

A User Identification Technique to Access Big Data Using Cloud Services

Manu A R
IEEE member, CORI Lab,
ISE Dept., PESIT, VTU,
Bangalore, Karnataka, India

V K Agrawal, Ph.D
distinguished Senior Scientist
with ISRO,
Bangalore. Director, CORI
LAB, PESIT, VTU, Bangalore

K N Balasubramanya
Murthy
Vice chancellor and Principal
PES University, Bangalore

ABSTRACT

Authentication is required in stored database systems so that only authorized users can access the data and related cloud infrastructures. This paper proposes an authentication technique using multi-factor and multi-dimensional authentication system with multi-level security. The proposed technique is likely to be more robust as the probability of breaking the password is extremely low. This framework uses a multi-modal biometric approach and SMS to enforce additional security measures with the conventional Login/password system. The robustness of the technique is demonstrated mathematically using a statistical analysis. This work presents the authentication system using the consumer authentication architecture diagrams, activity diagrams, data flow diagrams, sequence diagrams, and algorithms.

Categories and Subject Descriptors

Multi-factor biometric password generation and authentication.

Additional Key Words and Phrases

Multi-dimensional, multi-level security systems, multiple privilege levels, multi-factor passwords, multi-modal biometrics, cloud computing, big data intelligence, security and privacy.

1. INTRODUCTION

As the computing system is growing end users are progressively relying on shared community computers to carry out their business over the Internet. With the growing speed of network connectivity, system demands secure computing. The user, in order to gain access to organizational assets, principally has to catalog and register with the system admin providing the details of the entity. In general, a password generated and stored in the password database server. Then, a unique code assigned later for that particular individual.

At present, various entities accessing web-based computing services conventionally use single-factor authentication, which requires typing a username and password to

authenticate. This system has substantial susceptibility since the password can be hacked by the shared public computing system, which may be reused by hostile parties to carry out malicious activities. There are many techniques found in literature to authenticate the entity [1- 10]. This technique is still facing various pitfalls. It is not competent enough to avoid the intruder from hacking the data [1-10]. Certain business transactions with high-risk big data consider single-factor authentication to be insufficient [11 - 13]. It involves access to customer confidential information or the movement of data to other entities using shared network resources. Therefore, we need a better technique to reduce the probability of breaking the password. Multi-factor and multi-dimensional authentication techniques can be used for such an application [10 - 13].

The different types of authentication systems currently in practice include three broad canonical categories that are Knowledge-based (e.g. use of passwords), Tokens-based (e.g. a smart card, OTP) and Biometric authentication. Among these authentication techniques, the biometrics technique is based on the biological/genetic characteristics and/or behavioral traits that are known to be innate to a particular user [1-13].

Manu A R, IEEE member, is with CORI Lab, ISE Dept., PESIT, research Scholar, VTU, Bangalore, Karnataka, India. (email: manu.a.ravi@gmail.com, manu_ar@nitk.ac.in)

V K Agrawal Ph.D, Senior IEEE member, was distinguished Senior Scientist with ISRO, Bangalore. He is now with Director, CORI LAB, PESIT, VTU, Bangalore, Karnataka, India. (email:vk.agrawal@pes.edu)

K N Balasubramanya Murthy Ph.D, was Principal and Director in, PESIT, VTU. He is now Vice-Chancellor PES- University, Bangalore, Karnataka, India. (email:principal@pes.edu)

Multi-factor Authentication [11- 13]

Authentication can take place based on a combination of multiple factors of various types/modes. User validation with

verification, authentication and recognition based on – a) **something the user is acquainted** with - like passwords, PIN, name of a person, phone number, date of birth, his residential address, and his own facts. b) **Something the user has** e.g.: chip card/smart card-based ID/access cards, tokens, driving license, PKI certificates, virtual private network PIN, embedded security token card that produce one-time passwords (OTP) pins etc.). c) The physiological traits of the **client** like - physiological and behavioral characteristics – fingerprints, multi-modal biometric input, hand geometry and or d) **something that the user acts/does** - performs/operates/executes that constitute his behavioral aspects.

Multi-dimensional Scaling password generation system [1-13] involves three basic steps. They are: a) Identification of the dimensions of the perceptual space on which customers perceive objects and input to generate complex password. b) Input the respondents judged values of perception or preference of the objects to generate multi-dimensional password. c) Multiplexing the objects in the perceptual space. The output of n-dimensional scaling consists of multi-factor input with location of the objects/entities details contained in it. Here, user inputs multi-factor, multi-modal, multi-level, multi-layered information to the system using multi-lateral policy framework. At each consecutive level, the user multi-factor input combined with multi-dimensional output of the system and a unique password with PIN generated and sent to the customer via SMS or e-mails.

Cloud users with valid privileges are grouped as normal privileged users, super privileged users and special privileged users. These users are authenticated at many levels. Initially, cloud vendors organization level password authentication/generation is at level 1. If hostile organization/party attempts to hack the data using cloud services they are filtered at the initial stage. Level 2 is at the organizational level cloud user group authentication or the group's individual team password generation/authentication. This allows users to perform computation on big data using particular cloud services. At level 3 is cloud user group privilege authentication where user supplies his details for password generation with permitted privileges and levels of access to generate user privilege authentication. In this way, authentication system can include the fourth and the fifth levels etc. depending upon the hierarchy of the organization.

In this algorithm at level 4, we can generate final multi-dimensional password with multi-factor, multi-modal

One simple approach is to use complex authentication method with combinations using multi-factor and multi-dimensional password with multi-level authentication. This is done so that the probability of breaking such secret codes is reduced largely. This existent menace provoked our thoughts to propose this tailored authentication technique in securing big data computation for ensuring strict authentication. This technique may help large cloud user groups to perform computation on big data.

In this paper, Section 2 describes the algorithm of the proposed authentication system. Section 3 discusses the probability of breaking multi-level, multi-factor and multi-dimensional authentication systems. Section 4 gives the architectural and data flow diagrams of the proposed system. Section 5 presents the experimental details. Section 6 concludes the paper with possible future work of the proposed system.

2. PROPOSED ALGORITHM

The proposed algorithm is based on multi-factor user inputs to generate multi-dimensional password at multiple levels. In this algorithm, we have taken multiple factors for authentication at each level and combined them with different levels of authentications. The proposed algorithm is performed at six levels. Functions at each level are described here briefly.

biometric information. This authentication can be done by combining service level authenticated password, group level authentication and privilege level authentication along with the details of company logos, user personal details including signature, location, time and zonal details, virtual host ids, network ids, IP details and other details supplied by the user. In level 5, it applies appropriate encryption algorithm for image and text details acquired or input by the user. These input details are combined and the PIN is generated with multi-dimensional password and it is sent to the user's registered contact number through SMS or mailed to him through a defined mode of communication.

Finally, in level 6, we generate the CAPTCHA image (an alpha numeric image) and send it to user in order to prevent the automated spam. An SMS is sent to the user along with PIN. Once the user enters the PIN, multi-dimensional password and CAPTCHA image to the system, it authorizes and provides access to the privileged cloud service.

The algorithm is described Fig. 1.

Main Algorithm

Level 1:

Step 1: Cloud-Service-authentication - Organization to generate multi-dimensional password to perform computation on big data and allied Service.

Read organization service/unique parameter details and input multi-factor values.

Generate service authentication password (Algorithm A)

If cloud-Service type authentication is authenticated with organization then,

Go to step 2

Else

Go to step 7

End

Level 2:

Step 2: Organizational Cloud User Group authentication - group individual team multi-dimensional password generation.

Read organizational, team/group details, inputs multi-factor, multi-modal biometric and group ID's, host ID's, IP details, time and zonal location inputs and parameter details.

Generate Cloud User Group Authentication (Algorithm B)

If Cloud User Group Authentication is authenticated then,

Go to step 3.

Else

Go to step 7.

End

Level 3:

Step 3: User Privilege Authentication– User group/individual password generation.

Read User details for password generation with permitted privileges and levels of access.

Generate User Privilege Authentication (Algorithm C)

If Privilege Authentication is authenticated to the user then,

Go to step 4

Else

Go to step 6

End

Level 4:

Step 4: Generate Final multi-dimensional password with multi-factor, multi-modal biometric input.

Password = Concatenate Service Authenticate Password, Group Authentication and Privilege Authentication, and company image, signature and logos, user input details, location and zonal details, virtual host ids', network ids', IP details and other details supplied by the user.

If password with user and service privileged mode is authenticated then,

Go to step 5

Else

Go to step 6

Level 5:

Step 5: Apply appropriate encryption algorithm for images acquired/input and user input details combined and PIN and password generated, and is sent/mailed to user registered contact number or mail id thru SMS or mail-id or through defined mode of communication.

Level 6:

Step6: SMS the pin to user. Generate the CAPTCHA image, mail the same to the user and, once user enters the pin and CAPTCHA image, allow the user to access the privileged cloud service.

Step 7: End.

Algorithm A: Generate _ Cloud Service Authenticate (), to perform computation on big data and allied service
Begin
Step 1: Read input values
 Read input values of corporation names, business logos and other organizational details, geo - location and zonal details.
 Read corporation signature of and company ID, service ID, type and details, text input by the user
Step 2: Add images
Step 3: Convert into 3D/ N-Dimensional (preferred dimension by the user/organization) images
Step 4: Concatenate 3D/ N-Dimensional (preferred dimension by the user/organization) image with textual inputs in a predefined sequence
Exit ()

Algorithm B: Generate _Cloud User Group Authentication ()
Begin
Step 1: Read input values
 Read team image, with multi-modal biometric information-digital DNA sample input, thumb sign, audio video acoustics, iris scan input, physiological movements, digital signature, and user multiple text input.
 Read team name, id with host, system, IP, network ids'
Step 2: Add images
Step 3: Convert into 3D/ N-Dimensional (preferred dimension by the user/organization) images
Step 4: Concatenate 3D/ N-Dimensional (preferred dimension by the user/organization) image with textual inputs
Exit ()

Algorithm C: Generate _ Privilege Authentication () to perform computations on the big data and access allied services
Begin
Step 1: Read input values
 Read user name, id and DOB and other details, user input multiple personal detail texts
Step 2: Concatenate texts
Step 3: Concatenate with cloud service Authenticate and Group Authentication, and predefined privilege level
Exit ()

Fig. 1 The proposed algorithm for robust authentication

3. PROBABILITY OF BREAKING MULTI-LEVEL, MULTI-FACTOR AUTHENTICATION SYSTEM

In this section, we present the formal analysis of probability of breaking the authentication system.

Considering *the events at each level as independent of each other*, we get the probability of breaking the authentication system, as

$$pL = p_1 * p_2 * p_3 * p_4 * \dots * p_n \dots \dots \dots (1)$$

Where p_n is the probability at level n.

Or
$$pL = \prod_{i=1}^n p_i \dots \dots \dots (2)$$

Considering $p_1 = p_2 = p_3 = p_4 = \dots = p_n = p$, then

$$pL = p^n \dots \dots \dots (3)$$

Further, at each level if we have multiple factors and they are independent events, then the probability of breaking m inputs at i^{th} level is given by:

$$p_i = \prod_{j=1}^{m_i} p_{ij} \dots \dots \dots (4)$$

Where, m_i is number of multiple factors at i^{th} level and p_{ij} is the probability of breaking m inputs at i^{th} level with j^{th} event. Then total probability of breaking the password of n levels authentication system is given by:

$$pT = \prod_{i=1}^n p_i$$

$$pT = \prod_{i=1}^n \prod_{j=1}^{m_i} p_{ij} \dots \dots \dots (5)$$

If we assume for a specific case where $p_{ij} = p, \forall (i \in n \text{ and } j \in m_i \text{ and } m_i = m \forall i \in \{1, 2, \dots, n\})$, then

$$pT = (p^m)^n \dots \dots \dots (6)$$

For illustration, $p = 0.1, m = 5, n = 10$, then $p^T = (0.1)^{50} \cong 1 * e^{-50}$. This shows that for a reasonable value of m and n the probability of breaking the password is very low. Figure 10 shows the variation in p^T for different values of $m = 1$ to 5, $n = 1$ to 10 levels and $p = 0.5$. Table 1 shows the quantitative value of the probability of breaking the password, at various levels. The pT is also plotted in Fig. 2.

Table 1
The Value Of The Probability Of Breaking The Password, At N Levels

N (No of levels)	m= 1	m= 2	m= 3	m= 4	m= 5	
1	0.5	0.25	0.125	0.0625	0.03125	2
	0.25	0.0625	0.015625	0.003906	0.000977	3
	0.125	0.015625	0.001953	0.000244	3.05E-05	4
	0.0625	0.003906	0.000244	1.53E-05	9.54E-07	5
	0.03125	0.000977	3.05E-05	9.54E-07	2.98E-08	6
	0.015625	0.000244	3.81E-06	5.96E-08	9.31E-10	7
	0.007813	6.1E-05	4.77E-07	3.73E-09	2.91E-11	8
	0.003906	1.53E-05	5.96E-08	2.33E-10	9.09E-13	9
	0.001953	3.81E-06	7.45E-09	1.46E-11	2.84E-14	10
	0.000977	9.54E-07	9.31E-10	9.09E-13	8.88E-16	

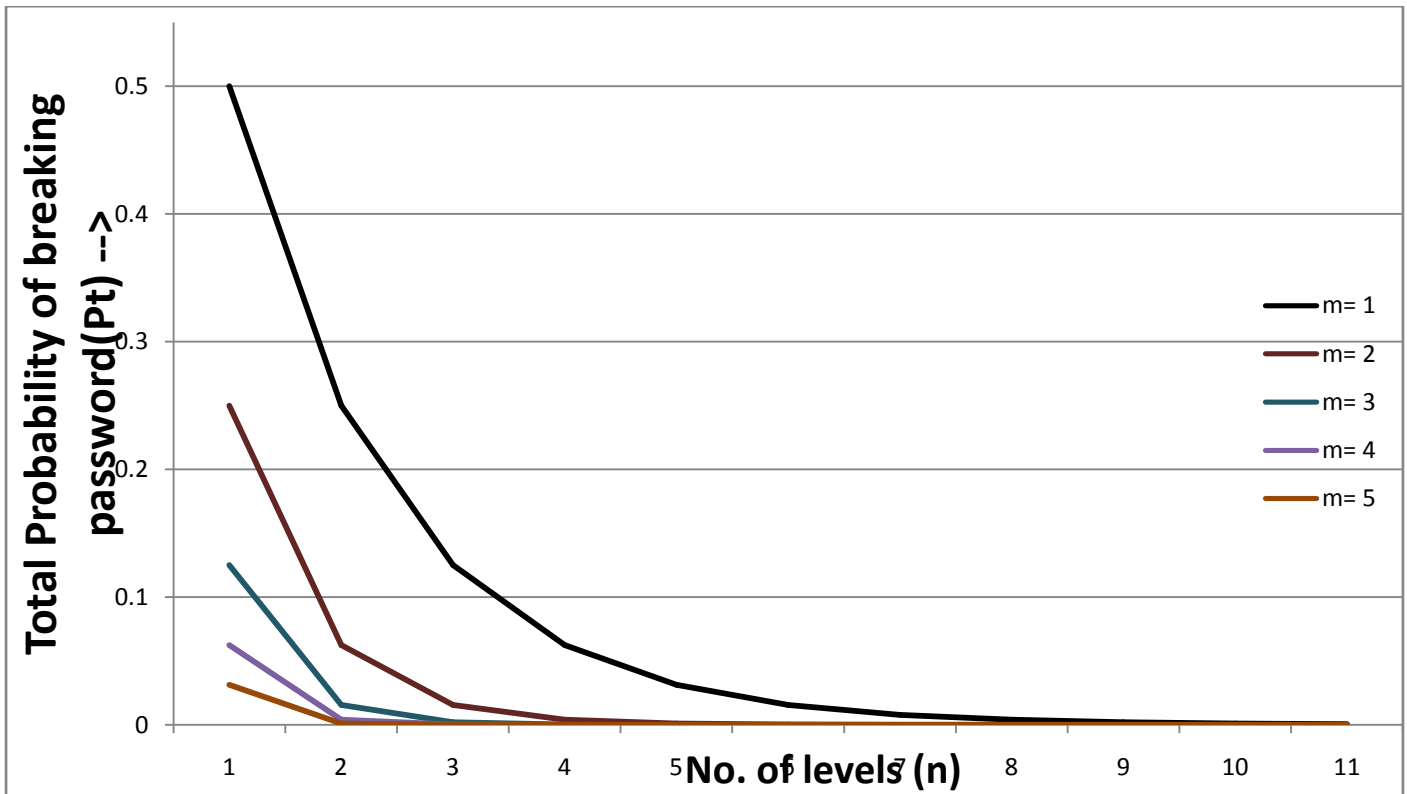


Fig. 2, Total probability (p^T) versus Number of levels (n) with m values

From the above discussion we find that multi-level multifactor password provides robust design against password hacking.

4. AUTHENTICATION SYSTEM ARCHITECTURE DIAGRAMS

DFD Level 0 shown in Fig. 3 depicts the overview of multi-dimensional and multi-level password generation system. Fig. 4a and Fig. 4b shows context level diagram password authentication system using centralized sever model to perform computation on big data. Fig. 5 shows Flow of Server Module for tracking the hacker.

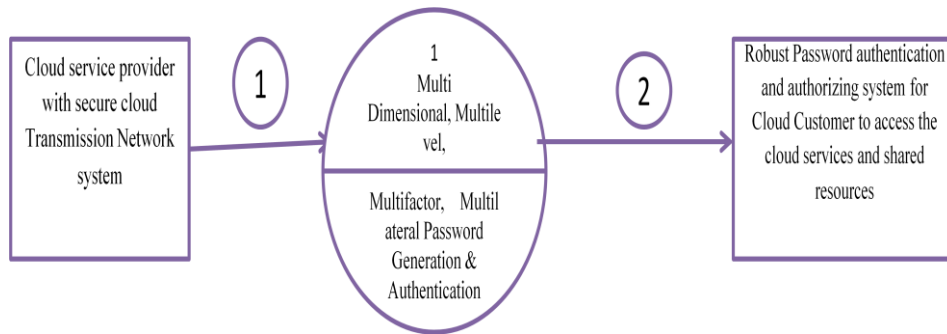


Fig. 3 Level 0 - Context Diagram

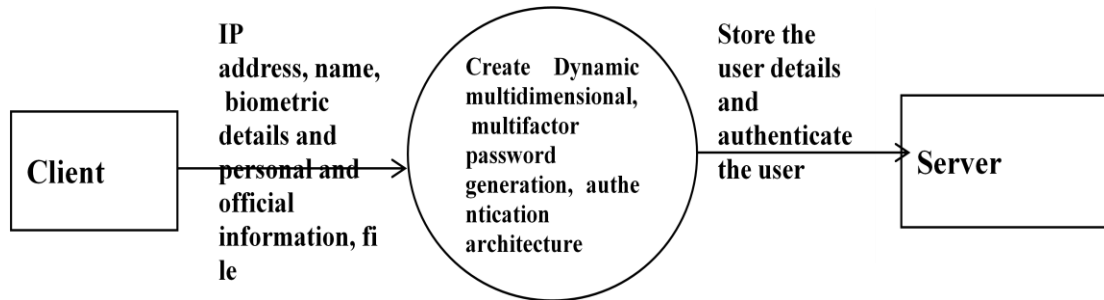


Fig. 4a: context level diagram password authentication system using centralized sever model.

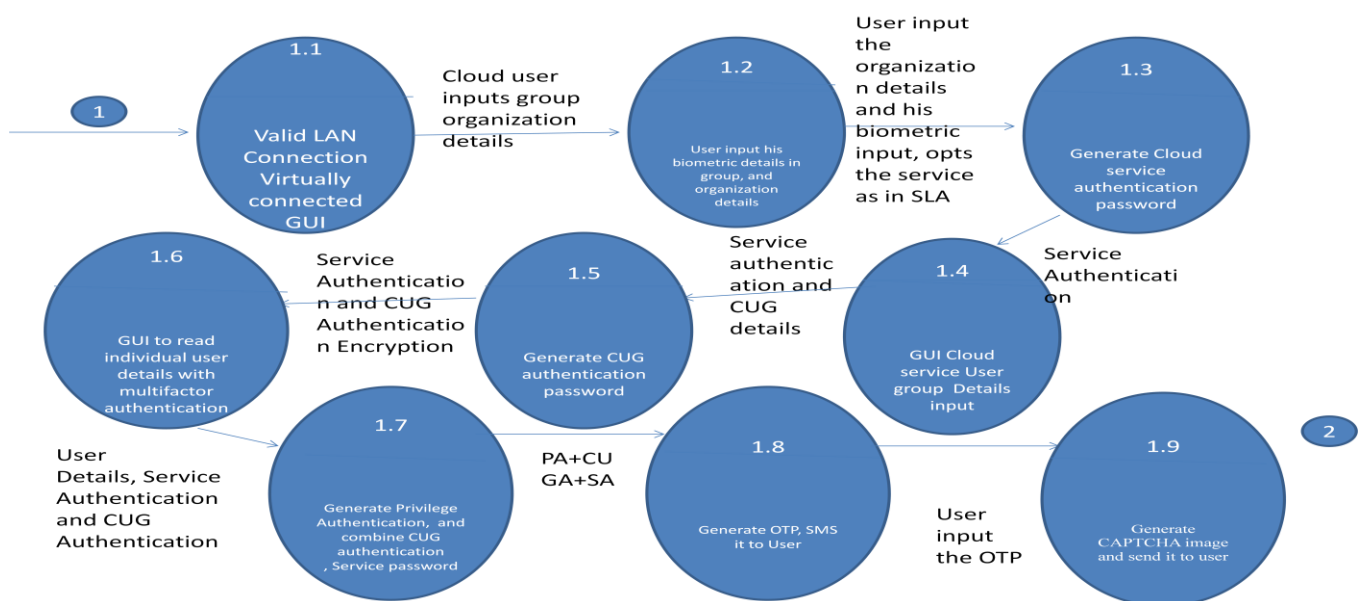


Fig. 4b: Dataflow diagram showing various stages of authentication system

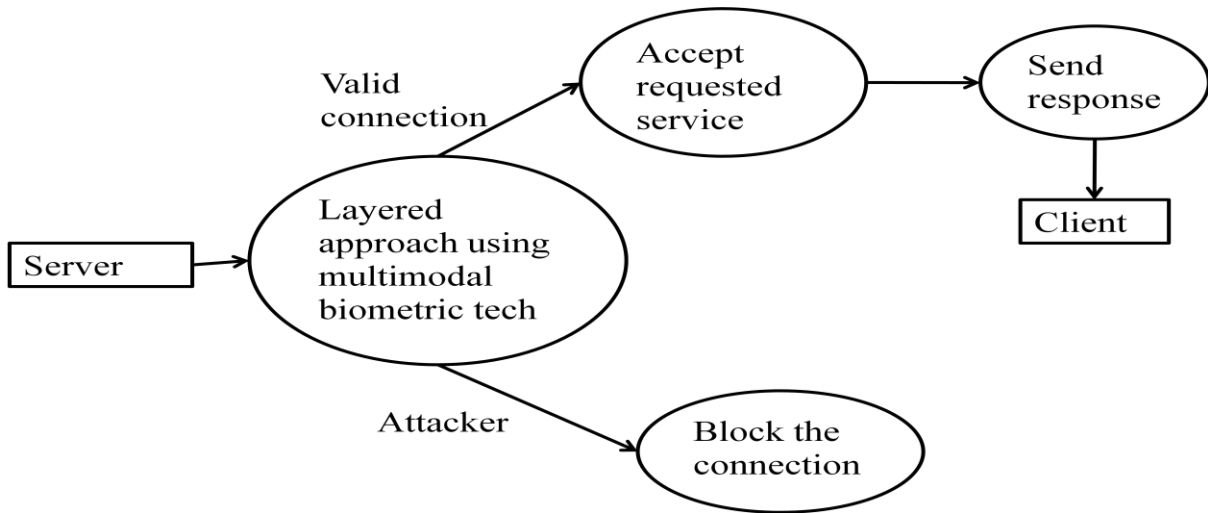


Fig. 5 Flow of Server Module for tracking the hacker

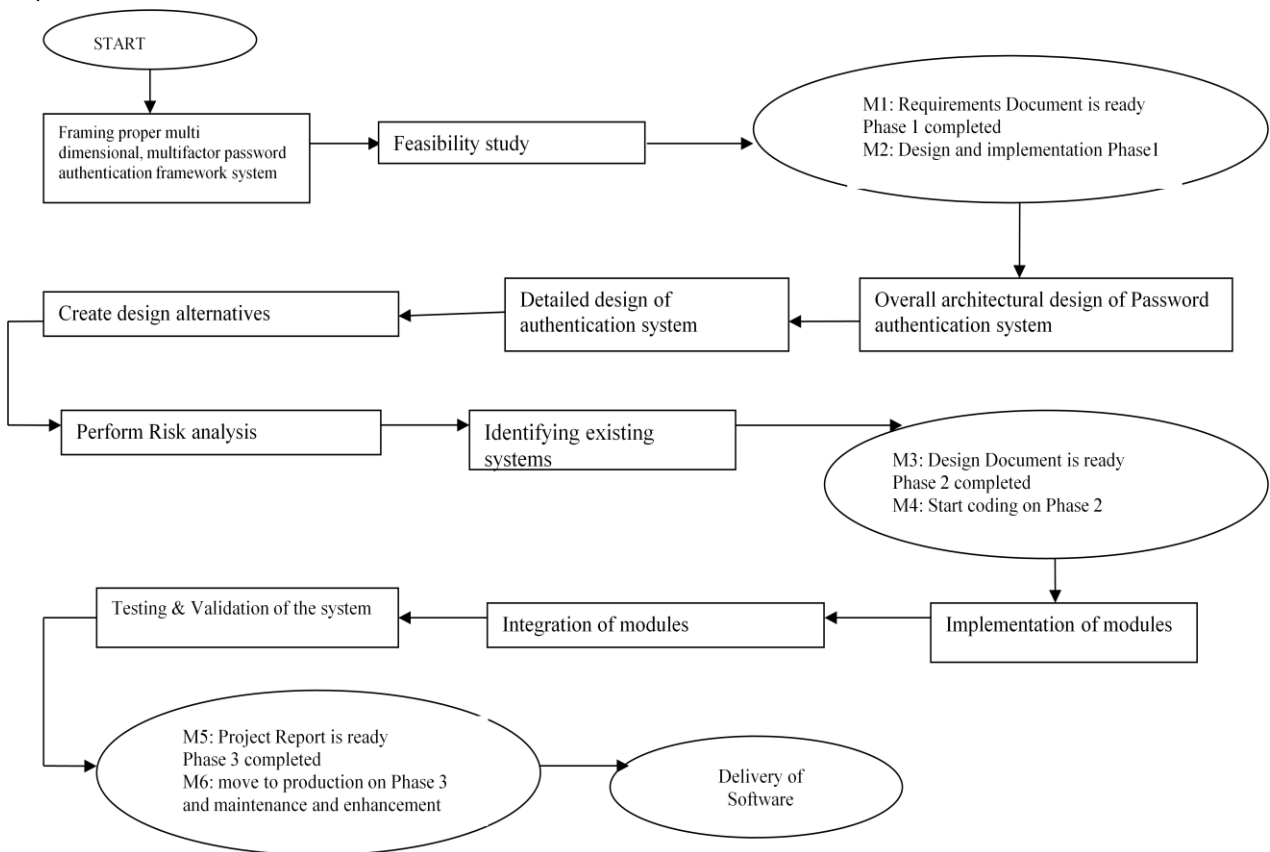


Fig. 6: Activity Chart

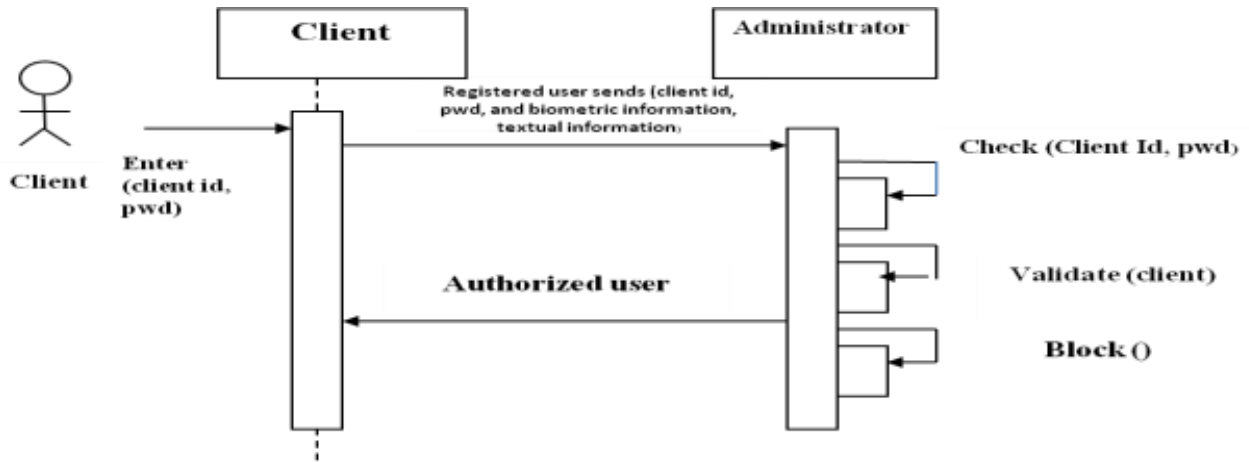


Fig. 7: Sequence diagram for client sending request for connection to administrator.

The activity chart gives us an overview of authentication system and gives details of how much time spent on each phase, milestones to be achieved and deliverables that are to be given. The activity chart is as

shown in Fig. 6, Fig. 7 shows sequence diagram for client sending request for connection to administrator. Fig. 8 shows the work flow diagram at module level.

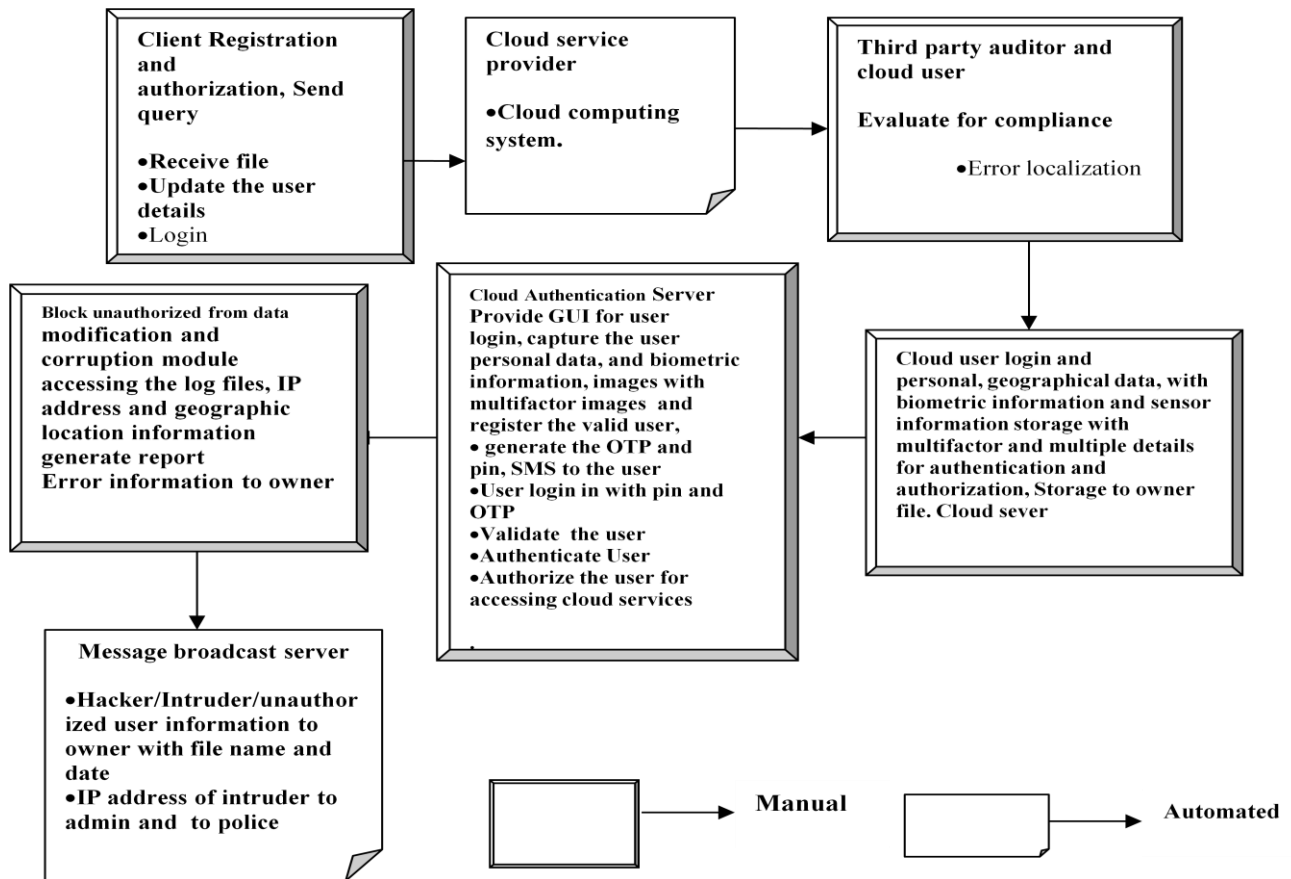


Fig 8 shows work flow diagram at module level.

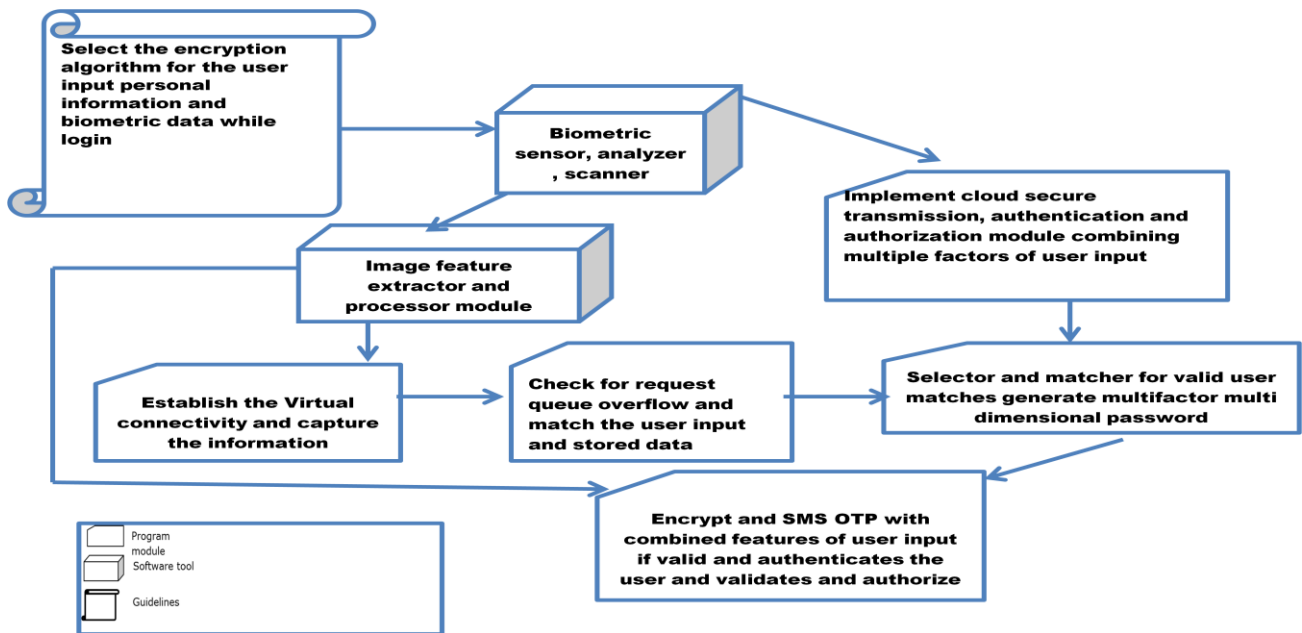


Fig. 9 Job flow diagram of cloud service authentication and authorization system

The job flow diagram of cloud service authentication and authorization system is, shown in Fig. 9 showing the expanded DFD of use case diagrams representing leveling diagram. The procedure of expanding a DFD is - called as leveling.

5. IMPLEMENTATION DETAILS

Strong password generation and authentication technique is used to perform computing on big data using cloud services. It utilizes multi-factor, multi-dimensional, multi-level, multi-sampling, multi-source, multi-mode, multi-layered user biometric data and his personal data for generating the password to identify the user. Multi-dimensional passwords get generated by considering many facets of input parameters of the cloud user and organizational data. Such as vendor details, consumer details, services ID, Security features with respect to network time protocol, VLAN features, track configuration service ID, event service ID, name space mapping, IP address, VLAN ID, device ID, interface ID, host name, signed privileges in SLA, logos, images, textual numerical ids, information, signatures, PIN and other biometric information of the user etc. It also enables periodic re-authentication specifying a time period with server based time and with interface id, which is connected virtually using allocated IP address of the server-client connected to the port.

We have to cluster the cloud services (entity cluster, business group, service cluster) using source and destination IP address to track and trail using time zone of the locality, geographic vicinity from where the cloud service is accessed via the mapping system and GPRS technology to supervise the communications. We can also use VLANS, elapsed communication time during accessibility to validate and track the user activities. Re-authentication of users for particular time slices in round robin clock schedule is used. Maintaining counters for login and access operation for security and identification can be done as it helps in automatic discovery of unauthorized access. This is - achieved by assigning a specific privilege level with associated rights and privileges to each

user name and password pair. Sensors are - used to check the sys-log communication for violation counters and they block threshold level when the limit reached.

Biometric sources of authentication are neither prone to being misplaced by the user due to his absentmindedness nor are they reproducible. The resources include physiologically active (sweat, genetic DNA print, fingerprint, shape of the face, facial thermo grams, vein patterns, hand geometry, iris, retina scan etc. via genetic biometric framework) and/or behavioral characteristics, passive biometrics of persons (dynamics of handwriting – signatures, voice prints-acoustics, gait) [6] using artificial intelligence and fuzzy neural network-based classification and analysis of results for authentication. Biometric framework uses biometric data captured using sensors like charged couple device (CCD) cameras, I-R cameras, fingerprint scanners and microphones, readers, genetic DNA sampling detectors etc. The proposed methodology uses digital-signal processing techniques, mathematical simulations, artificial intelligence, image processing and recognition, pattern recognition, natural language processing, digital signal processing and data mining algorithms which contribute to the building blocks of extractors.

The working system uses sensors to extort, differentiate and decide user information resulting in producing its own type of feature vectors called templates that are - used as generic setting standard classifiers. These classifiers are used to recognize thus produced vectors incorporating the pattern-matching algorithm. The classifier chart is a vector which is fitted into an associated identity with a certain degree of confidence viz. a score confidence measure. This could be a scalar value or a vector with claimed class label with several classifiers. The supervisor merges different scores to obtain final decision if the match is found. Then the system accepts this identity claim. If there is a mismatch then the system rejects the identity claim as shown in the below framework design in Fig. 10 and 11.

Fig. 10 and Fig.11 represent the framework of cloud - password authentication system at various stages. To get admittance to assets the customer ought to go by an objective screening course of action. Genetic DNA scanners will scrutinize a sole/unique segment of user's DNA/RNA helix structure in order to authenticate that user as is claimed by the user [7 8 9]. These days many laptop producers are including fingerprint and DNA scanners integrated on the current model of their machines. Some manufacturers have also in-built the scanner and sensor into their mouse. The scanner catches a snap of customer's fingerprint and compares it to the image file stored in the database and this provides an additional level of defense coupled with another form of multi-dimensional and multi-factor authentication [9, 10]. Certificate listener handler authenticates and initializes the registration of user along with key management certificate look up, certificate parsing, validation, proof material-mapping authentication, authorization, attributes of user's common name with surname, locality name, state, province, organization, organizational unit name, country name, email name ensure password strength, protection, detect delay in misusing,

establish trust, simple and composite delegation security authentication status.

Multi-factor web authentication systems rely on hardware-based security tokens that generate pass codes that are valid for about 60 seconds and ought to be input along with a password. Even if a malevolent party obtains a user's password, he or she would not be able to provide the relevant second element needed to complete the authentication process. This is costly, so instead, most websites ask users to undergo a one-time registration process during which users register one or more of their mobile devices with the website source. This is a trusted mechanism directly under the user's control which can get a verification code via SMS or another means to verify the user's identity. Any time a user signs into the website providing cloud service, a pass code/PIN is sent to the catalogued device. The user should enter the password and verification code/PIN to fully sign in and utilize the services. At any time, the client cannot obtain a text communication while travelling overseas. The solution might be an application for smart phones or tablet/laptops that can generate security codes on its own with simple steps to set up the app before starting the voyage.

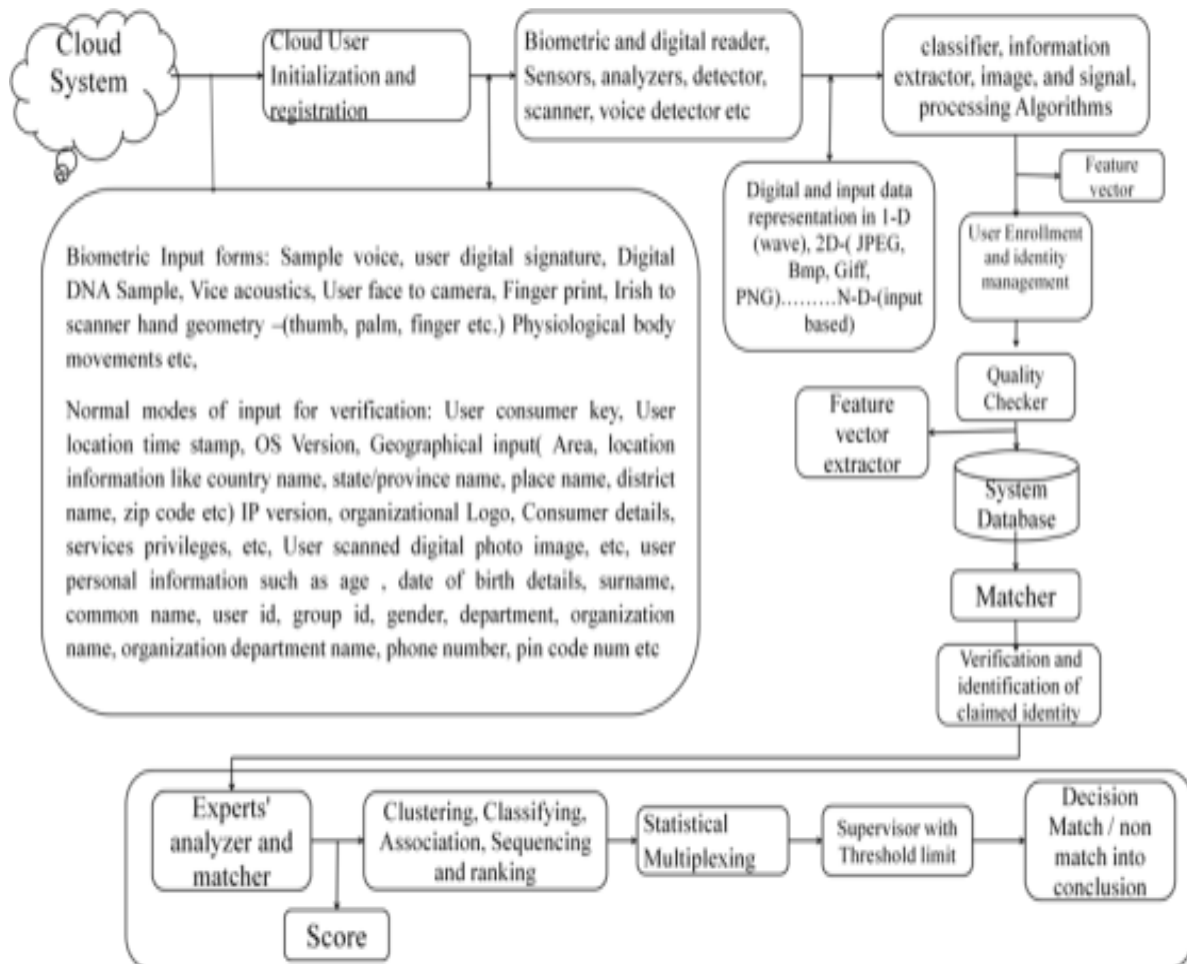


Fig. 10 Framework of cloud password authentication system at various levels

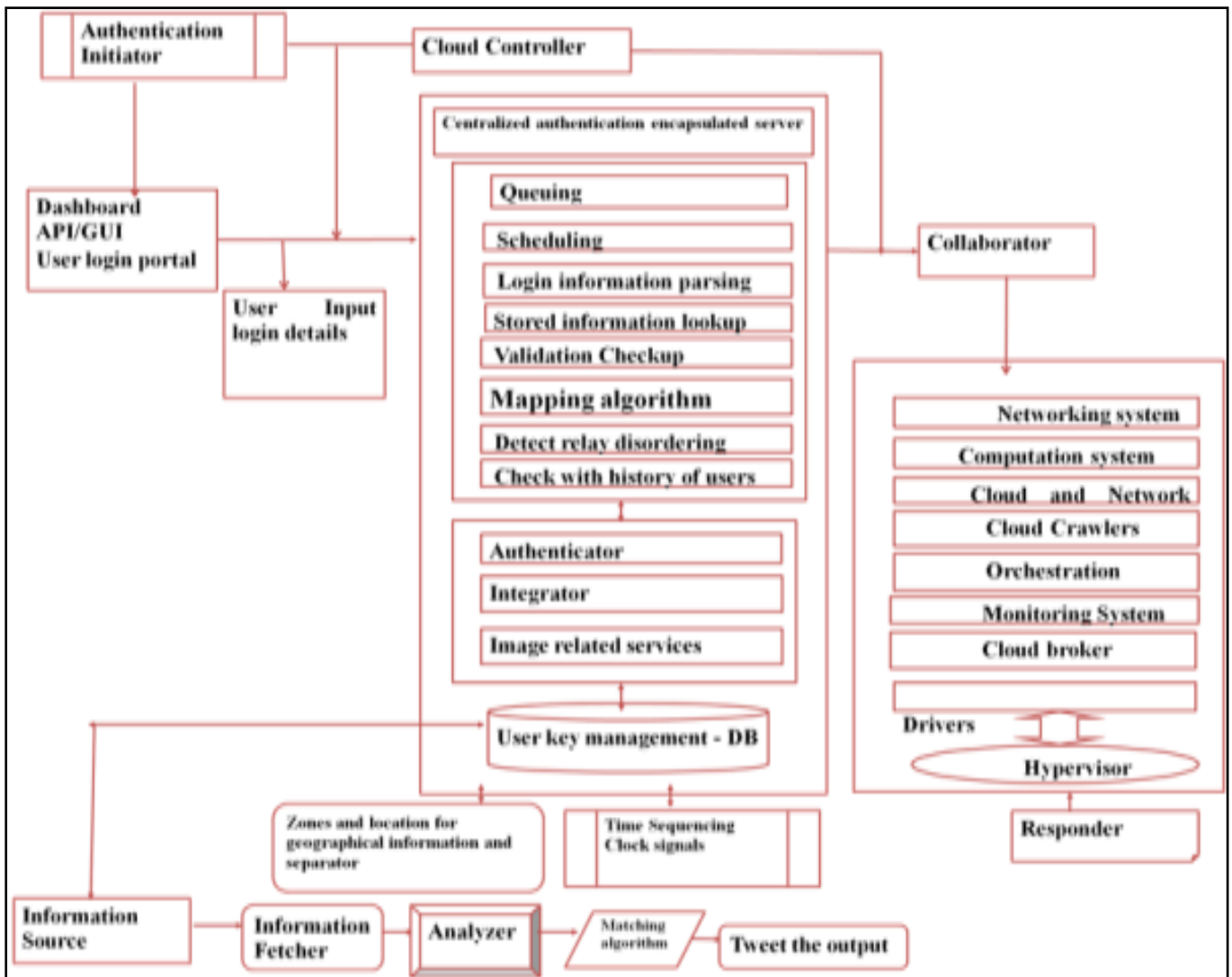


Fig. 11. Cloud architecture depicting multilevel, multifactor, authentication system

6. CONCLUSION AND FUTURE WORK

In this paper we have introduced a method of authentication and identification of end user in cloud environment. The methodology uses a multi-level, multi-dimensional, multi-lateral and multiple privileges, with multi-modal, multi-biometric and multi-factor approach of identifying the user in many levels with rank of organization so that strict verification, certification, validation, approval, and permission could be granted to access services to perform computation on big data. This practice helps in producing the password at various levels of institute so that stern authentication and authorization is achievable. Our future work will be carried out in adding intruder detection with strict access control method features to this technique.

ACKNOWLEDGEMENTS

Our, sincere thanks to **Dr. Shylaja S.S.**, HOD, Department of Information Science and Engineering, PESIT, Bangalore, and Management of PESIT, for their constant encouragement.

REFERENCES

- [1] Chris Roberts, "Biometric attack vectors and defences", computers & security, Elsevier, Science Direct, 26 (2007) 14 – 25, 2007 Elsevier Ltd.
- [2] "Biometric statistics in focus", Feature article in Science direct (Elsevier) 2010.
- [3] "The right biometric", survey, article in Science direct (Elsevier) Biometric Technology Today , March 2006.
- [4] M Y SIYAL, "A biometric based e-security system for internet-based applications", "<http://www.wseas.us-e-library/conferences/joint2002/451-208.pdf>" .
- [5] Koval, S. Voloshynovskiy, and T. Pun "Error exponent analysis of person identification based on fusion of dependent/independent modalities CUI-University of Geneva, Stochastic Image Processing Group, 24 rue General-Dufour, 1211 Geneva, Switzerland, "http://spie.org/samples/FirstPage/PSISDG/6819/68190P_sample.pdf", 2007.
- [6] Andreas Pfitzmann, "Biometrics – How to Put to Use and How Not at All?", Trust, Privacy and Security in Digital Business, Lecture Notes in Computer Science Volume 5185, 2008, pp 1-7 ,

- http://link.springer.com/chapter/10.1007%2F978-3-540-85735-8_1.
- [7] Anil K. Jain, Arun Ross and Sharath Pankanti, “*Biometrics: A Tool for Information Security*”, IEEE transactions on information forensics and security, VOL. 1, NO. 2, JUNE 2006.
- [8] Ayu Tiwari, Sudip Sanyal, Ajith Abraham, Svein Knapskog, Sugata Sanyal, “A Multifactor Security Protocol for Wireless payment secure web authentication using mobile devices”, ISBN: 978-972-8924-30-0 © 2007 IADIS, International Conference Applied Computing 2007, http://www.tifr.res.in/~sanyal/papers/Ayu_MultiFactorSecurityProtocol.pdf.
- [9] James Wayman, Anil Jain, Davide Maltoni and Dario Maio, “An Introduction to Biometric. Authentication Systems”, http://www.springer.com/cda/content/document/cda_downloaddocument/9781852335960-c1.pdf?SGWID=0-0-45-130566-p2275653.
- [10] Dinesha H A, Dr. V. K Agrawal, “Multi-dimensional Password Generation Technique for accessing cloud services”, Special Issue on: "Cloud Computing and Web Services", International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.3, June 2012, 31-39.
- [11] X. Suo, Y. Zhu, G. S. Owen, “Graphical passwords: A survey,” in Proc. 21st Annual Computer Security Application. Conf. Dec. 5–9, 2005, pp. 463–472.
- [12] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon, “Authentication using graphical passwords: Basic results,” in Proc. Human-Comput. Interaction Int., Las Vegas, NV, Jul. 25–27,2005.
- [13] Fawaz A. Alsulaiman and Abdulmoteleb El Saddik, “Three-Dimensional Password for More Secure Authentication”, Instrumentation and Measurement, IEEE Transactions , 03 April 2008, 57 , Issue:9, 1929 – 1938.