 2015.

[24] Qutqut, Mahmoud H., et al. "Comprehensive survey of the IoT open-source OSs." IET Wireless Sensor Systems 8.6 (2018): 323-339.

[25] "The friendly Operating System for the Internet of Things. Learn more.," RIOT. [Online]. Available: https://www.riot-os.org/#usage

[26] "Apache Mynewt," Apache Mynewt. [Online]. Available: https://mynewt.apache.org/

[27] "Supported Boards," Supported Boards - Zephyr Project Documentation, 14-Feb-2020. [Online]. Available: https://docs.zephyrproject.org/latest/boards/index.html

[28] "Android Things 1.0 Features and APIs : Android Developers," Android Developers. [Online]. Available: https://developer.android.com/things/versions/things-1.0

[29] Terry Warwick, "Overview of Windows 10 IoT Core - Windows IoT," Overview of Windows 10 IoT Core - Windows IoT | Microsoft Docs. [Online]. Available: https://docs.microsoft.com/en-us/windows/iot-core/windows-iot-core

[30] Embedded Linux. [Online]. Available: https://www.itu.dk/research/rces/emli.html

[31] Iot.eclipse.org, 2020. [Online]. Available: https://iot.eclipse.org/resources/iot-developer-survey/iot-developer-survey-2019.pdf

[32] "ThingSpeak for IoT Projects," IoT Analytics - ThingSpeak Internet of Things. [Online]. Available: https://thingspeak.com/

ISBN -978-93-5406-047-02

A Novel Approach To Sketch Based Image Retrieval using Unsupervised Learning and Shape Descriptors

Jisma Wilson
Student
jismawilson4@gmail.com
Sacred Heart College, Thevara

Arya Chandran
Student
aryachandran1999@gmail.com
Sacred Heart College, Thevara

Shailesh S
Assistant Professor
shaileshsivan@gmail.com
Sacred Heart College, Thevara

Dr. Regitha M.R.
Head Of Department
regitha.baiju@gmail.com
Sacred Heart College, Thevara

***Abstract*—The introduction of search engines for information retrieval lights a path to opportunities in the field of science and technology. Different types of inputs like text, images, sounds can be used as query for modern search engines. Even the features considered for the retrieval process is limited. In this paper, a sketch based image retrieval system based on Hu moment and other similarity measures are introduced and it works with the image processing and unsupervised machine learning. The objective of the system is to retrieve images based on the given query sketch. For every searching process, an index structure is very important and it is done using k-means clustering using Hu moment as feature and retrieval process is done by matching the query with image in the target cluster. For the similarity comparison, measures like Scale Invariant Feature Transform (SIFT), structural similarity and pixel similarity are used. Overall, the system gave 76.6% accuracy rate.**

***Index Terms*—SBIR, Unsupervised learning, k-means clustering, Hu moment, Indexing, SIFT, Retrieval**

I. INTRODUCTION

In the recent past, image retrieval has emerged to be a prominent research field due to the rapid increase in the collection of digital images for various purposes. The traditionally and most commonly used methods are: image meta search, content-based image retrieval, image collection exploration

Image meta search is the type of image retrieval wherein the user can input related keywords or text. When the description of an image is not known, an image can be used as a query in order to retrieve images from the dataset based on their characteristics. This method of image retrieval is known as Content-Based Image Retrieval (CBIR). Another method of image retrieval is image collection exploration which focuses on searching of images based on novel paradigms. [1]

Among these methods, currently the most widely used is CBIR. However, there is a difficulty in finding an image which is similar to what the user wishes to retrieve. To overcome this problem, researchers have come up with an efficient and user friendly method known as the Sketch Based Image Retrieval (SBIR).

SBIR is the method by which a user inputs a hand drawn sketch which is analyzed and the related images are displayed to the user. The various methods applied to analyze the sketch comes under two categories which are global and local techniques. Global techniques include Edge Histogram Descriptor (EHD), Histogram of Edge Local Orientation (HELO) and Angular Partitioning of Abstract Images (APAI) whereas Shape Context, Structure Local Approach (STELA) comes under local techniques. [1]

A sketch-based image retrieval system based on Hu moment and other similarity measures such as SIFT are introduced in this paper. This system works with image processing and unsupervised machine learning.

II. LITERATURE REVIEW

Jose M. Saveedra and Benjamin Bustos [2], in their paper, proposed a novel local approach based on detecting simple shapes called keyshapes rather than keypoints. Abdolahchalechale , Golshah Naghdy and Alfred Mertins [3], proposed a method of SBIR using Angular partitioning of the abstract image by exploiting the Fourier Transform to identify the features which are scale, rotation and translation invariant. Yang Cao, Changhu Wang, Liqing Zhang and Lei Zhang [4] introduced a real time search engine called the Mind Finder used to retrieve images using sketches based on the structure, semantic meaning and colour tone. It was the first index-based query-by-sketch method for large scale databases. Rui Hu, Tinghuai Wang and John Collomosse [5] presented a system based on SBIR which divides the image into several regions hierarchically and are represented in the Bag Of Regions (BOR) which includes shape of objects at multiple scales. Using the Gradient Field HoG (GF-HOG) descriptor, we extract shapes from each region so as to directly compare with the sketched query. For a faster SBIR, Li Liu, Fumin Shen, Yuming Shen, Xianglong Liu and Ling Shao [6] together implemented a novel binary coding method known

ISBN -978-93-5406-047-02

as Deep Sketch Hashing (DSH) which introduced a semi-heterogeneous deep architecture and integrated it into the end-to-end binary coding process. Shantanu Deshpande and Naman Goyal [7] developed a task on sketchy database and employed a residual network in the Siamese and Triplet architecture for sketch based image retrieval. Sini Thankachan and Smita C Thomas [8] introduced a system for image retrieval based on face-sketch queries. Scale Invariant Feature Transform (SIFT) is used to extract the features from the sketch. Binary hashing is used to compute the binary hash codes from which the hamming distance is obtained. Based on the hamming distance and Fine Grain Matching, the required images are retrieved. Yonggang Qi, Yi-Zhe Song, Honggang Zhang and Jun Liu [9] proposed a novel Siamese Convolutional Neural Network (CNN) for SBIR. The Euclidian distance between a query sketch and the images are calculated by comparing the positive and negative sketch image pairs samples. Based on the ranking of Euclidian distance, the images are retrieved. Jose M. Saveedra and Benjamin Bustos [10] put forward a method for SBIR based on Histogram of Edge Local Orientation (HELO). To make the system scale, translation and rotation invariant, principal component analysis and polar coordinates were applied.

III. PRELIMINARIES

A. K-MEANS CLUSTERING:

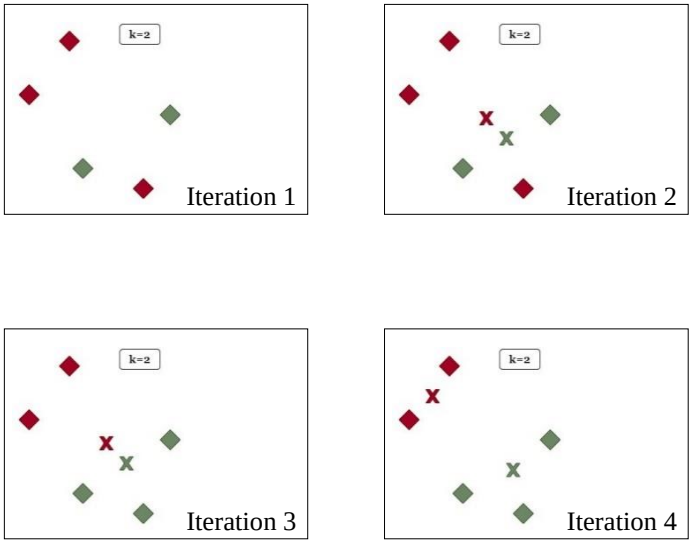


Fig. 1. Iterations of k-Means Clustering

Clustering is basically the division of data into different groups or subsets called as clusters. The main goal of cluster- ing is to maximize intra-cluster similarity and minimize inter- cluster similarity. k-means clustering is one of the simplest unsupervised machine learning algorithms. It is an algorithm to cluster n objects based on attributes into k partitions where k less than n. The process involves randomly assigning n data points to k different clusters (See Iteration 1) and computing the centroid for all the k clusters (See Iteration 2). All the data points will be then re-assigned to the closest cluster centroid (See Iteration 3). This process of computing the centroid and re-assigning the data points based on the centroid value is repeated until there is no switching of data points between different clusters (See Iteration 4). [11]

B. HU MOMENTS

Hu moments is a set of 7 numbers which are determined using the central moments. The first 6 moments are invariant to translation, scale, rotation and reflection but the 7th moment's sign changes for image reflection.

The moments are calculated as follows:

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
&\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
&\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\
&\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned}$$

[12]

id	Image	H[0]	H[1]	H[2]	H[3]	H[4]	H[5]	H[6]
K0		2.78871	6.50638	9.44249	9.84018	-19.593	-13.1205	19.6797
S0		2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S1		2.67431	5.77446	9.90311	11.0016	-21.4722	-14.1102	22.0012
S2		2.65884	5.7358	9.66822	10.7427	-20.9914	-13.8694	21.3202
S3		2.66083	5.745	9.80616	10.8859	-21.2468	-13.9653	21.8214
S4		2.66083	5.745	9.80616	10.8859	-21.2468	-13.9653	-21.8214

[12]

Fig. 2. Hu moment example

IV. METHODOLOGY

For sketch based image retrieval, we have proposed a method using the Hu moment property of images. To ease the

ISBN -978-93-5406-047-02

task, we have also applied k-means clustering for grouping the images based on their similarities. The procedure for the proposed method can be divided into two parts: index creation and image retrieval.

A. Index Creation

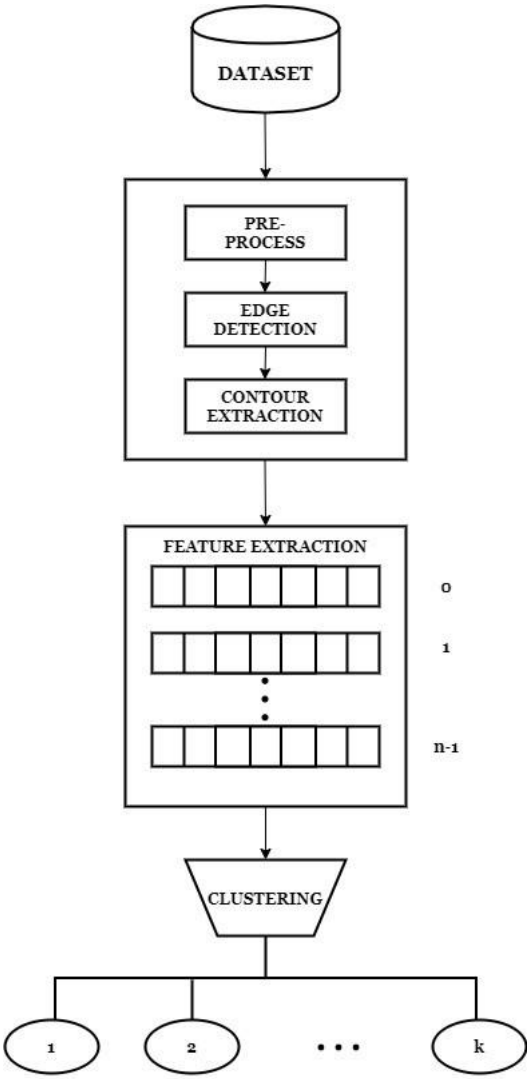


Fig. 3. Architecture of Indexing Process

An index is a data structure used to partition the search space for improving the complexity of searching. The overall architecture of the indexing process is shown in 3. Initially the system consider dataset of n different images which is the search space. Index creation involves clustering of the n images in the dataset to k different clusters. For this, firstly, we pre-process all the images by converting them into grayscale. Further, edge detection and contour extraction are applied to each image. The Hu moment of each image in the dataset is then determined. It is a set of 7 values which are calculated using central moments. Based on the Hu moments calculated, the images are grouped into k different clusters using k-means clustering.

B. Image Retrieval

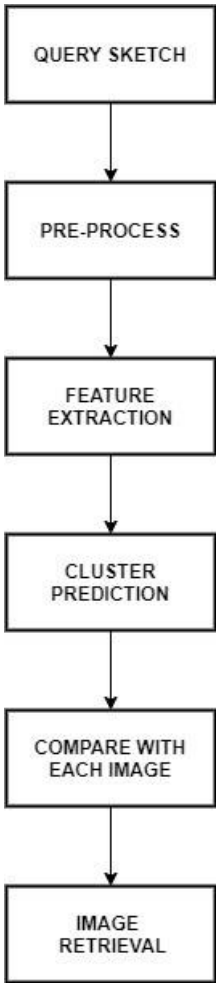


Fig. 4. Flow of Retrieval Process

Second part involves image retrieval based on the user supplied query sketch. The process of retrieval of image is shown in 4. For this, firstly a similar procedure of pre-processing and feature extraction is carried out on the sketch. Once the Hu moment of the sketch has been calculated, it is compared with each of the cluster's Hu moment value. The cluster whose Hu moment is closest is selected. Further, the query sketch is compared to each image in the selected cluster by executing Scale Invariant Feature Transform (SIFT), structural similarity and pixel similarity. SIFT [13] is a feature detection algorithm in computer vision to detect and describe local features in images. SIFT is rotation, scale and transition invariant. In SIFT, from a set of reference images, object keypoints are first extracted and stored into a database. Further, the objects identified from each image is compared with the features already stored in the database. Structural similarity is used to assess the quality of the image. It predicts the visual impact of luminance changes, contrast shifts, as well as any residual mistakes commonly identified as structural shifts. Pixel similarity, as the name suggests is used to compare the

ISBN -978-93-5406-047-02

pixels between images. The results of all three methods are evaluated and the images are provided to the user according to their ranking.

V. IMPLEMENTATION AND RESULT ANALYSIS

As an implementation for this technique we have developed a web application that can be used by two users where one is the admin and the other is the common user. Admin performs the index creation process and any user can input the sketch and retrieve relevant matching images with the help of the index created by the admin. Since we use k-means clustering for index creation, finding the k value is crucial. As a solution, we use the inertia curve. The average value of the curve is taken as the k value.

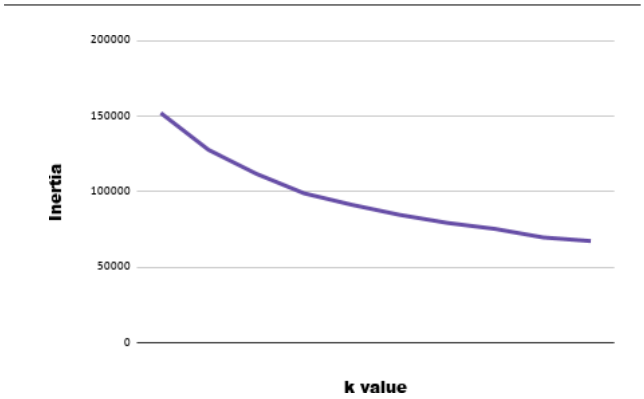


Fig. 5. Inertia curve

In the front end we have used Java Script, HTML and CSS. In the server side we have used Python and packages like OpenCV, Sklearn, Pandas, Flask etc. The results are shown in figure:

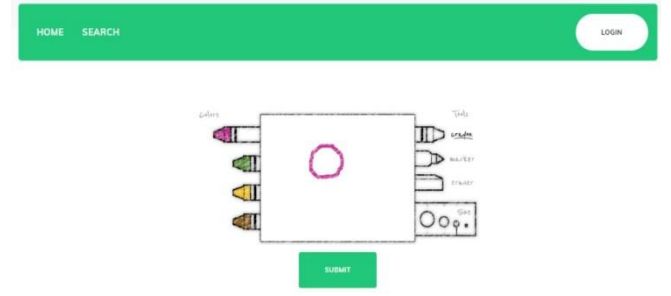


Fig. 6. User supplied query sketch

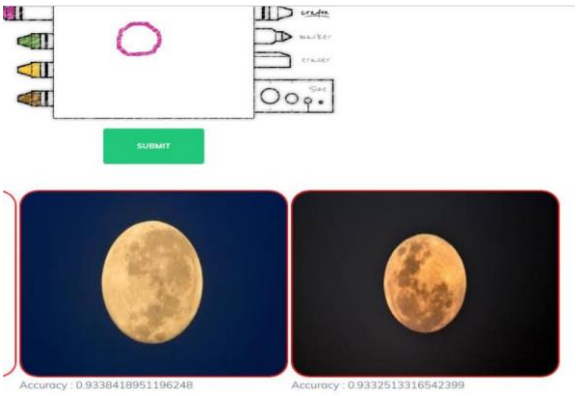


Fig. 7. User output

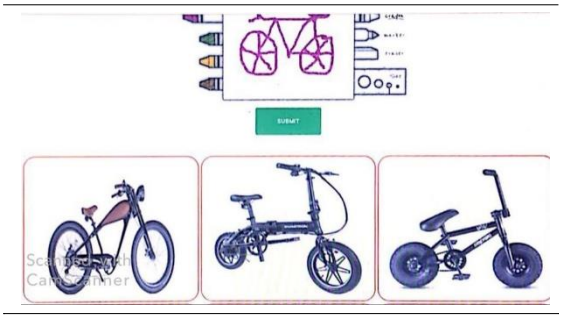


Fig. 8.

Once the system is implemented, to calculate the performance of the system, we have input 30 query sketches out of which 23 were successful, 4 were partially correct and the rest 3 were incorrect. From this, we are able to conclude that the system's accuracy rate is approximately 76.6%. The below table shows the percentage rates of different retrieval cases.

Retrieval	Successful	Partial	Incorrect	Total
Count	23	4	3	30
Percentage	76.6%	13.3%	10%	100%

TABLE I
PERCENTAGE RATE OF DIFFERENT RETRIEVAL CASES

VI. CONCLUSION AND FUTURE WORK

Several search engines were developed in the past years for information retrieval through various methods. In this paper, we have introduced a system for sketch based image retrieval that works with image processing and unsupervised machine learning. For searching process, an index is created using k-means clustering and Hu moment property. As the user inputs a query sketch, it is processed and compared with the images using SIFT and other similarity measures in order to return the relevant images according to their ranking. The system has an accuracy rate of 76.6%. For better performance in the future, hierarchical clustering methods or other indexing methods can be used. Similarity measures like Histogram Of

ISBN -978-93-5406-047-02

Oriented Gradients (HOOG) and other deep learning methods can also be used.

REFERENCES

[1] G. D. K and D. L. Latha, "a Survey on Sketch Based Image Retrieval," *ELK Asia Pacific Journal of Computer Science and Information Systems*, no. January 2015, 2015.

[2] J. M. Saavedra and B. Bustos, "Sketch-based image retrieval using keyshapes," *Multimedia Tools and Applications*, vol. 73, no. 3, pp. 2033–2062, 2014.

[3] A. Chalechale, G. Naghdy, and A. Merlins, "Sketch-based image retrieval using angular partitioning," *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2003*, pp. 668–671, 2003.

[4] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang, "MindFinder: Interactive sketch-based image search on millions of images," *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*, no. July 2015, pp. 1605–1608, 2010.

[5] R. Hu, T. Wang, and J. Collomosse, "A bag-of-regions approach to sketch-based image retrieval," *Proceedings - International Conference on Image Processing, ICIP*, no. September 2011, pp. 3661–3664, 2011.

[6] L. Liu, F. Shen, Y. Shen, X. Liu, and L. Shao, "Deep sketch hashing: Fast free-hand sketch-based image retrieval," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2298–2307, 2017.

[7] S. Deshpande and N. Goyal, "Sketch Based Image Retrieval," pp. 1–4.

[8] S. Thankachan and S. C. Thomas, "MindCam: An Approach for Sketch Based Image Retrieval," *International Journal of Information Systems and Computer Sciences*, vol. 2, 2019.

[9] Y. Qi, Y. Z. Song, H. Zhang, and J. Liu, "Sketch-based image retrieval via Siamese convolutional neural network," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2016-Augus, pp. 2460–2464, 2016.

[10] J. M. Saavedra and B. Bustos, "An improved histogram of edge local orientations for sketch-based image retrieval," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6376 LNCS, pp. 432–441, 2010.

[11] S. Kaushik, "An Introduction to Clustering and different methods of clustering," 2016.

[12] S. Mallick, "Shape Matching using Hu Moments (C++/Python)," 2018.

[13] R. A. Hughes, "Geoscience data and derived spatial information: Societal impacts and benefits, and relevance to geological surveys and agencies," *Special Paper of the Geological Society of America*, vol. 482, pp. 35–40, 2011.

ISBN -978-93-5406-047-02

Fifth National Conference on IoT - VIGYAAN 2020

Tools and Techniques used in IoT - A Review

Mr.Fredy Varghese
Asst.Professor , Department of Computer Science Naipunnya college of Management And Information Technology,Pongam,Thrissur,Research Scholor In VMRFDU Salem
E-mail: fredy@naipunnya.ac.in

Ms.Siji Jose
Asst.Professor , Department of Computer ScienceNaipunnya college of Management And Information Technology
Pongam,Thrissur,
E-mail: siji@naipunnya.ac.in

Dr.Sasikala P
Professor,
Department of Computer Science, VMKVEC, Salem
E-mail: rgsasi@gmail.com

***Abstract*—Internet of Things is a archetype in information technology which provides immeasurable services for the advancement of technological innovations.IoT applications enables seamless consolidation of the cyber-world with the physical world. In this paper we discuss about the various IoT tools and techiniques used for implementing the facilities in our daily life routine specially for smart homes.There are many tools which ease the development of IoT applications.There are many tools available for development.But here we are reviewing only the tools and platforms which are used for creating a smart home.**

Keywords—IoT(Internet of Things), unique identifiers (UIDs).

I. INTRODUCTION

The Internet of Things is penetrating every aspect of our daily life, so the IoT phenomenon is already around us: it is made up of the ordinary objects we use at home, at work or in the streets .[1]The difference is that all these objects and devices are computerized. They have embedded network connectivity, can communicate with phones and other gadgets, get information and remain under control. As the IoT trend is transforming into an industry, the need for reliable and comprehensive developer toolkits is increasing. IoT development tools empower teams with the ability to create applications, access specific networks, test hardware responses to application changes and manage updates.

II. LIRERATURE REVIEW

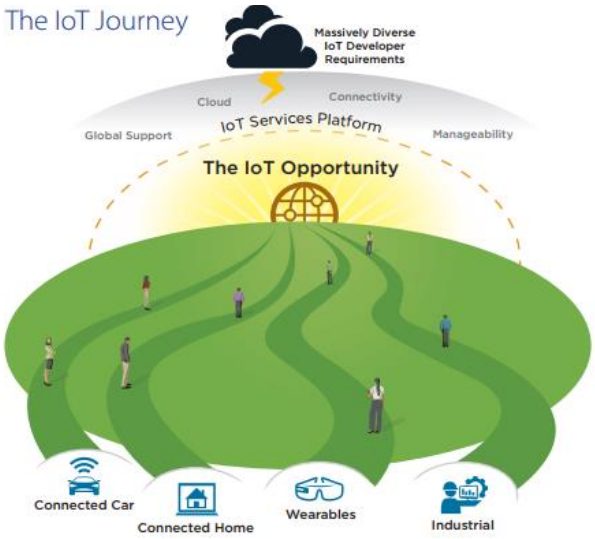
The Internet of Things is a novel paradigm shift in IT arena. The phrase "Internet of Things" which is also shortly well-known as IoT is coined from the two words i.e. the first word is "Internet" and the second word is "Things". The Internet is a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/IP) to serve billions of users worldwide[2]. It is a network of networks that consists of millions of private, public, academic, business, and government networks, of local to global scope, that are linked by a broad array of electronic, wireless and optical networking technologies [3]. Today more than 100 countries are linked into exchanges of data, news and opinions through Internet.

III. BASICS OF IOT

IoT devices can be anything and seen anywhere: door locks, plugs, lights, appliances, vehicles, wearables and other objects[3]. In addition to cyber systems IoT extends its use for industrial facilities as well as domestic use. Major two developments are:

Building Internet of Things (BIoT)

Industrial Internet of Things (IIoT).



The IoT development is based on the integration of hardware and software components so that the final product works according to the principle of the management of both an actuator and an endpoint. The actuator monitors the connected device, searching for a specific value that energizes the endpoint into action. More generally, an IoT system consists of four integral parts:

- Hardware (sensors and devices)
- Software (an application with a user interface)
- Connectivity (Bluetooth, Wi-Fi)
- Data processing algorithms

IV. IOT DEVELOPMENT TOOLS/PLATFORMS

A .*TOOLS*

i. Eclipse IoT

Inorder to build IoT Devices, Gateways, and Cloud Platforms Eclipse IoT technology is used. commonly,

ISBN -978-93-5406-047-02

Eclipse IoT is an environment of companies and individuals collaborating to establish an Internet of Things based on open technologies.[4] This mixture helps to make focal point on the development, promotion and adoption of open source IoT technology.[5] Creating an effective integrated development environment (IDE) for use in programming these applications presents some special challenges because a large number of different tool technologies have to be tightly integrated in support of development task flows. In order to meet these challenges, the Eclipse Platform was designed to serve as the common basis for diverse IDE-based products, providing open APIs (application programming interfaces) to facilitate this integration.

ii. Node-Red

Node-Red is a browser-based editor which makes it very easy to wire together flows using nodes in the palette that can be deployed to its runtime in a single-click[6]. Node-Red provides a built-in library that let you save helpful functions or templates for re-use.

iii . Canopy

Canopy make plainer the IoT cloud by acting as a cloud pass on between IoT devices and applications[5].Canopy makes fast and smooth development of IoT solutions for sundry markets ranging from end users, mercantile and trade .The Cloud service component of Canopy is an open source system that work anywhere, including both public and private cloud, hybrid cloud, LAN, or even on your personal systems**.**

iv . Tessel 2

Tessel 2 is a robotics development platform in IoT . It helps all the libraries of Node.JS to create useful devices easily within minutes using Tessel[6]. Interact with the physical world from sensing to actuation to connecting with other devices Tessel 2 works as a platform. Each module has an open source library on NPM, with instructions and tutorials available online. It's literally plug, npm install, and play.

***B*. IoT Hardware Platforms**

i. Particle.io

Particle.io is an end to end IoT platform enabling different platforms in IoT. It gives a precise, reliable infrastructure to build and control your IoT group[5]. To make the hardware connected with your devices in minutes – over Wi-Fi,Particle's cloud-connected microcontrollers are used.

ii. **Arduino Nano**

Arduino Nano is a experimental model board based on the ATmega328 (Arduino Nano 3.x)[5]. The Arduino Nano can be powered via the Mini-B USB connection.

C . IoT Software Platform

i. PlatformIO

It is an non-proprietary environment for IoT development. For the rapid professional development , C/C++ Intelligent Code Completion and Smart Code Linter are included in PlatformIO[6] . It will support both dark and light colors and multi-projects with Multiple Panes .

ii. prPL

The prpl Foundation create conjoint sections from the clever minds in security,stack and chip design, carrier and mobile communications design, enterprise and storage systems, user applications, and lot more[7].

iii. Tessel 2

Tessel 2 is a robotics development platform which uses all nodes of libraries.JS to create useful devices easily within minutes using Tessel. Tessel boards have its own capabilities to plugging modules[7]. It can access by npm install with few lines of code. It have different combinations of modules to build new systems.

iv. Programmable Wireless

Programmable Wireless provides IoT connectivity with much needed scalability to the organizations, and serves developers around the world.The Twilio Console or via the Twilio API, developers are effortlessly control, analyze, and monitor cellular connectivity. This platform can let add cellular data, voice, and SMS capabilities to your connected devices[5].To manage a large device fleets empowering you to control your IoT devices based on custom requirement REST APIs are used .The developers can build an entirely custom deployment through Programmable Wireless .

ii. Losant

In order to make quickly, easily, and securely build IoT solutions Losant is an IoT platform is used[6].By this tool you can experience enormous connectivity among a wide range of abstract hardware, several devices, store and evaluate the data, and take action in real-time.It is easy to build complex applications on top of the Losant platform.

V Home Automation Software

- **Control Any**

ControlAny offers good IoT software that manage and control operations like Home Automation, Energy Monitoring, Security Automation, and Infrastructure[7]. The main aim is to build smart homes and smart cities.

- **Ninja Sphere**

It is a platform that creates connections to operate all smart devices together. If any problem encountered the Sphere software application will help to fix problem by inform the users when they are away from home[7].Ninja Sphere works with various smart devices available such as WiFi lightbulbs, connected power sockets, Sonos media centers, and more.

- **OpenHAB**

openHAB is a vendor and technology sceptic publicly available autonetics software for home. With pluggable architecture, openHAB supports 200+ different technologies and systems and thousands of devices[8].

ISBN -978-93-5406-047-02

Conclusion

IoT has been gradually bringing a sea of technological changes in our daily lives, which in turn helps to making our life simpler and more comfortable, though various technologies and applications. There is innumerable usefulness of IoT applications into all the domains including medical, manufacturing, industrial, transportation, education, governance, mining, habitat etc. Though IoT has abundant benefits, there are some flaws in the IoT governance and implementation level. The key observations in the literature are that there is no standard definition in worldwide universal standardizations are required in architectural level Technologies are varying from vendor-vendor, so needs to be interoperable for better global governance, we need to build standard protocols. Let us hope future better IoT.

References

[1] Chen, Hao, Xueqin Jia, and Heng Li. "A brief introduction to IoT gateway." *IET International Conference on Communication Technology and Application (ICCTA 2011)*. IET, 2011.

[2] Mileva, Aleksandra, and Boris Panajotov. "Covert channels in TCP/IP protocol stack-extended version." *Open Computer Science* 4.2 (2014): 45-66.

[3] Ruan, Junhu, et al. "A life cycle framework of green IoT-based agriculture and its finance, operation, and management issues." *IEEE communications magazine* 57.3 (2019): 90-96.

[4] Wiegand, J. "Eclipse: A platform for integrating development tools." *IBM Systems Journal* 43.2 (2004): 371-383.

[5] Sodhro, Ali Hassan, et al. "Green media-aware medical IoT system." *Multimedia Tools and Applications* 78.3 (2019): 3045-3064.

[6] Yu, Jaehak, et al. "IoT as a applications: cloud-based building management systems for the internet of things." *Multimedia Tools and Applications* 75.22 (2016): 14583-14596.

[7] Noura, Hassan, et al. "One round cipher algorithm for multimedia IoT devices." *Multimedia tools and applications* 77.14 (2018): 18383-18413.

[8] Meffert, Christopher, et al. "Forensic State Acquisition from Internet of Things (FSAIoT) A general framework and practical approach for IoT forensics through IoT device state acquisition." *Proceedings of the 12th International Conference on Availability, Reliability and Security*. 2017.

ISBN -978-93-5406-047-02

Soft Set Theory-A Novel Soft Computing Tool for Data Mining

Santhosh Kottam,
Research Scholar,
Research and Development Centre,
Bharathiyar University,
Coimbatore,
Email: sankottam@gmail.com.

Dr. Varghese Paul
Professor & HOD
Department of Information Technology
Thrikkakara,Kochi, Pincode: 682022
Email: vp.itcusat@gmail.com

***Abstract*--- Presently various soft computing techniques are applied in data mining. They have certain benefits in their workplace. Regularly, the volume of data repositories is changing. More than that, these data are not always precise and crisp in nature. There exists certain ambiguity and vagueness in the data. While the time of knowledge discovery, our computing tools have to consider these limitations. Due to the lack of parameterization tools, conventional soft computing tools cannot deal with these difficulties effectively. To handle these difficulties we tried to bring out the possibilities of a new soft computing tool- soft set theory- in the data mining. First two sections of this paper cover the introduction and preliminaries of soft computing. In the remaining sections, the authors discuss the theoretical aspects of soft set and multi-soft set. Finally, the paper ends with a discussion on possibilities of soft set theory in different data mining functionalities. Soft sets have a lot of use in the fields of business, health, education, agricultural, and many more.**

Key words: Soft Computing, Soft set, Multi-soft set, Fuzzy set, Rough set, Neural network, Genetic algorithms, Classification, Clustering, Frequent itemset.

I. INTRODUCTION

Powerful techniques and methods for mining useful information from huge data repositories have emerged throughout the recent decades. These techniques utilize the capacity of PC to find large volumes of information in a speedy and successful way. Though, the information to search is vague and badly affected with ambiguity. In the case of non homogeneous data repositories such as movie and plain text, the information might be uncertain and inconsistent. Moreover interesting patterns and relationships are uncertain and not accurate. To make the data mining process more accurate and reliable, there should be effective methods for handling uncertainty, vagueness and errors. These methods should have approximate reasoning power and capacity for handling incomplete data. These methods are collectively known as soft computing.

Soft computing aim is to utilize the patience for partial reasoning, uncertainty, vagueness and partial truth in order to achieve amenability, stability, and inexpensive solutions. Common soft computing methodologies - genetic algorithms, fuzzy sets, rough sets and neural networks - are most broadly used in the overall information discovery and data mining procedures.

We introduce a new soft computing tool –soft set theory- for handling ambiguity and uncertainty in various decision making problems. This concept is introduced by famous Russian mathematician D. Molodtsov. Soft set theory provides sufficient parameterization methods for dealing with data vagueness. In the soft set theory, we do not have to bring in the idea of an exact solution and the initial image of the object is approximate in nature [1]. Also we examine the possibilities of soft set theory in different data mining functionalities like classification and association analysis.

II. PRELIMINARIES

A. Soft Computing

One of the major drawbacks of conventional data mining technique is that complicated problems cannot be precisely explained by mathematical models, and therefore it is very difficult to produce good results from such problems. As mentioned in the introduction part, soft computing handles incomplete facts, ambiguity, and approximate reasoning to solve complicated problems. According to founder of fuzzy logic theory Dr Zadeh [2] stated that "the steering principle of soft computing is to make use of the tolerance for vagueness, uncertainty, and incomplete truth to achieve tractability, robustness, low solution cost, better relationship with reality". Zadeh brings up that soft computing is not a single technique, it consists of a number of techniques like- fuzzy logic, rough set, neural networks, and genetic

ISBN -978-93-5406-047-02

algorithms. All these methods are not standing alone and competitive, but are supportive to each other and can be used collectively to find a solution for a given problem. We can conclude that soft computing objective is to solve complicated problems by using the ambiguity and uncertainty in decision making procedures [3]. A comparative picture of the conventional and soft computing based problem solving method is given in fig.1.

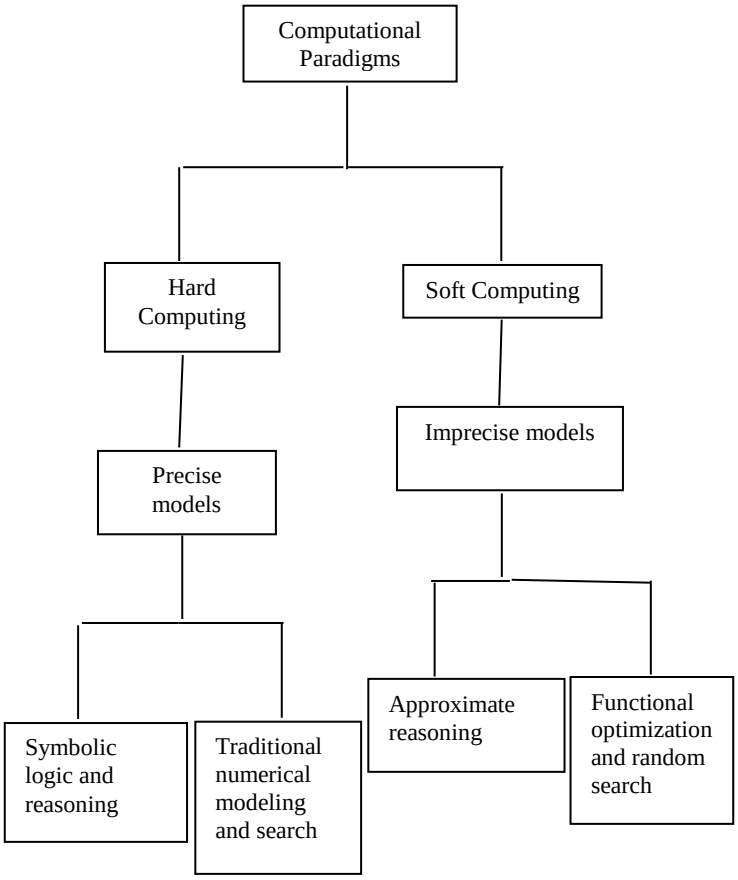


Fig. 1 conventional and soft computing approaches

The left branch represents the conventional hard computing method where a precise model of the problem under analysis is obtainable and conventional mathematical techniques are used to find the solution. The right branch depicts a soft computing approach where only an approximate model of the problem may be obtainable, and the solution depends upon approximate reasoning methods. Soft computing has many applications like- handwriting recognition, automotive systems and manufacturing, image processing and data compression, architecture, decision-support systems and many more. Soft computing is expected to take a good role in science and engineering, but ultimately its power may extend much beyond. In future soft computing performs an important paradigm shift in the style of computing. A modification, which represents the reality that the human mind, dissimilar to current day PC, retains a noticeable capacity to keep and process data which is commonly uncertain, ambiguity and vagueness. Following techniques are commonly used in soft computing- fuzzy set, rough set, genetic algorithms, neural networks and soft set.

B. Fuzzy Sets

This theory has an alternative name known as possibility theory. In 1965, Dr. Lotfi Zadeh invented fuzzy set as an alternative solution for traditional binary-value logic and probability theory. It allows functioning at a maximum level of abstraction and offers a method for handling data, which possess uncertainty and ambiguity. Fuzzy set is different from traditional set theory. It assigns a membership scale for each object and value lies between 0 to1. It has wide application and use in data mining functionalities [4].

C. Neural Networks

A neural network is a set of linked input/output units in which each link has a weight related to it. It consists of a huge collection of extremely interconnected processing objects called neurons, working together to find solutions for precise problems. Neural networks are competent for extracting meaning from complex or vague data, which can be used to find patterns and perceive trends that are too multifaceted. Neural network widely used in following data mining functionalities like- rule generation, rule assessment, regression, and clustering [5].

D. Genetic Algorithms

In knowledge discovery, genetic algorithms are used to evaluate the strength of other algorithms. These are easy to implement and have been used in the following data mining functionalities- association rule analysis, classification, and other optimization problems. Genetic Algorithms can be considered as an evolutionary process with a collection of possible solutions, from which, solutions with higher degrees of fitness are selected. At each level of process, these chosen solutions undergo-crossover and mutation - to produce a candidate of the next generation. Crossover supports in the exchange of generated knowledge in the form of genes between individuals and mutation supports in restoring lost or unexplored regions in search place [6].

E. *Rough Sets*

Rough set theory can be used for classification to find out structural relationships within vague or noisy data. It applies to discrete-valued attributes and therefore, continuous-valued attributes must be discretized before their use. It evaluates a given model from below and from above, using *lower* and *upper* approximations. A rough set learning method can be used to obtain a set of rules in IF-THEN form, from a *decision table*. It provides an effective tool for extracting knowledge from

ISBN -978-93-5406-047-02

databases. In rough set method, firstly creates a knowledge base, classifying objects and attributes within the created decision tables. Then, a data pre-processing step is initiated to remove some noisy data. Finally, the data dependency is assessed, in the reduced database, to find the minimal subset of attributes called reduct [5].

III. SOFT SET THEORY

Next we discuss a newly emerged soft computing tool - soft set theory. Theories like rough set, fuzzy set, and genetic algorithms are good mechanisms for dealing uncertainty and ambiguity. These theories have certain limitations caused by the insufficiency of parameterization. This fact is clearly true in the case of fuzzy set theory. Even though a fuzzy set is a good tool for handling uncertainty, there exists a difficulty like how to determine membership value for each object. There is no precise method for setting a membership value. We should not enforce a uniform method to set the membership value. The character of the membership value is particularly individual. For example, consider the statement $\mu_F(x) = 0.7$, everyone understands this statement in his own manner. So, the fuzzy set does not give a realistic approach for fixing a membership function. The basis for these limitations is, due to the insufficiency of the parameterization. Considering these facts, famous Russion Mathematician, Molodtsov contributed an idea of soft theory as a method for handling with uncertainties which is except from the above limitations. Soft set theory has wide applications in different domains.

A. *Definition (Soft Set)*

A pair (G, P) is called a soft set over V, where G is a mapping given by

$$G : P \to F(V)$$

In other words, a soft set over V is a parameterized family of subsets of the universe V. Every set $G(\varepsilon), \varepsilon \in P$ may be considered as the set of ε-elements of the soft set (G, P). A soft set is not a set.

Consider an example

V is the set of leaders considering an event.
P is the set of qualities of a good leader. Each parameter is a word or a sentence.
P= {Vision, Courage, Integrity, Humility, Strategic planning, Focus, Cooperation}
Here soft set (G, P) describes the "qualities of the leader" which an organization is going to measure. Suppose that there are six leaders in the batch V given by

V={l1,l2,l3,l4,l5,l6} and P={p_1,p_2,p_3,p_4,p_5,p_6,p_7}

where
p1 stands for the parameter 'Vision'
p_2 stands for the parameter 'Courage'
p_3 stands for the parameter 'Integrity'
p_4 stands for the parameter 'Humility'
p_5 stands for the parameter 'Strategic planning'
p_6 stands for the parameter 'Focus'
p_7 stands for the parameter 'Cooperation'

Suppose that

G (p_1) = {l2,l4}
G(p_2) = {l1, l3}
G(p_3) = {l3,l4, l5}
G(p_4) = {l1, l3, l5}
G(p_5) = {l1}
G(p_6) = {l2, l3, l4}
G(p_7) = {l2, l4, l6}

The soft set (G, P) is a parameterized family {G (p_i), i=1, 2, 3……, 7} of subsets of the set V and gives us a collection of approximate descriptions of an object. G (p_1) means "leaders (vision)" whose functional value is the set {l2, l4}. The soft set (G, P) is a collection of approximation as below:

(G, P) = {vision = {l2, l4}, courage= {l1, l3}, Integrity = {l3, l4, l5}, Humility = {l1, l3, l5}, Strategic planning= {l1}, Focus= {l2, l3, l4}, Cooperation= {l2, l4, l6}}

Where each approximation has two parts
(i) a predicate *p* eg: vision
(ii) an approximate value set *v* eg: {l2, l4}

TABLE I
TABULAR REPRESENTATION OF A SOFT SET

U	visio n	cou rage	Integr ity	Hu mili ty	Strate gic plann ing	Focus	Coop erati on
l1	0	1	0	1	1	0	0
l2	1	0	0	0	0	1	1
l3	0	1	1	1	0	1	0
l4	1	0	1	0	0	1	1
l5	0	0	1	1	0	0	0
l6	0	0	0	0	0	0	1

B. Definition (*Operations with Soft Set)*

Consider a binary operator *, for subsets of the set V. Let (G, A) and (H, B) be soft sets over V. Then, the operation * for soft sets is defined as

$$(G, A) * (H, B) = (I, A \times B),$$

Where $I(\alpha, \beta) = G(\alpha) * H(\beta), \alpha \in A, \beta \in B,$ and $A \times B$ is the Cartesian product of the sets A and B.

ISBN -978-93-5406-047-02

C. Definition (Equality of Two Soft Sets)

Two soft sets (G, A) and (H, B) over a common universe V are said to be soft equal if (G, A) is a soft subset of (H, B) and (H, B) is a soft subset of (G, A).

D. *Definition* (*Not Set of a Set of Parameters)*

Let P = {p_1,pe_2,pe_3, ...pe_n} be a set of parameters. The Not set P denoted by ⅂ P is defined by ⅂ P = {⅂p_1, ⅂p_2, ⅂p_3, ……., ⅂p_n} where ⅂p_i = not p_i , $\forall i$

E. Definition (*Complement Soft Set of a Soft Set*)

The complement of' a soft set (G: A) is denoted by (G, A)^c and is defined by (G, A)^c = (G^c, ⅂A) where G^c: ⅂A $\rightarrow$ F(U) is a mapping given by G^c (α) = V – G (⅂α), $\forall \alpha \in$ ⅂A.

F. Definition 4.6 (Null Soft Set)

A soft set (*G*, *A*) over *V* is said to be a NULL soft set denoted by ϕ , if $\forall \varepsilon \in A, G(\varepsilon) = \phi,$ *(null-set).*

G. Definition (Absolute Soft Set)

A soft set (G, A) over V is said to be absolute soft set denoted by $\widetilde{A}, if \forall \varepsilon \in A, G(\varepsilon) = V$

H. Definition (And Operation on Two Soft Sets)

If (G, *A*) and (H, *B*) be two soft sets then (G, *A*) AND (H, *B*) denoted by (G, *A*) ∧ (H, *B*) and is defined by (G, *A*)(H, *B*) = (I, *A* × *B*) where I(*α*, *β*) = G(*α*) ∩ H(*β*) for all (*α*, *β*) ∈ *A* × *B*.

I. Definition (Or Operation on Two Soft Sets)

If (G, *A*) and (H, *B*) be two soft sets then (G, *A*) OR (H, *B*) denoted by (G, *A*) ∨ (H, *B*) is defined by (G, *A*) ∨ (H, *B*) = (I, *A* × *B*) where I(*α*, *β*) = G(*α*) ∪ H(*β*) for all (*α*, *β*) ∈ *A* × *B*.

IV. MULTI-SOFT SET THEORY

A. Decomposition of Multi-valued Information Systems

We decompose a multi-valued information system T=(V, B, W, g) into |*B*| number of binary-valued information systems. The decomposition of T=(V, B, W, g) is based on decomposition of B={b_1,b_2,b_3….,$b_{|B|}$} into the disjoint-singleton attribute {b_1 },{b_2 },… {$b_{|B|}$ } . Let T = (V, B,*W*,g) be an information system such that for every b∈B , W_b = *g(V,B)*, is a finite non-empty set and for every *v* ∈*V* , |*f* (*v*, b)| =1. For every b_i under i^{th} -attribute consideration, b_i ∈B and w∈W_b , we define the map b^i_w :*V*→{0,1} such that b^i_w (*v*) = 1 if *g*(v, b) = w , otherwise b^i_w (v) = 0. The next result, we define a binary-valued information system as a quadruple T*'=(V, b_i, $W_{\{0,1\}}$, g).*The information systems T*'=(V, b_i, $W_{\{0,1\}}$,g), i=1,2,....|B|* is referred to as a decomposition of multi-valued information system *T* = (V, B,W, g) into |B| binary-valued information systems. Every information system T*'=(V,b_i, W_{bi},g), i=1,2,....,|B|* is a deterministic information system since for every b∈ B and for every v ∈V ,|*g* (*v*,b)| = 1 such that the structure of a multi-valued information system and |B| number of binary-valued information systems give the same value of attribute related to objects[7].

Proposition 1. *If* (H, P) *is a soft set over the universe S, then* (H, P) *is a binary valued information system R*= (S, *A*,$V_{\{0,1\}}$, *h*).

Proof. Let (H, P) be a soft set over the universe *S*, we define a mapping

H ={h1 , h2 ,….. , hn} , where

$$h_{i:}S \rightarrow V_i \text{ and } h_i(x) = \begin{cases} 1, & x \in H(yi) \\ 0, & x \notin H(yi), \end{cases} \text{ for } 1 \le i \le n$$

Hence, if A=P, V= $U_{ei} \in A$ $^{V}_{ei}$ where $^{V}_{ei}$ = {0,1}, then a soft set (H, P) can be considered s a binary-valued information system R= (S, *A*,$V_{\{0,1\}}$, *h*).
Consider a multi valued information system R= (S, A, *V*, *h*) and R^i = (S, a_i, V_{ai}, *h*), ι=1,2,....|A| be the |A| b*inary*-valued information systems.

From *proposition 1*.

$$Q = (S, A, V, h) = \begin{cases} R^1 = (S, a1, V_{\{0,1\}}, h) \iff (H, a_1) \\ R^2 = (S, a2, V_{\{0,1\}}, h) \iff (H, a_2) \\ \quad . \qquad\qquad . \\ \quad . \qquad\qquad . \\ R^{|A|} = (S, a_{|A|}, V_{\{0,1\}}, h) \iff (H, a_{|B|}) \end{cases}$$

$$= ((H,a_1),(H,a_2),\ldots..,(H,a_{|A|}))$$

We can conclude (H,P) = $((H,a_1),(H,a_2),\ldots..,(H,a_{|A|}))$ as a multi soft set over universe S representing a multi-valued information system R = (*S*, *A*,*V*, *h*).

V. APPLICATION OF SOFT SET THEORY IN DATA MINING

A. Decision Making

All kinds of decision making problems consist of following processes- formal modelling, reasoning and

ISBN -978-93-5406-047-02

computing. Traditionally, these methods have been deterministic, crisp, and exact in character. In conventional mathematics, we build a mathematical replica of an item and define the concept of the precise solution of this replica. There are many complex problems existing in the real world that involve data which are not always all crisp. Usually the mathematical replica is too complex and we cannot find the precise solution. Nowadays, we apply different soft computing methods to overcome different challenges posed by data mining. The main techniques of soft computing include - fuzzy set, rough sets, neural networks, and genetic algorithms. These methods face challenges from uncertain data handling. Soft set theory contributes sufficient parameterization tools for handling data uncertainty. Here, the initial picture of the object has an approximate nature, and we do not need to set up the idea of precise solution.

B. Classification

Classification is one of the important data mining functions and used in many decision making problems. Simply, it is a process of finding the class label of an unknown object. Classification is a two-step process. In the first phase, we build up a model from the training data set and in the second step; the derived model is used for finding the class label of unknown object. In the first phase, we use efficient and scalable algorithms.

Generate a multi soft set from the given data set. Apply operations and properties of multi-soft set on generated soft set until all records are classified.

C. Frequent itemset mining

Association analysis is a widely used functionality in knowledge discovery. Nowadays, it has a good number of applications in academia, industry, and research scholars. For making decisions in all areas, an interesting pattern among the items is very important. Usually, we find out frequent item sets from huge data sets. Different algorithms and methods exist for finding frequent item sets. Among them, Apriori and FP growth algorithms are commonly used. Apriori algorithm is well known for its simplicity. Since it is an iterative method, it will consume a huge amount of time to finish its execution. Similarly, FP-Growth algorithm emerged for eliminating the drawbacks of apriori algorithm. The generation of FP-growth algorithm requires a high volume of memory. Scanning the dataset twice reduces the full efficiency of FP- growth algorithm and produces a large number of FP-trees. Two overcome these limitations soft set theory can play a vital role.

D. Clustering

Clustering is one of the most prominent methods in data mining. Clustering is a process for identifying class and finding interesting correlations and patterns in the underlying data, which is needed in a number of data analysis tasks like - unsupervised classification and data summation, segmentation of large homogeneous data sets into smaller homogeneous, that can be separately modeled and analyzed. Rough set theory, introduced by Z. Pawlak in 1982, is a mathematical tool to deal with vagueness and uncertainty. One of the popular techniques in data clustering is based on rough set theory. The main approach of rough set-based data clustering is the clustering repository which is mapped as the decision table and this method can be executed by introducing a clustering attribute. Hence, from a database, we select only one attribute, which is the best way to partition the elements, is of primary importance for this approach. Currently, there have been researches in the area of implementing rough set theory in the selection process of clustering attribute. Relationship between the rough set and soft set theories are interesting. Combination of these two theories is used for determining a clustering attribute from a given set of attributes [8].

CONCLUSION AND FUTURE WORK

Soft computing techniques have a good role in data mining. Many research works are going in this direction and produced various techniques which have dependable accuracy and computation performance. In data mining following soft computing techniques are frequently used for mining knowledge huge repositories – Fuzzy set, rough set, and soft set. Soft set theory is eliminating deficiencies of the above theories. Except for soft set theory, others face the problem of data uncertainty. Reason for these problems is due to the lack of parameterization tools. Soft set theory has sufficient parameterization techniques. This property automatically increased the popularity of soft sets. It has a good number of applications in data mining. We studied these applications and presented the possibilities of soft set theory in data mining.

In our research work, firstly, we presented soft computing and it's important in computation modelling. Our study went through following soft computing methods-fuzzy set, genetic algorithms, neural network and rough set. We tried to bring out limitations of these techniques. Next, we presented theoretical aspects of soft set and one of its branch multi-soft set theory. We discussed different definitions, propositions and properties of these theories. At the end, our paper concluded with a discussion on possibilities of soft set theory in data mining.

In the future, we would like to extend our study to find more opportunities for soft set theory in data mining.

ISBN -978-93-5406-047-02

REFERENCES

[1] D. Molodtsov," Soft set theory- First results", An International Journal of Computers & mathematics with applications", Elsevier, Volume 37, Issue 4-5, Page 19-31, February-March1999.

[2] Zadeh LA. Fuzzy logic, neural networks and soft computing. One-page course announcement of CS 294-4. Spring 1993. University of California at Berkeley; Nov. 1992.

[3] Zadeh LA. Fuzzy logic, neural networks, and soft computing. Communications of the ACM 1994; vol. 37. no. 3. pp. 77-84.

[4] L A Zadeh, Fuzzy sets, Inor. and Control 8, 338-353, 1965.

[5] Sushmita Mitra, Senior Member, IEEE, Sankar K. Pal, Fellow, *IEEE*, and Pabitra Mitra, Data Mining in Soft Computing Framework: A Survey, *IEEE* transactions on neural networks, vol. 13, no. 1, january 2002.

[6] Jiawei Han, Micheline Kamber, Data Mining: Concept and Techniques, Elsevier, second edition, 2006.

[7] Tutut Herawan, Mustafa Mat Deris "On Multi-soft Sets Construction in Information Systems", 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009, Proceedings (pp.101-110).

[8] Hongwu Qin, Xiuqin Ma, Jasni Mohamad Zain, Tutut Herawan "A novel soft set approach in selecting clustering attribute", Knowledge-Based Systems, 36 (2012) 139–145, Elsevier.

AUTHORS PROFILE

Mr. Santhosh Kottam completed his Master of Computer Applications (MCA) from Madras University, Tamilnadu and B.Sc mathematics degree from M G University, Kerala. He is currently pursuing PhD in computer science at Bharatiyar University, Coimbatore. His research area is data mining and has more than 19 years of teaching experience, which includes UG and PG. He has been serving Federal Institute of Science and Technology (FISAT), Angamaly, Kerala as HOD & Assistant Professor (Senior Grade) in the Department of Computer Applications since May 2008. During this period of time, he has taught many subjects including Programming Languages, System Analysis and Design, Operating Systems, Object Oriented Modeling and design and Data Mining. He has published research papers in the International Journals, National and International Conferences. Mr. Santhosh received the best teacher award of the FISAT College during the year 2017 for his overall contribution.

Dr. Varghese Paul is completed B.Sc (Engg) in Electrical Engineering from Kerala University, M.Tech in Electronics and Ph.D in Computer Science from Cochin University of Science and Technology. Research Supervisor of Cochin University of Science and Technology, M G University Kottayam, Anna Technical University Chennai, Bharathiar University Coimbatore, Bharathidasan University Trichy and Karpagam University Coimbatore. Under the guidance, 29 research scholars had already completed research studies and degree awarded. Research areas are Data Security using Cryptography, Data Compression, Data Mining, Image Processing and E_Governance. Developed TDMRC Coding System for character representation in computer systems and encryption system using this unique coding system. Published many research papers in international as well as national journals and a text book also. Earlier worked as Industrial Engineer with O/E/N India Ltd Cochin, Communication Engineer with KSE Board, SCADA Engineer in Saudi Electricity Department, Head of IT Department CUSAT and Dean (CS, IT and Research) in Toc H Institute of Science and Technology. Certified Software Test Manager, Ministry of Information Technology, Govt of India. Life Member, Information System Audit and Control Association, USA (ISACA), Indian Society for Technical Education, India (ISTE) amd National Geographic Society, USA.

ISBN -978-93-5406-047-02

Fifth National Conference on IoT - VIGYAAN 2020

A Study on the Influence of E commerce Website Quality on Customer Satisfaction among Working Professionals in Kerala

Sarithadevi S
Asst. Professor,
Department of Computer Science, NIMIT
sarithadevi@naipunnya.ac.in

Abstract— **Due to rapid development in the field of technology, the use of e-commerce has expanded drastically in recent years. In an e-commerce company, website quality is a very significant determinant of customer satisfaction. This study is based on a survey done among working professionals in Kerala on their perception of website quality with respect to e-commerce websites like Flipkart, Amazon and myntra. This study is to understand the relation between different factors affecting website quality with the satisfaction of customers. The focus is on online ecommerce websites. This study is based on a quantitative survey of website quality with respect to online ecommerce websites. Correlations and regression with SPSS Ver. 16 was used to analyze the variables and their relationships. This study conceptualizes the relation between different factors affecting website quality with the satisfaction of customers. The strongest determinants of website quality are information quality, system quality, service quality, and website design.**

Keywords- Customer Satisfaction, Website Quality, Purchase Intension.

I INTRODUCTION

In the last few years, E-commerce has gained popularity. India ranks second in the world based on population, this implies a large consumer base. E-commerce companies offer immense growth potential because they have larger geography than traditional retailers. In the past decade, internet users have increased drastically and the data tariff has come down significantly making access to E-Commerce websites easy and less expensive. Now ecommerce websites are available in smartphone applications, and thus, can be accessed anywhere using a cell phone. E-Commerce websites are very easy to use and offer better options in terms of product choice and customer service. There is no face- to- face interaction between the customer and the company and the first interaction is via the website homepage. In an online booking website, the relevant information should be well-organized since the customers visit the website to find relevant information about a particular product and then choose from the options available.

II.LITERATURE REVIEW

Jarvenpaa and Todd (1997) stressed that service quality was most important in E-commerce websites. According to a Boston consulting group survey (2012), 41% of online shoppers stopped using online shopping website when they experienced failure of transaction. This study further pointed out that disappointed customers spent less money in online purchases. Customer satisfaction effected the money spent on the website and if a customer was satisfied with the overall website quality then he may buy value-added services. Security and personalization are very important for online shopping. Kim and Lim (2011), found that system quality and information quality as most important in user satisfaction. Delone and Mclean (2012) proposed a model to measure the quality of information system by referring the work done by scholars in the 2012s. Later, Delone and Mclean (2013) proposed an updated model for measuring information system success. Information quality is a key factor in the success of ecommerce websites. Relevant and easy-to-understand information significantly influences customer satisfaction. The information in the ecommerce website should be easily understandable and briefly explained. Understandability means ease of understanding and clarity of information includes frequent update of information. The pricing is also very dynamic and changes frequently. So regular update of information is needed to fulfil the customers' demands to get the best deals possible.

Information quality can be measured using information relevance and completeness of information. Initially, a customer visits the website as an information seeker, and then, if he finds the relevant product he purchases it. Lee and Lin (2015) found information quality as the most important factor affecting the buying behaviour of a customer.

ISBN -978-93-5406-047-02

System quality is the performance of a system in delivering information. E commerce websites should have an easy to understand system with a minimum number of parameters from which a customer can access the relevant information. Website system quality also refers to the efficient use of technology. But personalization by collecting sensitive information like saving debit card numbers will have a negative impact on customer satisfaction. Customers in online context are dissatisfied with poor navigation and unsecured payment gateway. In the modern world, customers are also dissatisfied by lack of personalization. Customers will stop using the website if the system quality is not up to their expectation even though the information quality is high. It is convenient for a customer to compare the products online than through traditional means. The results can be personalized by using various options in an ecommerce website for customers to view the required results. Service quality is very important for a ticket booking website in terms of customer retention. If the service quality is high then it is more likely that customers will return to the same website for future purchase. Service quality will also help customers in trusting the ecommerce brand and will subsequently help in increasing its reputation because consumers in an online context interact with unseen retailers. Also, an ecommerce website must be reliable. Reliability means keeping the promised service and responsiveness. 24x7 customer service, FAQs, and complaint management system are very essential in affecting purchase intention of customers in ecommerce websites. Website design is also important in accessing website quality. It is likely that customers will evaluate online store experience as an overall process, rather than on individual sub- parameters. Website design is the first impression when a customer logs into the system. Website design describes the appeal that a user interface design presents to the customers. Customer satisfaction is an emotional status of an individual. Bhattacherjee (2017) found that satisfaction is a major driver for continued purchase intention. Customer satisfaction positively affects purchase intention. A satisfied customer is more likely to revisit the website than a partially dissatisfied or dissatisfied customer. Customers who are satisfied will pose positive attitude towards the website. For an online ecommerce website business to be successful there should be satisfied customers spreading positive word of mouth. Online businesses are highly competitive and customers have high expectation in terms of service quality.

Hsu, Chang and Chen (2011) in "The impact of website quality on customer satisfaction and purchase intention: perceived playfulness and perceived flow as mediator" confirmed that website quality affected customer's perceived playfulness and perceived flow. Notably, this study found that service quality was more important than information and system quality in influencing customer satisfaction.

Sun et al. (2015) in "Consumption system model integrating quality, satisfaction and behavioral intentions in online shopping" developed and tested a consumption system model for online shopping that incorporated both product and e-service elements and online and offline stages of transaction process. The results showed that perceived e-service quality significantly affected customer satisfaction. However, only customer satisfaction had direct effect on behavioural intention and offline perceived product quality. The survey respondents were college students.

Pauline de Pechpeyrou (2018) conducted a survey on personalized selling online. It was found that personalized selling increased positive attitude on the website, thus getting more clicks on products displayed.

Lee and Kozar (2016) used AHP method to investigate the factors effecting website selection. Delone and Mcleon's model for IS success was used and websites selling commodity goods only were considered for the study. The findings of the study suggested that online shopping websites must provide more aesthetic and convenient shopping experience. Information relevance was the most important factor in website selection.

Kuan et al. (2018) found that website quality was directly related to initial and continued purchase intention. System quality was very important to convert a website surfer to a customer. However, the research showed that system quality had limited impact on purchase intention once the customer had significant trust in the website. E-Commerce websites should not just provide information, but also convert information seekers to online shoppers. As the shoppers repurchase from the same website, system quality is very important. Similarly, for a customer purchasing an ecommerce product, information quality is very significant.

Taylor et al. (2018) observed that personalizing online interactions improved customer relationship and increased purchase intention. This study also investigated the negative effect of privacy concern with behavioural intentions. Increase in perceived information control reduced the negative effect on privacy and engaged in positive intention. However, compensation increased trust.

By the introduction of e commerce websites, the people began to choose the online purchase instead of traditional system. "A study on the Influence of ecommerce website quality on Customer satisfaction among working professionals in Kerala" is trying to find out

1) What are the factors influencing customer satisfaction on e-commerce purchase

2) What is the customer's perception specially working professionals about website quality?

Information quality, System Design etc.

ISBN -978-93-5406-047-02

III OBJECTIVES

1. To study the factors influencing website quality of ecommerce websites
2. To understand if Service quality factor have an influence on Customer satisfaction of e commerce websites
3. To understand if information quality factor have an influence on Customer satisfaction of e commerce websites
4. To understand if website design factor have an influence on Customer satisfaction of e commerce websites
5. To understand if system design factor have an influence on Customer satisfaction of e commerce websites
6. To understand customer satisfaction in e commerce website have an influence on purchase intention.

IV HYPOTHESIS

H_{O1}: There is no significant relationship between website quality and Customer Satisfaction

H_{O2}: There is no significant relationship between service quality and Customer Satisfaction

H_{O3}: There is no significant relationship between Information quality and Customer Satisfaction

H_{O4}: There is no significant relationship between Website Design and Customer Satisfaction

H_{O5}: There is no significant relationship between System Design and Customer Satisfaction

H_{O6}: There no significant relationship between Customer Satisfaction and Purchase Intention

V SIGNIFICANCE OF STUDY

In an attempt to satisfy the basic need of the customers online e-commerce websites comes to play a crucial role. In recent years the purchase behaviour of customers has under gone major changes to a large extent the changes have led to increase competition particularly between the domestic and foreign e-commerce websites. Moreover the e commerce business is developed with severe competition, fluctuations and new challenges. The online e-commerce websites are facing with new competitors which are Flipcart, Amazon, and Myntra etc.

Now a day's customer become wiser and they require not only high quality but also a better and more professional service. The website company which can provide better quality service to its customers can sustain the competition in the market. In this point of view 'A study on the Influence of ecommerce website quality on customer satisfaction among working professionals in Kerala' is highly significant in the current scenario.

VI SCOPE OF STUDY

The basic idea of this study is to find out the influence of website quality on customer satisfaction. The customers are changing; their preference, attitude and demand are changing as per the change in information technology

The geographical scope of the study is among working professionals in Kerala state from which primary data is collected. The study covers the customer satisfaction, webiste quality and purchase intention through online e commerce websites.

VI METHODOLOGY

Primary data

Primary data was collected by online questionnaire using Google Form from the working professionals in Kerala who use e-commerce websites for their purchase.

Secondary data

The secondary data is collected from published articles and from internet

Sampling technique

Convenient sampling technique

Sample size

For the survey a sample size of 70 respondents from Kerala State was taken into consideration

Tools for analysis

The collected primary data were statistically processed, classified, tabulated and analyzed by using statistical and mathematical tools and techniques like Percentages, Mean, and Regression Analysis etc. Fig 6.1 shows conceptual model of website quality

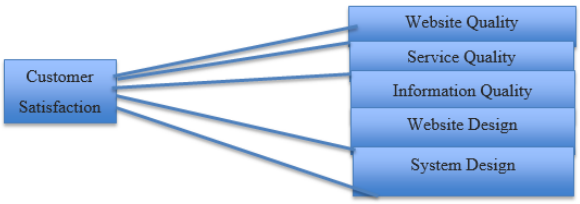


Fig 6.1 conceptual model of website quality

VII STATISTICAL ANALYSIS

Statistical Package for the Social Sciences (SPSS) was applied to examine the relationship and to assess the objectives of current research. For the purpose of data survey, descriptive and inferential statistics were employed. This study has employed descriptive statistics to review the biographical responses, to explain the website quality dimensions and describe the customer satisfaction. The Pearson correlation applied to define the degree of relationship between the variables in this study. I.e. Service quality, system quality and information quality.

Interpretation of Results

Demographic Profile of Respondents

ISBN -978-93-5406-047-02

The table presents the Gender wise composition of respondents

Table 1: Gender of Respondents

Gender	Frequency	Percentage
Male	40	57.14
Female	30	42.85
Total	70	100.0

Source: Primary Data

Out of the 70 respondents, 57.14 percent are males and 42.85% percent are female.

Regression Analysis and Hypotheses Testing

Regression analysis was conducted to measure the influence of WQ, SQ, IQ, WD and SD on CS.The independent variables are WQ, SQ, IQ, WD and SD and the dependent variable is CS. The main objective of regression analysis is to explain the variation in one variable (called the dependent variable) based on the variation in one or more other variables (called independent variables)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.626[a]	.392	.345	2.04053

a. Predictors: (Constant), SD, WQ, WD, SQ, IQ

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	171.863	5	34.373	8.255	.000[a]
	Residual	266.480	64	4.164		
	Total	438.343	69			

a. Predictors: (Constant), SD, WQ, WD, SQ, IQ

b. Dependent Variable: CS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.205	2.552		.472	.638
	WQ	.041	.114	.044	.358	.722
	SQ	.031	.108	.037	.282	.779
	IQ	.116	.152	.102	.767	.446
	WD	.411	.143	.379	2.865	.006
	SD	.300	.136	.240	2.200	.031

a. Dependent Variable: CS

Table : Cronbach's Co-efficient

Factors(Constructs)	Number of Items	Cronbach's Alpha
Website Design	4	.708
System Design	4	.774
Customer Satisfaction	4	.762
Website Quality	4	.704
System Quality	4	.884
Information Quality	4	.728

Source: Author's Calculation

VII FINDINGS

Tables represents the beta values, t values and significant values of independent variable WQ, SQ, IQ, WD, SD. The independent variables WQ (t = 0.044,p=.638), SQ (t = 0.037,p=0.772), IQ(t =0.102,p=0.779), WD (t=0.379,p=0.006), SD (t=0.240,p=0.031), are statistically significant at 1 percent significance level. It means that these five independent variables have significant positive effect on CS.

Hence **H01, H02, H03, H04, H05, H06 and H07 are rejected.** The beta coefficients give a measure of the contribution of each variable to the model. Higher the beta value, the greater is the effect of independent variable on the dependent variable. Among the independent variable WD has greater effect followed by SD, IQ, WQ, and SQ. So it can be concluded that, Website Design, System Design, Information Quality and Website Quality and System quality have significant influence on Customer Satisfaction.

XI LIMITATIONS OF THE STUDY

Despite the fact that very reliable results that may also be generalized have been arrived at, the researcher would like to point out some unavoidable limitations that have entered into the study. They are stated below

1. The study is limited to working professionals in Kerala; hence it's not applicable to all other places.
2. The findings of the study are based on the responses of the respondents which might have their own limitations. The attempted objectivity has naturally been constrained by the extent of respondent's readiness to give factual information although all possible efforts have been made to collect authentic information.
3. The time constraints of some of the respondents forced them to give casual responses

ISBN -978-93-5406-047-02

4. Another limitation was the time factor due to which the sample size had to restrict to 70 respondents.

The study is limited by the knowledge and experience of the researcher of the project

X CONCLUSION

This study is conducted to examine the influence of website quality of e commerce websites on customer satisfaction among working professionals in Kerala. On the basis of study, the conclusion drawn are the website design have high influence over customer satisfaction. The appearance of website in top position in search results is a convenience and indicator of quality. The navigability and organization of the website influence the purchase decision. The study identified that e commerce platforms provide delivery services to remote areas.All the people agrees that availability of a toll free customer care number is a basic service that must be provided to the customers.Provision of order tracking and returns/refunds adds a great value to e commerce. Most of them agrees with the relevance of information, user reviews influences the purchase decision. Difficulties in loading different elements affects purchase intention.As a customer, all are very much concerned about the security of the information they provide to the website.System design highly agrees that multiple payment mechanisms is needed.It is identified that an ideal an e commerce seller takes user feedback highly into consideration.All agreed on the fact that online sellers should establish a relationship a with regular visitors to make them loyal customers. The Regression results revealed that customer satisfaction is highly influenced by website design.Most of the customers regularly visits the websites Flipcart and Amazon for their online purchase.

REFERENCES

[1] Raja Raman, V. Essentials of e-commerce technology. PHI Learning Pvt. Ltd., 2009.
[2] Whiteley, E-Commerce: Strategy, Technologies and Applications, McGraw-Hill Education (India) Pvt Limited, 2001
[3] https://en.wikipedia.org/wiki/E-commerce
[4] https://www.toppr.com/guides/business-environment/emerging-trends-in-business/electronic-commerce/
[5] https://www.toppr.com/guides/business-environment/emerging-trends-in-business/electronic-commerce/

ISBN -978-93-5406-047-02

Fifth National Conference on IoT - VIGYAAN 2020

Internet of Things (IoT): Applications, Benefits, Challenges and Implementations in Banking Domain

Binju Saju[#] , Jayakrishnan S[*]
[#]Asst.Professor , Computer Science Department , NIMIT ,Pongam
[#]binju@naipunnya.ac.in
[*]Associate Professor and HOD, Computer Science Department , NIMIT ,Pongam
[*]hod-cs@naipunnya.ac.in

Abstract- **The influence of Internet of Things (IoT) and its closely related counterpart Internet of Everything (IoE) is so exponential that it has become an active part of our digital lives. Consumers become the greatest beneficiaries as IoT creates the next wave of digital revolution. Connected objects form a dimension beyond the electronic world in which Internet operates and extends to things and places. The purpose of this article is to explore the applications of IoT in finance and scrutinizes the impact brought about by digital trends and the new trends in the banking sector. In this work, challenges and implementations in banking domain is dealt in detail**

***Index Terms*—IoT, IoE, banking, challenges, Finance**

I. INTRODUCTION

The Internet of Things (IoT) envisions a fully interconnected world [1].The Internet of things (IoT) is the network of interrelated computing devices that connect and exchange data with one another via the Internet. The new and evolving behaviours and uses of customers, as well as the mounting corpus of data, demands inevitable digital transformations for the Bank's stakeholders [2]. The IoT vision enhances connectivity from "any-time, any-place" for "any-one" into "any-time, any-place" for "any-thing"[3]. ITU-GSI[23] has defined IoT as "the network of physical objects-devices, vehicles, buildings and other itemsembedded with electronics, software, sensors, and network connectivity that enables these objects to collect and exchange data".

Banks are actively investing in IoT technology nowadays. Financial institutions have an average IoT banking budget of $117.4 million, which comes up to about 0.4% of the revenue. Banks have always been great at identifying and adopting new technologies. They have realized the potential of IoT banking in providing unimaginable levels of data and customer insights. IoT banking can extend suggestions and latest offers on a regular basis based on the transaction history of the customers, thus effectively personalizing banking for individuals.

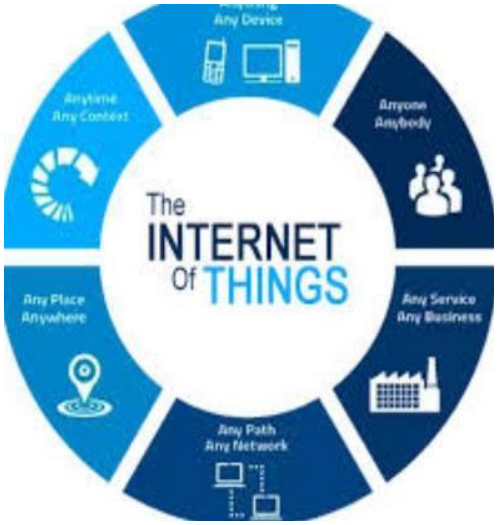


Fig [1] IoT

The purpose of this article is to present the different uses of IoT in finance and allied domains. It also analyses the impact of digital trends and IoT on the procedural scheme of a traditional bank. Currently, different sectors of the economy are making make use of various forms of technologies to simplify the job at hand, automate actions and automate repetitive processes. One such industry is the banking and financial sector where stakeholders are continuously looking for new ways of improving the quality of services and in turn the overall customer experience. Banks and financial institutions can now make use of IoT technologies to reduce fraud transactions, detection of risk, stay ahead of their competitors and provide better services to their customers. IoT plays a huge role in the financial sector - it has the

ISBN -978-93-5406-047-02

potential to changing the financial and banking sectors. In fact, research has shown that there will be high growth in the adoption of IoT in the financial industry with its market size expected to be $25 billion by next year. Banks generally deal with the gathering of data and massive data transfer. IoT will help banks and financial institutions save time in completing these tasks and moves towards a smarter work culture. The influence of IoT has increased through the introduction of mobile banking where customers can carry out various forms of transactions without physically going to the banks and self-assisted customer care services where issues are attended by a virtual assistant. IOT thus helps financial institutions improve customer experience, manage risk, secure funds from cyber-attacks, and strengthen the entire banking security system.

Fig [2] IoT in Banking

By 2020, it is predicted that there will be 26 billion connected devices globally- a 30-fold increase from 2009. [2] This exponential increase in the number of IoT devices will produce an incalculable rise in the number of interactions, transactions and exchanges, offering a virtually endless number of possibilities and opportunities, which we may not be able to measure or control. However, within this limitless exchange of information and data lies comparably limitless challenges and vulnerabilities.

To understand application of an IOT system in a better manner, let us consider this scenario of smart airport- an IoT-based airport management system in fig [3]. This is an airport management system based on the IoT paradigm, where passengers, baggage, plane or the departure lounge are considered as "things". This smart airport management system aims to automate passengers processing and flight management steps, in order to improve services, ease the work of airport agents and ensure a pleasant and safe journey for the passengers [4].

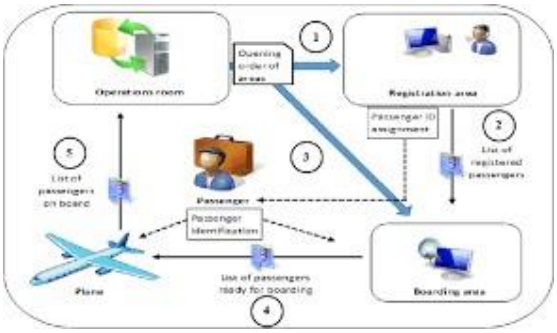


Fig [3] Iot based airport management system

In this research paper, we will be discussing mainly the applications, benefits and future challenges of internet of things (IoT) based on the work done by different researchers in the field... The main aim of this paper is to provide an overall view of the applications, benefits and challenges of Internet of Things (IoT) in the banking industry

II. LITERATURE REVIEW

When the web appeared in the late 90s, it was hard to imagine how much the internet would dramatically change the business ecosystem and consumer behavior. Within a few years, banks - like other industries - have seen their business models being reshaped to accommodate technology. Being "jostled" has the virtue of being forced out of one's comfort zone. Thus, the changes that unimaginable a few years ago were put in place (such as the electronic signature of documents). (According to Khanboubi F and Boulmakoul A) [2].

Historically, the term 'Internet of Things' was coined in 1998 by Kevin Ashton at the Massachusetts Institute of Technology (MIT) and defined as it "allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any path/ network and Any service" [30]. After that comes the mobile -Internet: the connection of mobile devices to the Internet then the people-Internet: connection which is supported by the social networks. Finally, it progressed to the Internet of Things: the interconnected objects world [31].

The main goal of digital technology is to personalize customer experience which means offering products and services at the right time, in the right packaging and into the right channel [32].

The Internet of Things (IoT), sometimes referred to as the Internet of Objects, will change everything including ourselves. The Internet has largely influenced education, communication, business, science, government, and humanity [36]. The Internet is considered one of the most important and powerful creations in all of human history and now with the concept of the things connected to internet, it can touch more lives. [37]

In IoT, an entity could mean a human, animal, car, logistic chain item, electronic appliance or a closed or open environment [38]. Interaction among entities is made possible by hardware components called devices [39] such as mobile phones, sensors, wireless transmitters, actuators or RFID

ISBN -978-93-5406-047-02

tags, which allow the entities to connect to the digital world [40].

Even though IoT offers a multitude of benefits, several security threats are observed [41–43].

III. APPLICATIONS, BENEFITS AND CHALLENGES OF IOT IN BANKING SECTOR

This section intends to list applications, benefits and challenges of IoT in Banking domain.

A. Applications of Internet Of Things in Baking Domain

Connected objects as a technology can help banks to stay competitive. Nowadays, consumers expect the banks to introduce innovations in products and services, especially the new generation banks who can provide the appropriate services to their new way of connected life [6]. Below, we explain 7 digital trends using IoT that have a direct impact on financial services: mobile banking, M-banking, crowdbased financing, virtual money, high frequency trading firm, cyber criminality, big data and IT analytics.

- *Mobile banking*

Account management on things**:** in the age of digital services, consumers are now claiming easy, seamless and instant access to all banking services. Thanks to the Internet of Things, it is possible for a customer to access his banking account from any digital interface. Also, biometrics represents all computer techniques designed to automatically recognize an individual based on his or her physical, biological or even behavioural characteristics. Biometric data is unique, permanent and specific to individuals (DNA, fingerprints, etc.). It allows an account access and banking service management from any device or 'object' having a digital interface.

Substituting physical signatures: Traditionally, several banking services rely on physical signatures. But, this that can be replaced by "Wet Ink" technology: the cloning on paper of physical signature made via any touch screen device. This eliminates the need for the customer to be physically present at the branch.

Real-time monitoring of collaterals and assets**:** IoT technology allows banks to monitor and track the status of the assets financed (car, appliance, industrial machinery,). Thanks to the digital identity of people and objects, the request for financing and the process of ownership transfer can be made automatic and entirely digital. This also allows the banks to monitor the quality of collateral by checking the condition of the assets and judging whether to keep hold of them physically or not. For example, if a bank finances an engine and the customer defaults on payment, the bank can deactivate it remotely.

- *M-banking*

Automated payment through a great number of endpoints: Internet of Objects, formerly limited to tasks such as counting the number of steps or measuring the heart rate, has been extended to allow payments. The future of payment focuses on diversification of means of settlement. Instant payment is the "buzzword" now and even the banking regulators are encouraging it. Contactless technology provides a fair degree of security and integrate well into the world of connected objects. On a similar note, biometrics technology offers a fairly comfortable level of security for increasingly innovative connected objects. This ability to integrate contactless on any object big or small, makes it appealing to customers. The future is expected to bring about innovations such that any object can become a means of payment. Many companies are already working on making this a reality. For example, Levi's and Jacquard by Google developing the "Commuter Trucker Jacket", a jacket that integrates contactless payment directly into the sleeve. Handheld devices to jackets – a phenomenal change in the way we make payments!

Wallet of things: A digital wallet is a device that can store money without the need for a bank account and make payments directly to any payment terminal. A device or equipment, powered or otherwise, will host an attached, prefunded wallet to facilitate instant payments.

Smart Contracts: The owner of Burj Khalifa in Dubai, the tallest tower in the world, disconnects the elevator system for tenants who are behind on their rent payments. This is only an example of how it's possible to connect an action (here, the deactivation of the elevator) to a condition (the non-payment of rent) automatically through smart payment contracts, with least human intervention. So, banks can now offer consumer products which are automated and instantaneous.

- *Crowd-based financing*

Crowd-based financing is a mechanism for collecting financial flows - usually small amounts - from a large number of individuals via an internet platform - to finance a project. Crowd funding can also benefit from the emergence of the IoT by using its new technologies, terminals and platforms. The quality of borrowers or their repayment capacity can be analysed by evaluating and scrutinizing the data from different IoT devices.

- *Virtual money*

Block chain technology holds great potential for the future. It could revolutionize many sectors of the economy, starting with banking and insurance. Block chain allows decentralized storage and a more transparent and secure transmission of data. It looks like a large database that contains the history of all the exchanges made between its users since its creation. Block chain can be put to use in three ways: for the transfer of assets (currency, securities...), for better traceability of assets and products and for executing contracts automatically ("smart contracts"). It can be used on IoT platforms to cope with digital challenges: having an analytical model to track the records the generated during an IoT process, enforcing strict identification rules to enhance security and finally

ISBN -978-93-5406-047-02

facilitating instantaneous payments between devices and network participants [7].

- *Risk Mitigation in Trade Finance*: RFID is the de facto standard in tracking high-value assets in transit. A more accurate and fine-grained tracking of the asset can be achieved through IoT. For instance, monitoring temperature of the container for shipments involving temperature sensitive goods such as pharmaceuticals and medicinal molecules. If one of the parameters being monitored goes out of the limits, an alert can be sent to the shipping company or logistics service provider.
  Implementations like these can help in risk mitigation and allows the bank management to make informed decisions for scenarios involving trade finance.

• *High frequency trading firm*

Smart algorithmic models: With the quantum of data flowing to the internet from smart devices increasing day by day, there is an opportunity to build better algorithmic models allowing maximum gain and targeted manoeuvres. The companies that make the first step will have a comparative advantage over others.

- *Cyber criminality*

Biometrics is a technology that makes it possible to recognize an individual based on their unique physical and behavioural characteristics. This technology is at the centre of banking system security. Generally speaking, there are two essential steps in the user journey that require different levels of security: authentication and validation. Banks are transforming their security systems to accommodate newer technologies like biometrics. In this sense, Barclays has pioneered to create a system based on venous impression (which is a safer bet than fingerprint as it's practically impossible to duplicate) to validate transactions, make payments or subscription to offers [8].

- *Big data and IT analytics*

Personal financial management**:** PFM can be considered a twenty first century product. It originated in the US in the early 2000's. PFM works on this basic principle - allow customers to have an accurate picture of their accounts, their incomes and their expenses. The principles underlying PFM stands common for all banks. PFM solutions create a panoramic view of all the flows in and out of banking accounts. Banks can provide better services by drawing insights from this data. It is also possible to study the consumption patterns of the customer and activate/deactivate any specific product or service offered by the bank.

Know Your Customer: KYC is a distinguishing feature of our banking system. KYC refers to the procedures of identification and customer information acquisition. KYC stands for "Know Your Customer". KYC procedures are implemented to conform to the regulations and prudential requirements set by the Government to prevent money laundering and financial fraud. The data so collected can be put to use in marketing. IoT will help create a unique, global digital identity for each customer and helps to provide services tailored to their needs. For example, credit card service with promotions and offers from businesses where the customer shops regularly.

- *Improve security*: Security is considered a top priority in every financial institution. IoT technologies help to improve security by detecting and preventing fraud even before they occur. With IoT, financial institutions can track the location of a financial crime, identify the type of device used in carrying out such crime and even get to the root of it on time. IoT also prevents fraud through authentication where wearable devices can be used to gather the biometric data of a customer; the combination of such biometrics and strong cryptography make it possible for the information to be secure in the bank's database.

- *Enhances customers experience through privacy***:** IoT enhances the customer experience by bringing the banking services to them. Customers won't have to wait in long queues at the bank. The technology also helps to improve the quality of self- service kiosks and virtual assistants. IoT uses an array of micro sensors that allow banks to collect and manage financial data in a secure manner.

- *Determine the right placement of banks and ATMs:* Equipped with the data provided by IoT, banks can find ideal locations for their branches and ATM machines that matches the customers' demand. This will stop the misspending on developing branches in areas where it not actually required.

- *Voice assistants:* Banks are extensively using virtual assistants to provide quick and easy services. It was Capital One that first integrated its services with Amazon's Alexa in 2016, which allowed the financial data of the client to be processed by the voice assistant. This provided the clients a quick and easy channel for real time access to their accounts and to query the status of their credit cards, loans, etc.

- *Data Analysis:* Mobile apps and sensors continuously collect data about the user. Based on this data, banks and financial institutions can analyse customer behaviour and make better decisions which are in the best interest of the clients.

- *Proactive Services:* In banking and finance systems, detection of service faults and finding timely solutions to them can be done with the help of IoT. The intersection of these technologies allow the tracking of past transactions, customers' behaviour and identify suspicious activity in accounts.

- *Manage customer relationship***:** The improvement in services through IoT establish a healthy relationship between

ISBN -978-93-5406-047-02

banks and their customers leading to a better customer experience. IoT helps banks to provide services that can meet the needs of the customers. [9]

B. Benefits of adopting IoT in the financial sector

Fintech companies that are looking forward to introduce IoT into their digital infrastructure might put their decision on hold as they are doubtful of the benefits to be derived from such an investment. The applications of IoT in banking and other domain have been identified and laid out in detail; but, the financial sector is yet to identify the ways to take full benefit of this emerging technology.

- *Improves users' financial habits*: IoT devices, especially in the form of wearable can promote positive financial habits and provides means to check excessive spending. The example of Interact IoT, the first IoT bank, can be cited here. They made use of shock wearable as a part of an educational program for their users. The user has the option to set a credit card spending limit and a wearable will track their transactions throughout the day. The user will be alerted, if he/she is nearing the limit. The wearable is capable of sending a shockwave to the user's wrist, if in case they ignore the alert. The shockwave serves as a reminder not to spend anymore.

- *Raises the quality of the banking experience*: The Internet of Things provides personalized services to the clients and offer them valuable insights. The usual procedure of visiting a bank can be replaced by a system of scheduling an appointment online and tracking its status through mobile phone. The customer won't have to wait in lines anymore. As for the bank or service provider, the availability of information regarding transactions made by the account holder and frequently asked questions will help them to serve the customer better. Citibank has been a trailblazer to come up with a Bluetooth-empowered system with IoT beacons that provides 24/7 access to the ATMs for the bank's customers.

- *Expanding the range of services outside banking*: Banks can venture into service that does not fall under traditional banking services and extend the range of services outside banking and deposits. One such successful example is of the U.S Bank which and their IoT initiative that motivates clients to stay fit. Completing fitness achievements will earn the user bonuses and financial rewards.
- *Increases the efficiency of branch banking:* The influx of mobile banking reduced the importance of branch system. Financial service providers combine the power of Internet of Things and banking applications to make sure that the traditional branch banking adds value to the system. A customer visiting branch can be identified with biometric sensors and the main system will be given an alert.
- *Better credit card experience*: The Internet of Things offers the possibility of developing interactive credit card. IoT extends the concept of credit card beyond a plastic card - an interactive digital display can be used by clients to direct questions to the bank in real time; A user can change the credit card limit settings at the shopping site, and so on.
- *Upgraded ATMs:* ATMs can sometimes give the perception of an old-school technology and can be frustrating to the customers. The Internet of Things can fix this connection by making better connection between ATMs and bank branches and developing ATM machines that are fully manageable with a smartphone. The use of motion sensors enables a user to find the closest ATM to their current location. Any faults in the ATM machines will be communicated to the branches so that an engineer can be assigned immediately to fix it, thus reducing machine downtime.
- *Automated business processes*: IoT systems can act as a means of automation in banking and financial sector providing services like instant loan processing and collateral monitoring. IoT can take care of request handling, automatically disable credit cards if in case payments are not made in time, transferring asset ownership and so on. The Internet of Things as a means of automation will enable instant loan processing and collateral monitoring.

C. Challenges of adopting IoT in the financial sector

Protection of IOT (Internet of Things) is complex and a difficult activity [24]. Although adoption of IoT technologies offers many benefits for the banking and financial institutions, some major challenges may be faced by financial institutions when implementing IoT. Most of the identified challenges are reported in [25], [6], [27], [28] and [29] studies

Some of these challenges include:

- *Technology Adoption:* There is no "one-size-fits-all" IoT solution. IoT applications in a business must be custom tailored to its core business objectives to reap the complete benefit of this revolutionary technology. Such a personalization demands substantial investment of time and resources and the right IoT data management solution provider who can understand the organization's hardware and software requirements. According to Microsoft's 2019 IoT signal's report, 38% of respondents stated that the major challenge for IoT adoption is the complexity or technical challenges since there's a lack of industry expertise.
- *Data Fusion Management and Analysis*: Although institutions accept the fact that that consumer data forms the basis to understand their customers, many organizations does not have proper infrastructure in place to comprehend and interpret the specific impact this data has on their business. B2B data management can pose problems as many companies lack the subject knowledge on this matter and the skill set for efficiently managing the information collected through IoT sensors. Another such hurdle is cross-border enterprisegrade data

ISBN -978-93-5406-047-02

transfers, given the fact that banks lack the appropriate infrastructure to fulfil the transfer securely and efficiently.

- *Multidimensional Data Privacy and Security Access:* Information stored in IoT devices can be vulnerable to threats like first-hand attack, gossip attack, observation attack, inference attack, automated invasion attack [21]. To counter these attacks, a proper security policy must be enforced. Critical security issues cantered around IoT have grabbed the attention of several companies in the domain. Availability of a large number of IoT hubs connected to the web will give better odds to invade the system, as many devices still have security holes.

  Smart-home devices and CCTV cameras get compromised first and the attacker will and deploy these IoT gadgets against their own servers. IoT security is expected to evolve depending on the influence IoT has on our lives.

- *There are no set/common standards for equipment maintenance:* There are different manufactures who build IoT hardware equipment. Each of these devices will follow a different approach to maintenance. A lack of common standard in IoT hardware can be pressing problem in many situations. This cannot be essentially solved by the manufacturers agreeing to adopt a common standard. Even if the manufacturers reach such an agreement, there will be other technical issues to handle. Hardware. The only solution would be to have only one supplier of these IoT devices. But, this is highly impractical on the grounds that the monopoly will tear the economy to ruins. Technology conventions incorporating network and communication protocols, and data-aggregation conventions, are the collection for activities that handle, process and store information obtained from several sensors. These enhance the data by increasing the scale, scope, and frequency of data available for analysis [33, 27].

- *Data encryption and key management:* A smart environment will have to address the encryption of the data it stores or manages. Encryption algorithms like AES, RSA, and DH use keys of longer in length whereas Elliptic Curve Cryptography algorithm uses shorter length key. ECC is preferred over other algorithms for use in IoT devices as the computing power available is limited and using a lengthy key might degrade the performance of the system [10].

- *Connectivity:* Ensuring connectivity with several devices can be a challenge for the future of IoT, as the currently existing systems will undergo a great degree of change. Currently, a centralized, server/client architecture is the common implementation used to authenticate, authorize and connect several terminals in a network [33].

- *Intelligent Analysis & Actions:* The final step in the implementation of an IoT system is the revelations derived from the data provided for analysis. Several cognitive technologies and models are put together to conduct data analysis. There are certain parameters that encourages intelligent actions to be incorporated in IoT. Some of them are lesser device cost, enhanced device functionality, the machine "influencing" human actions through behavioural-science rationale, deep learning tools, machines' response in unusual scenarios, improvement in information security and privacy and device interoperability [35].

- *IoT technologies are complex:* People lack understanding of how IoT works because the concepts of this technology can be quite complex. An IoT system requires connection to be established between many devices through a network. If a device fails or a break happens in the connection, the whole system will break down resulting in losses. In order to avert such disasters, banking and financial institutions must choose a qualified and reputed vendor for sourcing the IoT hardware and a competent software development company to build the application.

Sly No	Challenges	References
1	Managing device diversity Scale, data volume and performance, Flexibility and evolution of applications Data privacy Need for medical expertise	[17]
2	CPU capacity Memory of the system constrained over network performance like bandwidth	[7]
3	Data exchange Availability of resources Privacy	[4]
4	Hardware implementation and design optimization issues	[13]
5	Security challenges	[14] [4],[15], [16]
6	Interoperability	[4],[17],[18]
7	Technical challenges: Modelling relationship between acquired measurement and diseases. Software implementation of medical analytic schemes	[19]
8	Real time processing System predictability	[15]
9	Data integration	[17], [18]
10	Unstructured, growing and diverse data at exponential rate	[20]
11	Security in the Internet of Things	[10]

ISBN -978-93-5406-047-02

12	Multidimensional Data Privacy and Security Access	[22]

D. Implementations of IoT in the banking and finance industry

Some of the earliest and innovative IoT projects implemented in the financial sector are listed below:-

- *Supply Chain Optimization*: Commonwealth Bank of Australia (CBA) has achieved great feats in the adoption of the Internet of Things. CBA, in the last year, successfully executed a block chain-enabled global trading experiment where distributed ledger technology, smart contracts, and IoT were used in combination to propagate new supply chain efficiency alongside their current operations. The benefit that CBA derived from implementing this technology combination was accurate tracking of their shipments and better analysis of metrics related to the container in transit. As a result, the company was able to achieve a higher level of data transparency and shipment efficiency regarding the location, condition, and authentication of goods being transported. [22]
- *Interactive credit cards*: Marketing communication of the bank or its partners can be customized according to the interests of the customer through interactive credit cards. It can also improve the speed of delivering customer service to credit card holders. Although smart credit cards were introduced in Mobile World Conference 2018, we are yet to witness a wide-scale adoption.
- *London Test Bed:* - An organized effort in adopting IoT technology was initiated by Clydesdale and Yorkshire Banking Group by opening an IoT tested in London to prepare itself for the shift in smart financing and the competitive demands of personalized banking experiences. They are working on a project to train voice assistance products like Amazon Alexa so that the customers can have more power and flexibility managing their personal finances.
- *Innovation within Retail Banking:* IoT has been hailed as a technology that can revolutionize the retail sector by Forbes, a leading business magazine. A complete revamp of the retail sector is expected in the upcoming years. More than 70% of retail businesses have realized the potential of IoT and is ready to make it a part of their organization. Banks, in particular, have made excellent use of this innovative technology to achieve faster trading and processing times. Investing heavily on digital products and reducing the transaction costs online has led to the acquisition of many new retail customers.
- *Banking on wearable:* The wearable market is finding great traction owing to its low price and functionality. Banks have been able to integrate payment apps to wearable like Fit Bit, Apple watch and others. Some banks even have their own specialized devices and payment solutions.
- *Tailored Marketing*: Customers have better awareness of the market now and they actively demand personalized solutions that meet their varying needs. The banks will have to acquire the right information about client's needs, economic conditions and buying behaviors, which helps them create a customized banking solution. With the help of IoT, all consumer activities can be tracked by the bank to come up with a solution which meets their need.
- *Tailor-Made Auto Insurance*: Insurance Companies have shown signs of welcoming the IoT technology, by offering devices that plug into the on-board diagnostic port of cars and send driving behavior data back to them. The insurance company can assess the driver habits, for which the owner will be rewarded with discounts. Modern automakers, like Tesla Motors, are working towards a whole new level of digitization in the automotive industry. Tesla Cars even have a Linux-based OS that automatically upgrades features "over the air". Such vehicles will be collecting a lot of data about the vehicle and the driving. Insurance company can use this metrics to deliver tailormade insurance to customers based on driving habits, engine health and general wear and tear of the vehicle. Insurance companies will also have access to the GPS data on the actual speed of the vehicle in speed sensitive zones (such as schools or residential areas), which can be used to gain insights into the likelihood of accidents and price insurance premiums appropriately.
- *Branching out to connected cars:* Banks can also extend their digital services to smart cars so that the customer's experience is improved. For instance, Idea bank has an introduced the smart cars fitted with an integrated security deposit box and ATM.
- *Blockchain-based smart contracts*: Blockchain find great application in banks to secure records of authenticated transactions. For instance, the Commonwealth Bank of Australia (CBA) was the first to complete a global trade transaction between two banks by leveraging the capabilities of block chain, smart contracts and the IoT.
- *Smarter branches*: IoT technology can help make branches smarter. In a way, these smart branches can be a threat to the retails bank branches but definitely improves customer experience.
- *Chatbots for improved customer's experience:* Chatbots helps the banking and financial institutions to automate Customer service inquiries received by a bank or financial institution can be automated with the help of chatbots. There is active development going on in improving the functionalities of chatbot and several startups have been associated with it.
- *IoT enabled Smart Payment Contracts:* Smart contracts can be defined as computer programs that facilitate, verify, or enforce the negotiation or performance of a contract. Riding on IoT technology and powered by smart contracts and digital identity, payments can be made partially or fully self-executing and self-enforcing. Examples would include payment after a 7 days' trial period for home appliances or controlling access to a house based on timely payment of rent. IoT-assisted smart contracts can help in process automation and mitigation of operational risks. This model can also guide the development of new products which offer better customer experience.

ISBN -978-93-5406-047-02

IV. CONCLUSION

It is safe to say that the Internet of Things is strengthening its foothold in the banking and financial services industry. In this paper, we provided an overview of IoT services and technologies in banking sector. A number of research challenges have been identified, which are expected to become major research trends in the next years. IoT in the financial and banking sector is still at its infancy; but, there is no denying that IoT holds great potential to revolutionize the sector. Banks and financial institutions who are early to identify this opportunity will benefit greatly and will enjoy an advantage over other firms. Please include a brief summary of the possible clinical implications of your work in the conclusion section. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. Consider elaborating on the translational importance of the work or suggest applications and extensions.

REFERENCES

[1] Internet of Things (IoT): number of connected devices worldwide from 2012 to 2020 (in billions). https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[2] Khanboubi F. and Boulmakoul A... A roadmap to lead risk management in the digital era. ASD 2018: Big data & Applications 12th edition of the Conference on Advances of Decisional Systems, At Marrakech Morocco. 2018.

[3] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[4] Smart airport: an IoT-based Airport Management System. Samia Bouyakoub, Abdelkader Belkhir, Fayçal M'hamed Bouyakoub, Wassila Guebli less

[5] G, g, KUO - HUIYEH, (Senior Member, IEEE), "A_Secure_IoT-Based Healthcare System with Body Sensor Networks" IEEE, The journal for rapid open access publishing. doi: 10.1109/ACCESS.2016.2638038

[6] Petracek N. Is Blockchain The Way To Save IoT? FORBES. [Online]. 2018 [cit. 2018-09-02]. Online: https://www.forbes.com/sites/forbestechcouncil/2018/07/18/isblockchain-the-way-to-save-iot/#65d086d25a74

[7] Mckinsey. Transforming a bank by becoming digital to the core. Mckinsey and company. [Online]. 2018 [cit. 2018-10-08]. Online: https://www.mckinsey.com/industries/financial-services/ourinsights/transforming-a-bank-by-becoming-digital-to-the-core

[8] SIA partners. Biométrie : vers un monde bancaire plus sécurisé ?. Finance and strategy by SIA partners. [Online]. 2017 [cit. 2018-1112].Online:http://finance.sia-partners.com/20170920/ biometrievers-un-monde-bancaire-plus-securise#_ftn1.

[9] Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for realising the Internet of Things. Cluster of European Research Projects on the Internet of Things, European Commision, 3(3), 34-36

[10] Hui Suo, Jiafu Wan, Caifeng Zou, Jianqi Liu, 2012. Secruity in the Internet of Things: A Review. International Conference on Computer Science and Electronic Engineering, IEEE Computer Society, 648651.

[11] http://www.tcs.com/resources/white_papers/Pages/Internet-of-Things-Medical-Devices.aspx

[12] Shu-yuan Ge, Seung-Man Chun, Hyun-Su Kim and Jong-Tae Park, "Design and Implementation of Interoperable IoT Healthcare

[13] Georges Matar, jean-marc Lina,Georges Kaddoum, Anna Riley, "Internet of Things in Sleep Monitoring: An Application for Posture Recognition Using Supervised Learning". doi :10.13140/RG.2.2.21729.30561

[14] Robert S .H. Istepanian, Ala Sungoor, Ali Faisal, Nada Philip, "INTERNET OF M-HEALTH THINGS " m-IOT" ", Assisted Living 2011, IET Seminar on, 16 April 2012. doi: 10.1049/ic.2011.0036

[15] Vasileios Tsoutsourasm Dimirta Azariadi, Konstantina Koliogewrgi, Sotirios Xydis and Dimitrios Soudris, "Software Design and Optimization of ECG Signals Analysis and Diagnosis for Embedded IoT Devices", doi:10.1007/978-3-319-42304-3_15

[16] Rashmi Singh, "A Proposal for Mobile E-Care Health Service System Using IOT for Indian Scenario", Journal of Network Communications and Emerging Technologies (JNCET) Volume 6, Issue 1, January (2016) © EverScience Publications. ISSN: 2395-5317

[17] http://www.tcs.com/resources/white_papers/Pages/Internet-of-Things-Medical-Devices.aspx

[18] Boyi Xu, Lida Xu, Hongming Cai, Lihong Jiang, Yang Luo & Yizhi Gu, "The design of an m-Health monitoring system based on a cloud computing plat form", Talor & Francis 2015. doi: 10.1080/17517575.2015.1053416

[19] Hyun Jung La Han Ter Jung, and Soo Dong Kim *, "Extensible Disease Diagnosis Cloud Platform with Medical Sensors and IoT Devices", 2015 3rd International Conferenceon Future Internet of Things and Cloud. doi : 10.1109/FiCloud.2015.65

[20] Pallavi Chavan , Prerna More, Neha Thorat, Shraddha Yewale & Pallavi Dhade, "ECG - Remote Patient Monitoring Using Cloud Computing", Imperial Journal of Interdisciplinary Research (IJIR) Vol-2,Issue-2 , 2016 , ISSN : 2454-1362

[21] Dieter Uckelmann, Mark Harrison, Florian Michahelles, 2011. An Architectural Approach Towards the Future Internet of Things. Architecting Internet of Things by Springer,1-22

[22] A. Kumara and O. Shoghlia –"A review of IoT applications in Supply Chain Optimization of Construction Materials" DOI: 10.22260/ISARC2018/0067

[23]International Telecommunication Union (ITU). (2015). Internet of things global standards initiative. Retrieved January 8, 2017, from http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx.

[24]The Challenges Facing with the Internet of Things Helmi O, Ehsankanani, Sokeh MA, Sepidnam G. Int J Sci Stud 2017;5(4):527-533.

[25]R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future Internet: The Internet of Things architecture, possible applications and key challenges," in Proc. 10th Int. Conf. FIT, 2012, pp. 257–260.

[26]. A. Gluhak et al., "A survey on facilities for experimental Internet of Things research," IEEE Commun. Mag., vol. 49, no. 11, pp. 58–67, Nov. 201.4.

[27] Z. Sheng et al., "A survey on the IETF protocol suite for the Internet of Things: Standards, challenges, and opportunities," IEEE Wireless Commun., vol. 20, no. 6, pp. 91–98, Dec. 2013. [28]. J. A. Stankovic, "Research directions for the Internet of Things," IEEE Internet Things J., vol. 1, no. 1, pp. 3–9, Feb. 2014

[29] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A vision of IoT: Applications, challenges, and opportunities with china perspective," IEEEInternet Things J., vol. 1, no. 4, pp. 349–359, Aug. 2014 [30] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé. Vision and challenges for realizing the internet of Things. European Commission Information Society and Media, Luxembourg, Tech. Rep. [Online]. 2010 [cit. 2018-11-20]. Online: http://www.internet-ofthingsresearch.eu/pdf/IoT_Clusterbook_March_2010.pdf. [31]Charity P, Chi Harold L., Srimal J., and min C.2015. IEEE Access. The journal for rapid open access publishing. A Survey on Internet of Things from Industrial Market Perspective. DOI: 10.1109/ACCESS.2015.2389854

[32]Alf D. and al. Four Ways Banks Can Radically Reduce Costs. BCG. [Online]. 2018 [cit. 2018-10-02]. Online: https://www.bcg.com/publications/2018/four-ways-banks-can-radicallyreduce-costs.aspx.

[33] Vermesan, O., Friess, P., Guillemin, P., Gusmeroli, S., Sundmaeker, H., Bassi, A., & Doody, P. (2011). Internet of things strategic research roadmap. Internet of Things-Global Technological and Societal Trends, 1(2011), 9-52
[35] Theoleyre, F., & Pang, A. C. (Eds.). (2013). Internet of Things and M2M Communications. River Publishers.
[36] M. A. Ezechina, K. K. Okwara, C. A. U. Ugboaja. The Internet of Things (Iot): A Scalable Approach to Connecting Everything. The International Journal of Engineering and Science 4(1) (2015) 09-12.
[37] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.
(references)
[38] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions,"
Future Generation Computer Systems, vol. 29, no. 7, pp. 1645–1660, 2013.
[39] S. De, P. Barnaghi, M. Bauer, and S. Meissner, "Service modelling for the internet of things," in Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on. IEEE, 2011, pp. 949–955.
[40] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's intranet of things to a future internet of things: a wireless-and mobility-related view," Wireless Communications, IEEE, vol. 17, no. 6, pp. 44–51, 2010. [41] I. Cha, Y. Shah, A. U. Schmidt, A. Leicher, and M. V. Meyerstein, "Trust in m2m communication," Vehicular Technology Magazine, IEEE, vol. 4, no. 3, pp. 69–75, 2009.
[42] J. Lopez, R. Roman, and C. Alcaraz, "Analysis of security threats, requirements, technologies and standards in wireless sensor networks," in Foundations of Security Analysis and Design V. Springer, 2009, pp. 289–338.
[43] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Computer Networks, vol. 57, no. 10, pp. 2266–2279, 2013.

ISBN -978-93-5406-047-02

Browser Security: Attacks and Detection Techniques

Dr.Sarika S
Assistant Professor
Department of Computer Science
Naipunnya Institute of Management and Information Technology,Koratty,Pongam

Abstract-**Nowadays, the growth of the technology reaches its height and everything goes online. At the same time Cyber attacks are increasing to breach the information system of another individual or organization. Phishing is an increasingly common cyber threat in IoT . It is a malicious and deliberate attempt of sending fraudulent communications that appear to come from a reputable source or mimicking a webpage which looks like an original website . The goal is to steal sensitive credentials like login information and credit card details or to install malware on the victim's machine. Browser-based cyber threats have become one of the biggest concerns in networked architectures. The most prolific form of browser attack is tabnabbing. This paper presents an overview of tabnabbing detection techniques and proves the efficiency of agent based tabnabbing detection by comparing with the state of the art tabnabbing detection techniques.**

Keywords: Browser security , Social engineering , Tabnabbing attack , Phishing , software agents.

I. INTRODUCTION

The digital world is changing at a tremendous speed and social media communication tools have profoundly changed our interactions with the world around. The speedy transition to today's digital world has been boosted by network digitalization. The new communication technologies bring new opportunities, but also open up door for a number of risks. Technological advancement accelerates growth of social, cultural, managerial, political and organisational environments but also hasten the origin of new threats. In the present day, almost every person has his or her web identity. Online activities and social networking have grown to such an extent that governments safe keep the entire personality or identity of a citizen on web servers. Illegal activities too are increasing in the virtual world. People are subjected to direct and indirect security threats, for which they need some extra safety measures. Attackers will use a mix of social engineering and other sophisticated tools to trick users which require more skill and work hours.

The web has become a staple for information sharing and processing, and web browsers are the popular interface for deliverance of information. Web browsers are at the heart of the security problems that affect users as they are an appealing target for attackers. They have a huge and complex trusted computing base but face the challenge of keeping their users safe while using web applications.

Web browsers are the full-blown software suites, for locating and displaying webpages on the internet. The World Wide Web operates on a client/server model where a web client running on the web browser of a computer contacts a web server for information retrieval. As we use web browser as a the initial link to the rest of the internet, they are the crucial point of vulnerability for attack or exploit to happen. Protecting a browser from today's cyber threats is an important and daunting task. The browser based attacks originate due to poor security coding of web applications. Many web applications try to enhance the browsing experience by enabling various functionalities, but this might be unnecessary and may leave the client susceptible to being attacked. Each functionality added to the browser has its own unique vulnerabilities or weaknesses. The browsers store the statistics of online activities using cookies, cached pages, and history. Cookies are small data files maintained in user's web browser to record the user's browsing activity. Cached pages are stored copies of websites recently visited by the user. They are used to improve the system performance of user but also might be accessed by unauthorized parties. Browsers save the history of all the websites visited by the user to get quickly to the websites he visits more. Now, users choose browsers with most security patches. But it is impossible to know how secure they are until hackers have poked at them. Windows Internet Explorer, Mozilla Firefox, Apple Safari, Google Chrome, and Opera are the major browsers that people use to surf internet today. Thus, it is important to increase the security of the browser while browsing or mailing so we are better defended without divulging sensitive information to bad guys or opening the door for attackers.

The main objective of this paper is to give an overview of various browser security threats with main focus on tabnabbing attack. Further, an analysis is done with the existing anti-tabnabbing techniques with the proposed agent based method to detect tabnabbing.

The paper is structured as follows. Section 2 details about various browser security threats. Section 4 explain about the state-of-the-art tabnabbing detection techniques. Section 5 gives a brief idea about agent based method for tabnabbing detection.Implementation is explained in section 6 and experimental evaluation is done in section 7.Further, a comparative analysis is done in section 8 and section 9 concludes the paper.

II. BROWSER SECURITY THREATS

A. Clickjacking

Clickjacking[1] is a malicious technique of tricking web users into revealing confidential information or taking control of their computer while clicking on seemingly innocuous web pages. The principle behind these attacks is

ISBN -978-93-5406-047-02

that users can be tricked into clicking on things that they do not see or are aware off. Usually an invisible frame is loaded, along with some content, and laid over a simple game, or something similar, that gets the user to click multiple times at specific places. While a user thinks they have clicked somewhere in the game, they actually click on an invisible layer which perform some other action. Thus, the attacker is "hijacking" clicks meant for their page and routing them to another page, most likely owned by another application, domain, or both. Clickjacking, also known as UI redressing, is possible not because of a software bug, but because seemingly harmless features of web pages can perform unexpected actions.

B. Cross Site Scripting (XSS)

Cross Site Scripting (XSS) [2] attacks are a common type of injection attack where an attacker injects a malicious piece of code into an otherwise benign site. A cross site scripting vulnerability may be used by attackers to bypass access controls and to impersonate users. XSS attacks occur when an attacker uses a web application to send malicious code, generally in the form of a browser side script, to a different end user. Flaws that allow these attacks to succeed are quite widespread and occur anywhere a web application uses input from a user within the output it generates without validating or encoding it.

XSS can also be used by an attacker to send a malicious script to an innocent user. The end user's browser will execute the script as it doesn't show any malicious nature. As the executed script masquerades a trustworthy source, it can access cookies, session tokens, or other private information maintained by the browser and can also rewrite the content of the HTML page.

XSS attacks do not target a victim directly instead they steal a user's session cookie or exploit the web application that the user would visit. Alternatively, the attacker may use a fraudulent website as a vehicle to deliver a malicious script to the victim's browser. These phishing sites operate under the real domain, sending credentials back to the attacker. JavaScript is most widely used for malicious scripting.

C. Cross Site Request Forgery

Cross Site Request Forgery [3] is abbreviated as CSRF or XSRF, is an offensive practice of attacking a website in which an attacker masquerades as a legitimate user and sends malicious scripts to a website where the user is recognized as authentic. The attack is executed in such a way that a hacker inserts malicious codes into a link on a website that seem to be from a trustworthy source. When the user clicks on the link, the embedded code is submitted as a client web request and is executed on the user's computer. CSRF is also known as one-click attack or session riding as it happens from the valid session of user's browser. This attack exploits the trust that a website has in the user's browser and the trust that a user has for a specific site. The danger is not with the victim's browser or the site hosting the CSRF but in the affected web application. Whenever a request comes from a user's browser with a valid session, the web server has no way of knowing if the request was from an authorized user or not. The web server just process the request what it is received but it might have been issued without the knowledge of user from a website with a hidden script. CSRF attacks can be launched with the help of social engineering techniques for tricking the users of a web application to perform state changing requests like transferring funds, changing email address etc. These attacks specifically target web applications like email clients, social media, e-commerce applications, and online banking interfaces.

D. Session Hijacking

Session hijacking [4] is a kind of phishing attack where user's activities are monitored clearly until they log into a target account like the bank account and establish their credentials. At that point, the malicious software takes control and can undertake unauthorized actions, such as transferring funds, without the knowledge of the user. The goal of a session hijacking attack is to transfer the user's authenticated session to a different machine or browser, enabling the attacker to continue working in the victim's session. To achieve this, the attacker hijacks the session that the user has established with the target application.

E. Social Engineering

Social engineering [5] is the criminally fraudulent process of tricking users and gaining access to secure systems to acquire sensitive information such as usernames, passwords and credit card details by masquerading as a trustworthy entity in the web. In social engineering, the attacker uses human interaction to make an innocuous feeling to obtain or compromise information but in real, it isn't. Social engineering preys on common aspects of human psychology and it requires minimal technical knowledge. They range from wide scale attacks, which are crude and can normally be easily identified, through sophisticated multifaceted personalized attacks which use a range of social engineering techniques and are almost indistinguishable from legitimate interactions. The tactics of social engineers can be expected to evolve, to take advantage of new technologies and situations. One of the most common and effective form of social engineering is phishing.

F. Phishing

Phishing [6], the most prolific form of social engineering is an art of harvesting sensitive information from users by making counterfeited websites which impersonate legitimate ones. Phishing employs both social engineering and technical subterfuge to steal personal and financial account credentials of consumers. Social engineering is the act of manipulating people into performing actions or divulging confidential information. While similar to a confidence trick or simple fraud, the term typically applies to trickery or deception for the purpose of information gathering, fraud, or computer system access; in most cases the attacker never comes face-to-face with the victim. Technical subterfuge schemes plant crime ware onto PCs to steal credentials directly, often using

ISBN -978-93-5406-047-02

Trojan keylogger, spyware etc. According to the recent APWG Phishing Activity Trends Report, both deceptive phishing and spear-phishing that targets specific business employees have been on the rise. Deceptive phishing refers to any attack by which fraudsters impersonate a legitimate company and attempt to steal people's personal information or login credentials. *Spear phishing* is an e-mail spoofing fraud attempt that targets a specific organization, seeking unauthorized access to confidential data. Through the past decades, the number of victims has increased exponentially as phishers improvise tactics by exploiting loopholes in software.

1)Tabnabbing

Tabnabbing [7] is a phishing attack within a browser and it targets a user who keeps many tabs open at a time. The name 'tabnabbing' was coined in early 2010 by Aza Raskin, creative lead of Mozilla Firefox. The attack is simple to implement and silently tracks the victims. As the user navigates through a bunch of open tabs, phishers set up a rogue website which looks exactly like the genuine one and load the inactive tab with the counterfeited webpage. When the user switches back to the tab, it appears to be a site frequently used by the user. As the user sees an innocuous looking page and does not remember how each tab looked like before tab switch, he will give his credentials to the honest looking page and is trapped. Tabnabbing is a dominant browser attack which is likely to deceive even the most incredulous web surfers as it exploits user's trust and inattention in browser tabs. Unlike other attacks, this deception technique happens on the behind and is able to change favicon, title, and layout of a webpage with some other site familiar to the user. Tabnabbing cannot be avoided by using HTTPS instead of HTTP in web address.

The various steps of execution of tabnabbing attack is as follows:

- The user is visiting a webpage which looks perfectly genuine. The user opens multiple webpages like news, mail account or a social networking site in other tabs of the browser.
- The user changes his tab to another or he is forced to switch to another tab when the page takes time to load.
- When a tab is unattended for some time and is out of focus, the favicon, title, and layout of the page is replaced with some other site familiar to the user (a frequently used site by the user).
- As the user pays less attention to which site he was working with before, he will give his credentials to the honest looking site and is trapped.

The attack can be launched using different techniques. One of the methods to launch tabnabbing is via scripting support. This is done by exploiting JavaScript embedded in the webpage. Using JavaScript, it is possible to observe mouse and keyboard events to detect the user activity on that page. The JavaScript can be modified with an impersonation of a popular website, when the page has lost its focus and hasn't been interacted with for a while. The inactivity of a tab is identified using JavaScript *onblur* and *onfocus* events. Once the *blur* event is fired, the script replaces the favicon, title and the page layout with a well known login look-a-like. When the user returns to the tab after a while and see a legitimate looking webpage, he has no need to think that the displayed page is fake. The user is prompted to enter the login details and the data thus collected are redirected to the attacker. The attack can be made more effective by monitoring the websites that the user has loaded in the past or in other tabs, and loads a simulation of the same sites. CSS history probing technique [8] can be used to correctly generate the exact URL that the user has visited recently.

In order to execute the attack, it is not always required to change the tab. The attack can be launched when the user minimizes his window or he is not interacting with the system for doing some other activities. The attacker can detect the inert tab using JavaScript code and replaces the currently focused page with a deceptive site. In some cases, he can fool the user without even changing favicon and title. In this case, the background details of the webpage will be loaded slowly to make an impression to the user that it is taking time to render the webpage. In worst case of the attack, URL itself is unchanged which can even fool the most security conscious web surfers who always check the URL before proceeding their web activities.

Alternate ways of demonstrating this attack rely on the use of HTML refresh meta tag (HTML meta tags, 2017) in predetermined time intervals. This technique does not need scripting support instead the webpage refreshes itself in the background and turns into a fake page. In this variant, the phisher could use *window.onerror* events to recognize the user's login activity. This method does not rely on users to switch a tab, as the phisher assumes that the user might have been changed the tab after a predefined time interval.

Tabnabbing attack can also be performed with the use of *iframes*. An HTML <iframe> tag [9] specifies an inline frame which provides a provision for nesting of documents. In other words, it helps to insert another document within the current HTML document. They can be placed anywhere in the web document.

III. ANTI-TABNABBING TECHNIQUES

Mozilla has released a Firefox plugin called Account Manager [10] for online identity management. Account Manager lets us store the logins which are already created, suggesting them whenever they can be used. It makes the logins more secure by generating random passwords too. NoScript [12] and YesScript [13] are Firefox add-ons, preventing websites from running JavaScript, Java, Flash or other plugins. It provides powerful protection against malicious scripts, XSS, CSR and clickjacking attacks. But they do not provide protection in other browsers.

NoTabNab [14] is a Firefox add-on which protects users from tabnabbing attack by using the positioning of HTML elements of a webpage. This add-on looks out open tabs and track favicon, page title and layout changes for every single tab until a new URL is loaded. The attribute values are

ISBN -978-93-5406-047-02

recorded only for the topmost page element on these tabs and critical CSS information is attached to these records. When a tab switch occurs, the operation is repeated and the add-on compares these values with the previously recorded one. If an impersonation has happened, it alert the user about changes in its layout, favicon or title to mimic another page by highlighting the address bar in yellow or red according to the warning level. The problems related to this technique are follows. As the method doesn't keep data about all elements on the page, it is possible to trick the add-on by putting very small invisible elements under grid points and thereby bypassing the controls. Also the method is not effective if the document contains many iframes within each other as it needs to check each of these elements recursively and compare them with the already recorded data. Another problem in this technique is related to resizing the browser, as only some web pages are designed to re-layout themselves.

A signature based detection mechanism [15] has been presented to deal with tabnabbing. The method defines a set of rules to scrutinize vulnerable JavaScript code. JavaScript can be used to observe mouse and keyboard events to detect user activity on a page and to change the favicon, title and content of the page. They use a signature based detection mechanism to deal with tabnabbing attack with an assumption that dynamic iframes are never used with *onfocus* and *onblur* event of JavaScript. First, the source file is converted into a text file and then into tokens. These values are given to the rule based system which is checked for vulnerabilities. The JavaScript code is considered as vulnerable if both *onblur* and o*nfocus* events are used with dynamically generated iframes within the prescribed time and also a mouse click is not used to detect if an iframe is in focus or not. But this paper focuses only on iframe elements which is not always necessary for a tabnabbing attack.

Tab-Shots [16] is a browser extension which uses visual appearance of a webpage to detect tabnabbing. The method works by remembering what each tab looked like, whenever a tab is changed by recording the favicon and screenshots of the presently focused tab at regular time periods. Then the screenshot is separated into fixed-size tiles. Each tile of the present snapshot is compared to its counterpart in the stored data. If an exact match is not obtained, the non-matching area is marked by a colored overlay. One probable shortcoming of this technique is the difficulty in detecting small changes in a page.

TabsGuard [17] combined heuristic based metrics and data mining techniques to detect tabnabbing. The approach keeps track of the changes made to the structure of a page during the time when the page is idle and uses five heuristic-based metrics to measure the degree of changes made to the tree representation of each webpage whenever a tab loses focus. In this method, three parameters such as title, favicon, and the HTML Document Object Model (DOM) tree of the page are fetched at two time frames T1 and T2. T1 is the time when the page is fully loaded and T2 is the time when the user returns back to that tab after a tab switch. The values of each parameter are compared at two time frames to monitor the changes made to the page between these time frames. For comparing the HTML DOM trees of the page, the heuristic-based metrics such as common paths, cosine similarity, tag frequency distribution analysis, input fields added to the page and iframes in the page before and after tab switch are used to perform the comparison with respect to syntactical similarity. The comparison results are then analyzed using data mining techniques to find the outlier score. When the outlier score of a webpage with respect to the degree of changes become higher than the average outlier score, the page is detected as tabnabbing. Otherwise, it is detected as legitimate. If a page is detected as tabnabbing, an alert message is displayed to the user and adds the detected page to a local blacklist in the user's browser.

TabSol [18] is based on the hash value comparison of the webpage at different instants of time when the webpage is in focus and when the webpage regains focus after a tab switch. TabSol uses a URL whitelist which contains a list of legitimate pages. When a webpage is opened in a browser, the method checks whether the URL (address) is present in the domain whitelist or not. If the webpage is included in whitelist, the page is safe. Otherwise, the page is a potential phish candidate and is unsafe. For the suspected pages, compute the webpage hash value using SHA-1 algorithm and is stored. When the user performs a tab switch and switches back to the previous tab, TabSol recomputes the hash digest of the page and compares with the stored digest. If a match is found, the webpage is secure and can be appended into the domain white-list. Otherwise, if the comparison result is different, the page is considered as doubtful. Now, the method check for login forms in the suspicious page. If a login form is detected, then the page is classified as phishing and otherwise as genuine.

The tabnabbing detection system (TDS) proposed by Al-Khamis and Khalafallah [19] is a combination of content based and non-content based techniques which considered visual and structural features of a webpage. The method is implemented as an extension to Google Chrome browser. In this method, a profile is created for each opened tab which consists of important elements. When the tab is on focus from a nap, a second profile of the same tab is created and is compared with the old one to find any suspicious activity. The detection mechanism proceeds in two layers. The first layer compares the title, URL, and favicon and the second layer compares the HTML structure of the webpage to detect an attack. In order to overcome undetected cases, the method also incorporates user education which will enhance internet users protection.

TabSecure [20] is an anti-tabnabbing technique to handle phishing attack launched through fake websites and deceptive emails. This method has three modules, phishing website detection, tabnabbing detection and email phishing detection modules. The phishing website detection module communicates with the browser monitor engine which watches the activities of the browser. The browser monitor performs IP address check as well as web page content analysis. When a URL is entered in the browser’s address bar,

ISBN -978-93-5406-047-02

it is checked against a database of phishing sites kept by the browser monitor engine. If no match is found, then the browser monitor invokes the feature detection engine where the webpage undergoes a web page content analysis for finding suspicious links embedded in it. The email phishing detection module extracts emails from the mail box and is then converted to a .txt file. Once the mails are extracted, they are sent to a bayesian classifier to separate legitimate and non-legitimate emails. In tabnabbing detection module, the content of a webpage is compared when a tab is on focus and when the tab loses focus after 5seconds. If a change is noticed, then it is considered as a phishing attempt and an alert is displayed to the user with a popup. The method offers an accuracy of 93% but the complexity and time consumption of the system is more.

Existing anti-tabnabbing methods detect the layout change and warn the user only when the tab is on focus after being nabbed and also they focus mainly on the change in page layout, title and favicon but not much attention is given to change in URL. Beyond that, they fail to detect parallelism in attack detection.

IV. AGENT BASED METHOD

The work proposed in this paper use a heuristic method where structural features of a webpage are analyzed as explained in [20][21]. The proposed method is a multi agent system uses agents to concurrently monitor the change in webpage layout at regular intervals in all tabs of a browser and alerts the user during the attack wherein he can act accordingly. The method also provides a mechanism to monitor fraudulent URLs and thus combat three types of phishing attacks simultaneously. The approach uses textual features of a webpage to recognize a phishing attack and is able to capture visually similar or dissimilar phishing targets as it is considering the resemblance score of the webpage features for classifying the current page as fake or authentic. As shown in Figure 5.2, the proposed framework consists of four operational agents when a webpage is opened in a browser tab. The agents in this system are T-agent , U-agent, M-agent and I-agent.

The autonomous agents arranged in three levels continuously monitor the presence of attack at regular intervals in multiple tabs. The agents are hierarchically organized with the possibility to share and delegate activities and/or responsibilities. The operation of the proposed system proceeds in two main phases.

- Feature Extraction
- Feature Comparison

The workflow of the proposed system is illustrated in Figure 4.1 and is summarized as follows. The major functionality of the agent based system is performed by level 1 agents (T-agent and U-agent). When webpages are loaded in the browser tabs, T-agents in each tab perform Feature Extraction and Feature Comparison every 60 seconds. Feature extraction is a quantitative way of capturing 5 tuple information from a website such as text, image, favicon, title and URL. The feature extraction value of tuples during the first pass are stored as the expected value. Further a resemblance score is calculated for each tuple by comparing the subsequent feature extractions along with the expected value. Using these scores, a single resemblance score is derived that captures the similarity between the stored version and currently opened version of a webpage. Based on the resemblance score, webpages are classified as legitimate or phishing according to a threshold *t*.

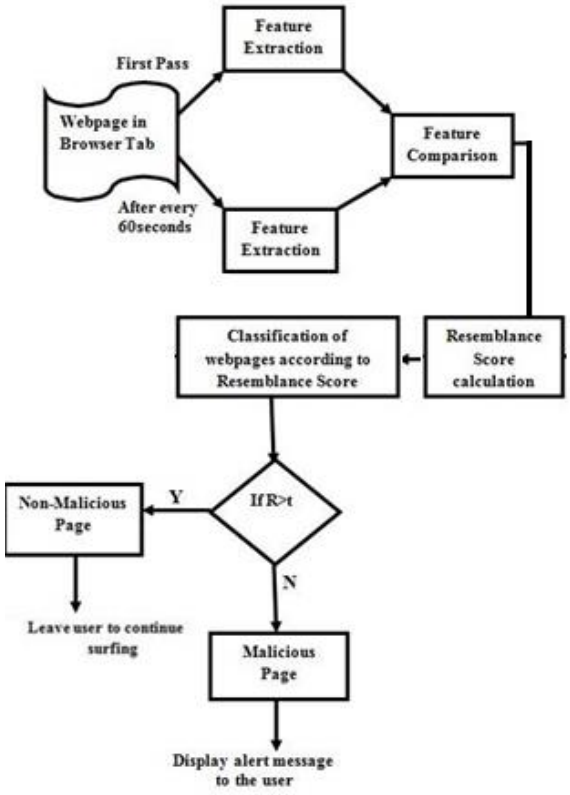


Fig. 4.1. Operational workflow of the proposed method

V. IMPLEMENTATION

The implementation of the proposed method uses JADE software framework (JADE, 2016) in java platform. The experiments are performed using Core i3 @2.20 GHz processor, 4GB of RAM memory, JDK 1.8 and JADE 4.2.0 in Windows 7 platform.

The JADE application consists of a set of agents with unique names and IDs. Agents execute tasks and communicate and cooperate with other agents by exchanging messages and beliefs. Agents run on a platform provide various services, such as message delivery, agent migration and resource management. A platform is composed of one or more containers, which can host multiple agents.

Google chrome was selected as the browser as it is vulnerable to modern type of attacks. The method currently has a simple user interface, displaying an alert message to the user if a webpage is deemed as phishing.

ISBN -978-93-5406-047-02

VI. EXPERIMENTAL EVALUATION

The positive data set consists of a set of common webpages with login forms such as money transaction sites, banking sites, web mail clients, credit cards, social networking sites etc. as tabnabbing targets webages which can provide confidential information of users. The approach used 2000 unique webpages with login forms from different sources.

To identify login forms, this method has used a heuristic based algorithm using HTML DOM. A login form is the typical sign of any webpage used by phishers which is characterized by FORM tags, INPUT tags and login keywords. INPUT fields are provided to enter user input and the login keywords give an appeal that the user is interacting with a login form. The method verifies that a webpage has any login form and then it is added to the dataset. This initial screening helps to avoid unnecessary method execution in webpages without having login forms. This approach adapted a login form finder implemented in CANTINA+ (Xiang *et al*., 2011). The proposed method has gathered 34 login and search keywords which can reveal its type.

For negative dataset,9 tabnabbing pages which are the fake versions of gmail, facebook, twitter, eBay, flipkart, hotmail, paypal, sbi and bradesco). To make a list of blacklisted URLs, a collection of real phishing sites from phishtank [23]are taken. By simulating the attack, the relevant features from the webpages opened in various tabs are captured and recorded. The feature extractions conducted further in every 60 seconds use this recorded value for comparison phase. The output of feature comparison is a resemblance score of the original webpage with its currently opened version. This process is continued with all the webpages in the dataset.

In order to separate legitimate and phishing pages, the resemblance score set is partitioned according to a threshold value *t*. In this framework, the value of *t* is set to 4 to get an accurate result. If resemblance score is greater than 4, the webpage is considered as genuine, otherwise as phishing and an alert message is displayed to the user.

The effectiveness of the method is assessed using False Positive Rate (*FPR*) and False Negative Rate (*FNR*).FPR and FNR for various values of the threshold *t* are computed which is shown in Figure. They are calculated using the formulas given below:

$$FPR = \frac{FP}{(FP+TN)} \quad \ldots\ldots (6.1)$$

$$FNR = \frac{FN}{(FN+TP)} \quad \ldots\ldots (6.2)$$

where FP is false positives,FN is false negatives,TP is true positives and TN is true negatives.

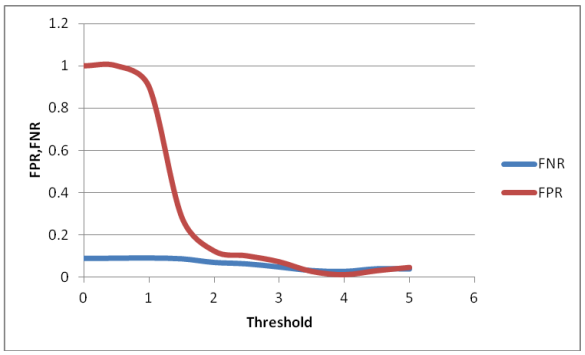


Fig 6.1 FPR and FNR in different thresholds

In figure 6.1, FPR and FNR is plotted with respect to varying threshold from 0 to 5. The results show that at threshold 4, the framework shows better result.

Figure 6.2 shows the percentage of false detections from negative dataset. It is seen that, the impersonated versions of email services (hotmail and gmail) give better result with no false detections. The percentage of false detections was mainly from paypal. This shows that the method could detect all the cases of tabnabbing launched using email services.

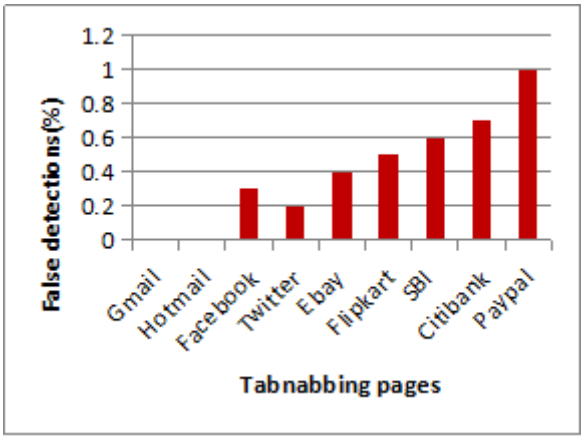


Fig 6.2 Number of False detections in negative dataset

A. Analysis of agent performance

In the proposed method, multiple agents are designed to operate in a complex environment to detect sophisticated phishing attacks. Moreover, the agents are equipped with computational behavior to perform tasks. Here, T-agent and U-agent are purely computational.

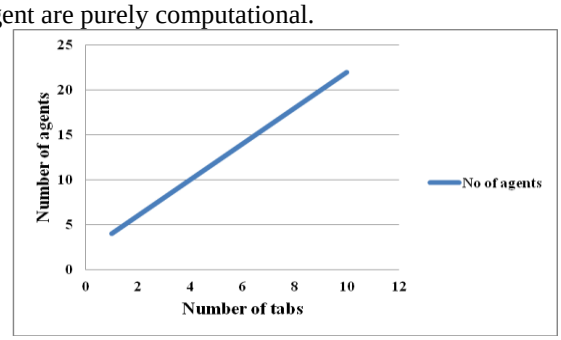


Fig 6.3 Number of agents vs Number of tabs

ISBN -978-93-5406-047-02

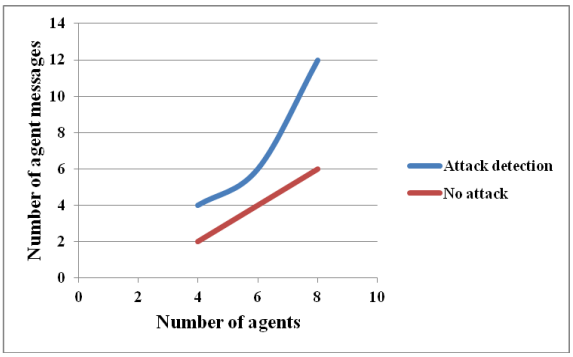


Fig 6.4 Number of messages sent vs Number of agents

There are four agents in action while a browser tab is open. Figure 6.3 shows graph with the number of agents active plotted with the number of opened browser tabs.

The architecture of the proposed MAS is hierarchical and distributed which significantly reduces communication cost and increases efficiency. The number of messages used for inter agent communication increases linearly with the number of agents but the number of messages sent between agents are minimized to reduce the communication overhead. Frequent communications are between level 1 agents.

Figure 6.4 shows the number of messages sent between agents in the case of attack detection and when no attack is detected. The number of sent messages will be more when an attack is found as there should be communication between each level of agents. When no attack is found, only the level 1 agents need to communicate for calculating the resemblance score of the currently opened webpage.

To evaluate the performance of the system in parallel attack recognition, multiple tabs are opened in parallel and ran the attack in each window. It has been noted that the method was able to detect attacks while running 10 browser tabs in parallel with good response time (in milliseconds). Beyond that, there is a slight decrease in efficiency as it may cause delay in the system.

Table 6.1 shows the time taken by the method in milliseconds to calculate the resemblance score of three categories of webpages.

Category I consists of simple webpages (only textual contents with no images), Category II consists of medium webpages (with text and images) and Category III consists of complex webpages(with more images). The results show that the proposed method consumes more time in image comparison than textual comparison. Beyond that, the execution time is also increasing linearly with increase in the number of tabs. Still then, the method is able to finish its execution with a maximum time of 1580msec which is less than page load time of a normal website. The proposed method has used the following websites in different categories for calculating the response time.

Category I	www.gmail.com
Category II	www.facebook.com
Category III	www.bradesco.com

Table 6.1. Response time of the proposed method according to number of tabs

Number of tabs	Response time in milliseconds		
	Simple webpage	Medium webpage	Complex webpage
1	285	538	730
2	285	550	760
3	290	590	820
4	300	630	870
5	316	690	916
6	330	760	1020
7	382	811	1092
8	402	890	1180
9	440	960	1320
10	480	1070	1580

VII. COMPARATIVE ANALYSIS

Table 3.1 provides a comparative analysis of the existing anti-tabnabbing solutions as well as the proposed agent based approach with respect to the following criteria. The various anti-tabnabbing techniques are compared in terms of the efficiency of techniques in multiple attack prevention, parallel attack recognition in multiple tabs of a browser etc. This table clearly shows that the proposed method stay unique from other methods and provides a novel solution to parallel attack recognition of multiple attacks in multiple tabs.

VIII. CONCLUSION

Browser attacks have become very common and are likely to succeed against systems that are not designed to resist them specifically. This paper provides an overview of some of the techniques used by attackers to deceive users using browser as a platform. Further, a discussion is made on a recent browser attack called Tabnabbing and the sequence of steps to execute it. Tabnabbing is a modern and more sophisticated kind of phishing attack. It is a phishing within a browser and it no longer relies on persuading user to click on a dodgy link. This targets a user who keeps many tabs open at a time. In addition, the paper has done a review on existing anti-tabnabbing solutions and a comparative analysis of the existing anti-tabnabbing techniques with the proposed method.

ISBN -978-93-5406-047-02

It has been seen that the proposed method offers parallelism in attack detection which is not supported by any other techniques. The proposed method is able to resist multiple phishing attacks with few resources in hand and provides user intervention while implementing the technique where other methods fail.

Table 7.1 Comparative analysis of existing anti-tabnabbing techniques with proposed method

Method	Technology in use	Browser	Use of Blacklist	Multiple attack prevention	Parallel attack recognition
NoTabnab(2010)	HTML layout	FireFox	X	X	X
Approach by Suri et al.(2011)	Signature based system	None	X	X	X
TabShots(2013)	Visual layout comparison by capturing screenshot	Chrome	X	X	X
TabSol(2014)	Hash value comparison	Chrome	X	X	X
Tabnabbing Detection System (TDS)(2015)	Profile creation and comparison using visual and structural features, User education	Chrome	X	X	X
TabsGuard(2015)	Textual layout comparison using HTML DOM	FireFox	√	√	X
TabSecure(2016)	Browser monitor	None	√	√	X
Proposed method	Structural features comparison, Multi Agent System	Chrome	√	√	√

REFERENCES

[1] Huang, L. S., Moshchuk, A., Wang, H. J., Schecter, S., & Jackson, C. (2012, August). Clickjacking: Attacks and Defenses. *USENIX Security Symposium*, pp. 413-428.

[2] Popa, R. A. (2016, February). *Server Side Security:Cross Site Scripting*. Retrieved February 2016, from http://www-inst.cs.berkeley.edu/~cs161/sp16/slides/2.8.XSS.pdf

[3] Yaakob, R., Joozdani, M., Abdullah, M. T., & Abdullah, A. (2013, July). Overview of cross site request forgery and client-side protection. *International Journal of Computer Technology and Applications*, *4*(4), pp. 706-709.

[4]Johns, M. (2011). Session Hijacking Attacks. In *Encyclopedia of Cryptography and Security*, pp. 1189-1190.

[5] Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015, June). Advanced social engineering attacks. *Journal of Information Security and applications*, *22*, pp. 113-122.

[6] Jagatic, T. N. (2007, October). Social Phishing. *Communications of the ACM*, pp. 94-100.

[7] Raskin, A. (2010). *Tabnabbing: A new type of phishing attack*. Retrieved from http://www.azarask.in/blog/post/a-new-type-of-phishing-attack/.

[8] Felten, E. W., & Schneider, M. A. (2000, November). Timing Attacks on Web Privacy. *7th ACM conference on Computer and communications security*, pp. 25-32.

[9] HTML meta tags. (2020, January). Retrieved from http://www.tutorialspoint.com/html/html-meta-tags.htm.

[10] HTML iframe Tag. (2020, January). Retrieved from W3Schools: http://www.w3schools.com/tags/tag-iframe.asp

[11] Mozilla Phishing protection. (2020, January). Retrieved from http://www.mozilla.com/en- US/firefox/phishingprotection/

[12] No Script. (2020, January). Retrieved from No Script-JavaScript/Java/Flash blocker for a safer Firefox experience: http://noscript.net/

[13] YesScript. (2020, January). Retrieved from YesScript Firefox Add on: https://addons.mozilla.org/enUS/firefox/addon/4922.

[14] Unlu, S. A., & Bicakci, K. (2010, October). NoTabNab: Protection Against The Tabnabbing Attack. *eCrime Researchers Summit (eCrime), IEEE*, pp. 1-5.

[15] Suri, R. K., Tomar, D. S., & Sahu, D. R. (2012, July). An Approach to Perceive Tabnabbing Attack. *International Journal of Scientific and Technology Research*, pp. 90-94.

[16] De Ryck, P., Nikiforakis, N., Desmet, L., & Joosen, W. (2013, May). TabShots: Client Side Detection of Tabnabbing Attacks. *8th ACM SIGSAC Symposium on Information*, pp. 447-456.

[17] Hashemi, H. F., Zulkernine, M., & Weldemariam, K. (2014, August). TabsGuard: A Hybrid Approach to Detect and Prevent Tabnabbing Attacks. In *Risks and Security of Internet and Systems*. *Springer International Publishing*. pp. 196-212.

[18] Singh, A., & Tripathy, S. (2014, December). TabSol: An efficient framework to defend Tabnabbing. *International Conference on IEEE, Information Tech- nology (ICIT)*, pp. 173-178.

[19] Al-Khamis, A. K., & Khalafallah, A. A. (2015, November). Secure Internet on Google Chrome: Client side anti-tabnabbing extension. *Anti-Cybercrime (ICACC), First International Conference on IEEE*, pp. 1-4.

[20] Joshi, P., & Chatterjee, M. (2016, June). TabSecure: An Anti-Phishing Solution with Protection against Tabnabbing. *International Journal of Com- puter Networks and Applications (IJCNA)*, *3*(3), pp. 63-69.

[21] Sarika, S., & Paul, V. (2017, April). Parallel Phishing Attack Recognition using Software Agents. *Journal of Intelligent & Fuzzy Systems*, *32*(5), pp. 3273-3284.

[22] Sarika, S., & Varghese, P. (2018). Intelligent Agents in Securing Internet. *Journal of Internet Technology*, *19*(3), 753-763.

[23] Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011, September). Cantina+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Transactions on Information and Sys- tem Security (TISSEC), 14*(2), pp. 21-43.

[24] PhishTank - Out of the Net, into the Tank. (2020, January). Retrieved from PhishTank: http://www.phishtank.com

[25] Sycara, K. P. (1998, June). Multiagent systems. *AI magazine*, *19*(2), pp. 79-92.

ISBN -978-93-5406-047-02

COURSES OFFERED

- B.Com Finance (2 Batches)
- B.Com Computer Application
- B.Com Co-operation
- BBA
- M.Com
- B.Sc Computer Science
- BCA
- M.Sc. Computer Science
- BA English Language and Literature
- B.Sc. Hotel Management and Catering Science (2 Batches)
- B.Sc. Hotel Management and Culinary Arts
- Craftsmanship Course in Catering Management

ISO 9001-2015 Certified

facebook.com/nimitpongam

twitter.com/nimitpongam

www.instagram.com/nimit_pongam/

ALMA CONNECT
https://naipunnya.almaconnect.com/

https://www.youtube.com/channel/UCEiCYfZpE0NRK6H-y5Lk6ig

NAIPUNNYA INSTITUTE OF MANAGEMENT AND INFORMATION TECHNOLOGY (NIMIT)

Pongam, Koratty East,Thrissur District, Kerala State - 680 308.
Ph : 9605001987, 0480 2730340, 2730341. Website: www.naipunnya.ac.in
Email - mail@naipunnya.ac.in