



A User Identity Management System for Cybercrime Control

T. Alese^{a,*}, O. Owolafe^b, A. F. Thompson^b, B. K. Alese^b

^aDepartment of Computer Science, Federal University of Technology, Akure, Ondo State, NIGERIA.

^bDepartment of Cyber Security, Federal University of Technology, Akure, Ondo State, NIGERIA.

Abstract

The rapid growth in the number of online services has led to an increasing number of different identities managed by each user that makes people feel overloaded and suffer from password fatigue. This poses a serious problem and makes people unable to control and protect their digital identities against identity theft. As organizations grow and adds services such as ecommerce and global remote access of services, controlling who is accessing what kind of information is also becoming a difficult task. This research therefore, presents the development of a user identity management system for cybercrime control. The four stages of an identity management life cycle were developed using some mathematical tools. A two-factor authentication technique was used in developing the system, the traditional username and password was also included with biometric features for robustness. A simulation was run on the model for users ranging from 10 to 1000 using life wild dataset, and accuracy was found to be 98.01%.

Keywords: cybercrime, user identity, PythonIDM, identity theft, two-factor authentication

1. INTRODUCTION

The development of Internet and the widened access to computer technology has created new opportunities for work and business activities. This technological advancement has produced radical shifts in the ability and economics of reproduction, distribution, control, and publishes of information [1]. The fast speed of information transmission across Internet and computer networks has made information dissemination almost instantaneous and affordable. Also, the rise in technological innovation and online communication has not only produced a dramatic increase in the incidence of criminal activities, but has also resulted in the emergence of somewhat a new variety of computer-related criminal activities. Thus, complexity of the increase in the incidence of criminal activities and the possible emergence of new varieties of cybercrimes pose challenges for legal system and law enforcement [2].

Cybercrime is an obstacle that may shut the door of progress against any business entity and the nation. According to [3], cybercrimes are described as one of the fastest growing criminal activities across the globe. It covers a large range of illegal activities and one of the most prevalent cybercrime is identity theft, which is a deliberate and fraudulent practice of using someone else's

“identity” (private or personal data and information) to gain a financial advantage and other benefits in an impostor's name [4].

In [5], it was stated that identity theft could be in form of criminal identity theft, financial identity theft and child identity theft. The term identity can be described as the set of permanent or long-lived temporal attributes associated with an entity [6] while [7], further defined an entity to be a physical or logical object which has a separate distinctive existence either in a physical or a logical sense. This identity is subject to cyber-attack and as digital data become more prevalent, users try to secure their information with encrypted passwords and ID cards. However, the misuse and theft of these security measures are also on the rise. Exploring the vulnerabilities in identity cards results in cards duplication, fake production and misuse. In [8], it was stated that Nigeria, Ghana and South Africa top the list of cybercrime in Africa. From the foregoing, the findings reflect the menace state vis-a-vis. There is a need to curb and proffer solution to cybercrime and consequently, user's identity management.

The world has gone digital and as a result it has altered many industries and change the way people use and enjoy the Internet. Consequently, technological development has become the spur for change today, and as in other technologically turbulent periods. Old methodologies and business models are gradually giving way to the development of new consumer-behavior and business models which are necessary features of market economies where there is consumer choice,

*Corresponding author (Tel: +234 (0)703 351 3174)

Email addresses: sharonalese@gmail.com (T. Alese), oiyare@futa.edu.ng (O. Owolafe), athompson@futa.edu.ng (A. F. Thompson), bkalese@futa.edu.ng (B. K. Alese)

transaction costs, and heterogeneity amongst consumers, producers and competitors. These technologies allow for enormous gains in efficiency, productivity, and communications, but there is strong evidence that access to such technology, with all its opportunities and benefits, can put businesses and families at increasing risk of exploitation and cybercrime. The vulnerability or possibility of exploitation by those who wish to take advantage of the technological innovations results to the subject of crime and security, which has become very critical. Nowadays, people are worried that an impostor or hacker is going to run up charges on their credit cards or fleece their bank accounts while their backs are turned. As pointed out by [9], all the thief needs to do is gain access to some vital information such as Social Security Number to become a real public enemy. For example, the banking sector is not immune to banking cybercrime. As the sector continues to grow, its growth is hampered by high cases of economic and financial crimes [10]. Fraud and money laundering have had adverse impacts on our national development and particularly on the financial system. The act have caused damage to the reputation and the image of the country, loss of FDI, poor infrastructural development, dwindling confidence and distortions in our political as well as financial systems, among other things. Although, cybercrimes are relatively easy to commit, investigating and prosecuting them are complex and time consuming. For this reason, attention is shifted to the study of user identity management system. There are currently a number of different incompatible Identity Management (IdM) solutions that resulted from separate isolated projects. The lack of cohesion and co-operation among these efforts lead to different notions and semantics for many central topics of IdM, and some fundamental concepts of IdM such as Identifiers. Such differences in these core topics introduce inconsistencies and such inconsistencies in turn introduce confusion, making it difficult to gain an understanding of IdM. This necessitated the present study on management of users' identity to control cybercrime [11].

2. REVIEW OF RELATED WORKS

Authors in [12] worked on Cyber Crime Detection and Control using the Cyber User Identification Model. The work proposes a paradigm shift from mere cyber attack detection and control, to the detection/identification of the cyber criminal, to be anonymous. Their research focused on the identification of cyber users on an enterprise network (Bank and Shopping Mall).

The authors in [7] worked on Identity Management using the Petname Systems. Concerns about the difficulty of managing identities on the user side lead to his research and he wanted to develop an aid to help users in managing service provider's identity. The unipet system in his work used the last two parts of a domain name to act as the Pointer. For example, if the domain name is `www.google.com`, the Pointer of the UniPet is

`google.com`. However, if the domain name consists of different Top Level Domain (TLD), then other parts of the domain name also need to be part of the pointer. For example, if the last two parts of the domain name are `com.ng`, obviously it is not a unique pointer as there are many more websites that may contain these two parts.

The researcher in [13] carried out a research on "An analysis of identity theft: Motives, related frauds, techniques and prevention." The paper provides a review of the unique effective techniques for sustainable development of prevention methods that have been offered to people and businesses. He summarized the most effective ways for individuals and organizations to protect them against identity theft. He analyzed four major factors of identity theft and also examined the major outcomes. He discussed the various techniques used by thieves to attack individuals and organizations and provided prevention techniques to protect key data and information against identity theft.

According to him, it is not possible to eliminate cyber-crime from the cyber space but possible to check them, therefore, he proposed that defensive strategies should combine security awareness, training, technical control, and an effective information management strategy.

3. DESIGN

The proposed Identity Management (IdM) System consists of the following parties:

Client/user receives services from a service provider through provision of partial identity with the identifier and the associated credential from the connecting identity provider. A client can be in form of any entity but this project is based on the assumption that each client is a person/user. The user profile for this project will involve a set of user attributes that will be used for customizing the service and possibly restrict access to portions of the service. The control of access to the service provided depends on accuracy of user profile information, which assists in making appropriate policy decisions. An access control entails granting of access to particular resources, and the auditable enforcement of that policy.

Service Provider (SP): An organization that provides services to clients or other SPs. It could also be referred to as the Trusted Party. This work adopted the use of the notation "SP" to denote the set of service providers.

Identity Provider (P): An organization that provides digital identities to allow clients to receive services from a SP. In this work, notation "P" is used for the set of identity providers.

The life cycle of the proposed IdM system, is made up of four (4) phases; registration, identification/authentication, authorization, and de-registration.

(a) Registration Phase:

This is the first phase and it involves a user "e" who wants to access services provided

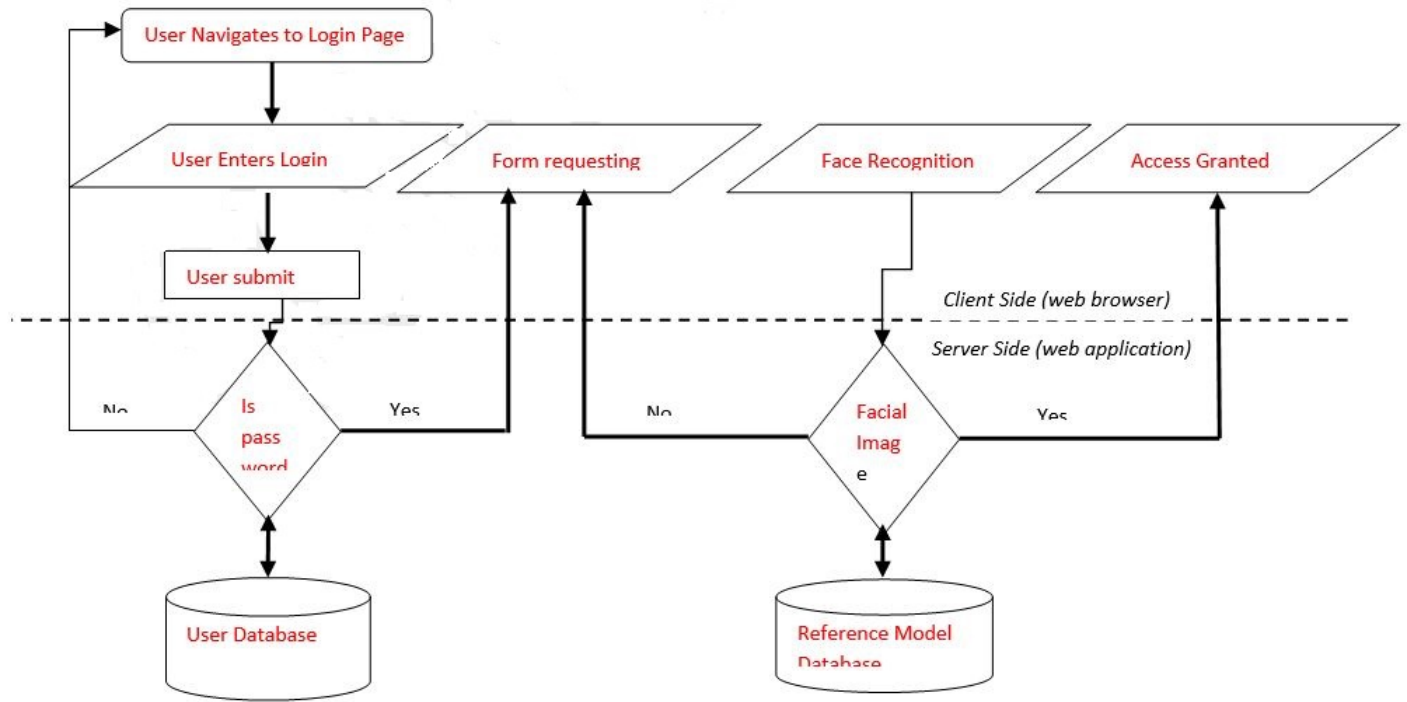


Figure 1: An overview of the structure of the Identity management system.

by service provider. It entails user registration at the respective identity provider “P” through provision of personal information. At this phase, the identity provider either generates the corresponding values of the identifier and the credential automatically on user behalf or the user may select an uncommon (unique) value for the identifier of the service and a value for the related credential. The data (value) for partial identifiers may also be provided by the user. The end of this process witness the update in the “P” through creation or insertion of new data in the set of entities and attribute values.

The process is modeled as shown in Eqs. (1) and (2):

$$E'_p = \{e\} \cup E_p \tag{1}$$

$$V'_p = \{e\} \cup V_p \tag{2}$$

where e represents a newly registered entity, E_p denotes an existing set of registered entities, E'_p represents the updated set of entities, that results from the addition of e , to E_p for the context p , V_p stands for the set of attribute values for E_p , V represents set of attribute values for e , V'_p denotes the updated set of attribute values (including the compulsory unique identifier value) formed by the addition of V to V_p for the context p .

The registration of the user will make addition of user information to become part of the registered domain

(b) **Identification /Authentication Phase:**

This second phase of the proposed system involves the recognition and confirmation of the registered user before accessing service. Identification is the process of finding an association between an identifier value and the user, while authentication is the process of proving the association. Therefore, authentication includes algorithm definition, which will be taking as input, the supply of user’s identifier and attribution of values. This algorithm then returns either a successful result if the entity (the user) can be identified or returns an unsuccessful result if otherwise.

$$Auth = \begin{cases} 1 & \text{if } f_{vx} \text{ matches } \{e\} \\ 0 & \text{if otherwise} \end{cases} \tag{3}$$

where vx represents an identifier that uniquely identifies a given entity; and Auth means authentication. Once an entity (user) is authenticated, she becomes a member of the set of Authenticated entities, $Auth_p$ of the particular provider, p .

(c) **Authorization Phase:**

The third phase requires making decision by the proposed system in order to ensure the performance or non-performance of certain action by entity (user) on a specific resource in a particular context based on an identifier values(s). Authorization only takes place if an entity is authenticated in the p , meaning the entity is in the set Auth.

Many systems combine the identification and

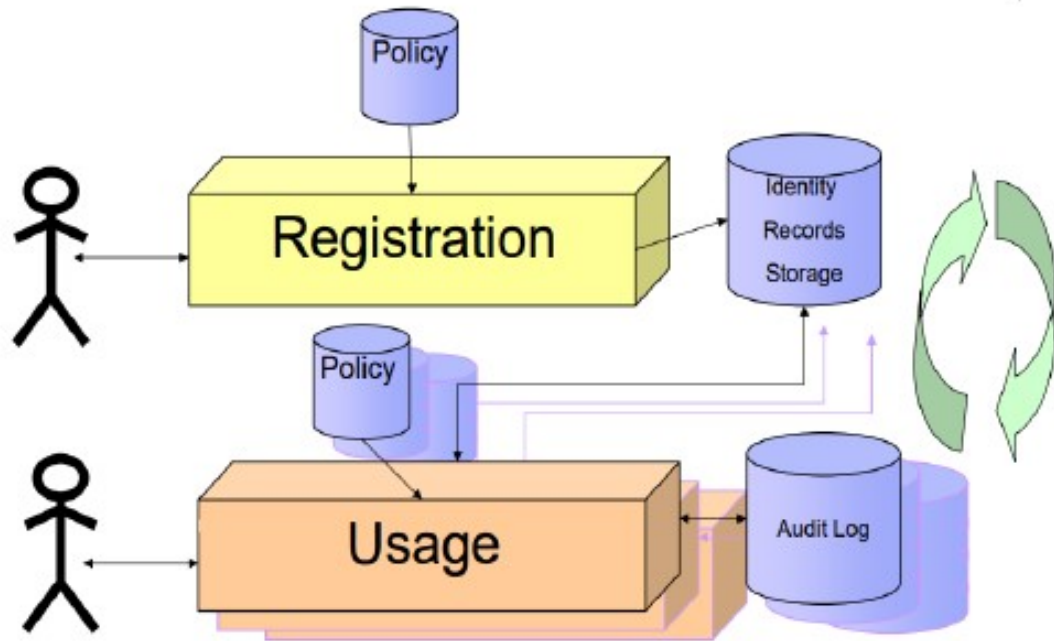


Figure 2: Workflow model of the proposed system.

authentication process into a step.

$$\text{Auth} = \begin{cases} 1 & \text{if } f_{vx} \text{ matches } \{e\} \\ 0 & \text{if otherwise} \end{cases} \quad (4)$$

where vx represents an identifier that uniquely identifies a given entity and Auth means authentication.

(d) **De-registration Phase:**

This is the final phase of this IdM system. It allows users to be de-registered from an identity provider p . This process usually removes the entity and the association between the entity and both the identifier as well as attributes from a specific context (or Identity Provider). De-registration is the reverse process of registration.

The process is modeled as shown in Eq. (5) and (6):

$$E'_p = \frac{E_p}{e} \quad (5)$$

$$AV'_p = \frac{AV_p}{AV'} \quad (6)$$

where E_p represents the updated set of entities, AV_p denotes the updated set of attribute values.

The life cycle of the proposed IdM system discussed above shows that the operations of registration, identification, authentication and de-registration occur at the identity provider (P) while authorization and service provisioning take place at the service provider (SP).

Identity management system (Fig. 1) that is developed has two parts namely registration phase and authenticating phase. The registration process is in two phases (Fig. 1), first is the basic personal information stage where username, password and few other details are provided. The second stage is the biometric phase, which uses face image for complete authentication.

I. **Registration Phase:**

Here the user commits his strong identifiers to be used later as proofs of identity these are basic personal information such as username, password and few other details.

II. **Authenticating Phase:** Already enrolled users are identified using the registered credentials. Users will be allowed to have access to resources based on their identifiers and credentials during service operation.

Multifactor authentication technique is used in this proposed IDM system: what you know (username and password) and who you are (face recognition and detection).

Figure 2 captures the workflow of the proposed system that comprises of Registration and Usage phase. Figure 3 shows the overview of the face recognition phase. The camera capture client image and apply local binary pattern on it, at the verification phase, image is matched with the already registered image, if it matches, it grants access, if not access is denied.

Application software was developed to simulate the proposed Identity Management system using Python and Django database management system as backend engine. Django Framework was employed in the development of this solution because

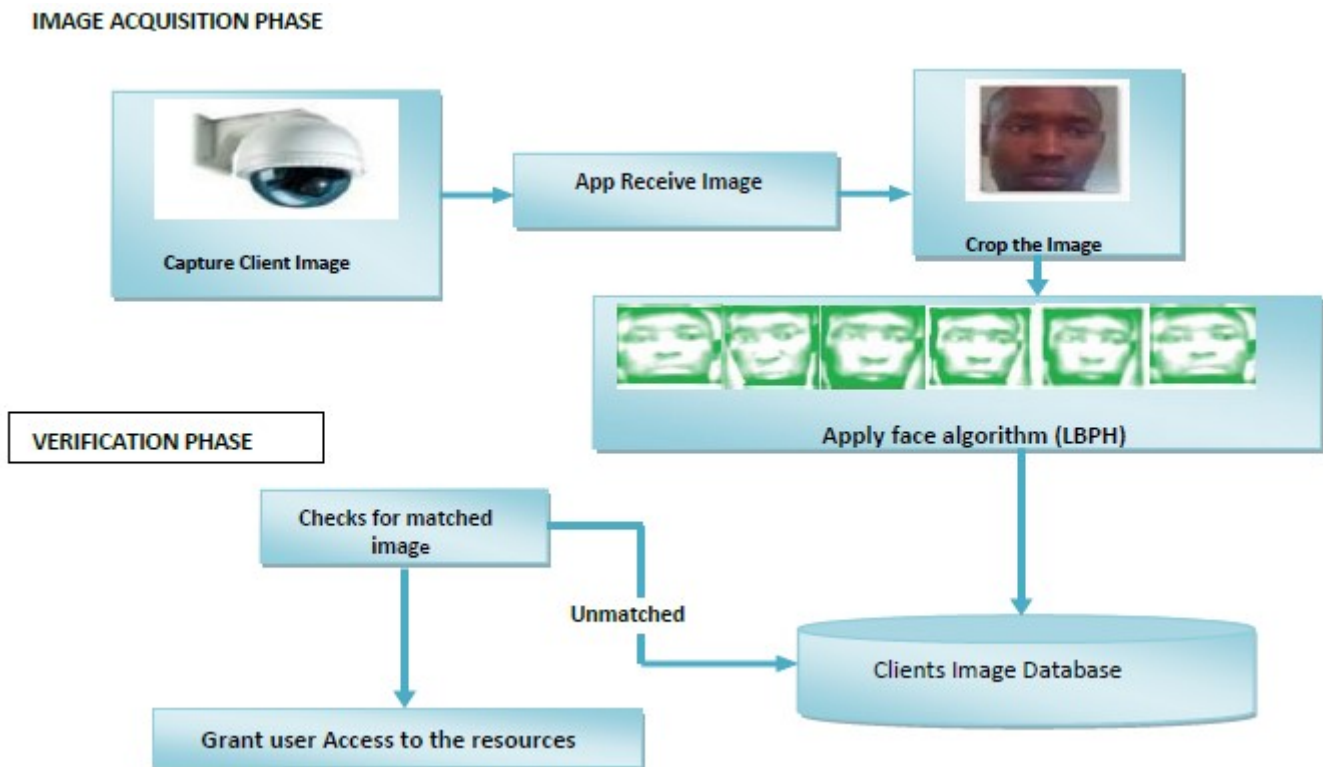


Figure 3: Overview of the face recognition phase.

of its robustness as a web framework. The configuration of the entire system was done from a single-entry point called the settings.py file. Here the connection to the database was defined, the session and cookie management systems and other key areas that are of importance.

A custom IdM solution was built which uses a two-step approach, namely: the username/password sign-in and the face recognition stage.

The IdM system handles both authentication and authorization. Authentication verifies a user as who they claim to be, and authorization determines what an authenticated user is allowed to do. Here the term authentication is used to refer to both tasks. The auth system consists of:

- i. Users
- ii. Permissions: Binary (yes/no) flags designating whether a user may perform a certain task.
- iii. Groups: A generic way of applying labels and permissions to more than one user.
- iv. A configurable password hashing system
- v. Forms and view tools for logging in users, or restricting content
- vi. A pluggable backend system

3.1. Training the Algorithm

First, we needed to train the algorithm. To do so, we used a dataset with the facial images of the people we want to recognize. It is required to also set an ID (it may be a number or the name of the person) for each image, so the algorithm will use this information to recognize an input image and give you an output. In our case the primary key generated for the user during the first stage of registration was used, as this remains unique throughout the life cycle of the application. Images of the same person must have the same ID.

3.2. Applying the LBP Operation

The first computational step of the LBPH is to create an intermediate image that describes the original image in a better way, by highlighting the facial characteristics. To do so, the algorithm uses a concept of a sliding window, based on the parameters radius and neighbors. Figure 4 shows this procedure.

From Fig. 4, suppose we have a facial image in grayscale, we can get part of this image as a window of 3×3 pixels. It can also be represented as a 3×3 matrix containing the intensity of each pixel (0~255). Then, we need to take the central value of the matrix to be used as the threshold. This value will be used to define the new values from the 8 neighbors. For each neighbor of the central value (threshold), we set a new binary value. We set 1 for values equal or higher than the threshold and 0 for values lower than the threshold. Now, the matrix will contain only binary values (ignoring the central value). We need to concatenate each

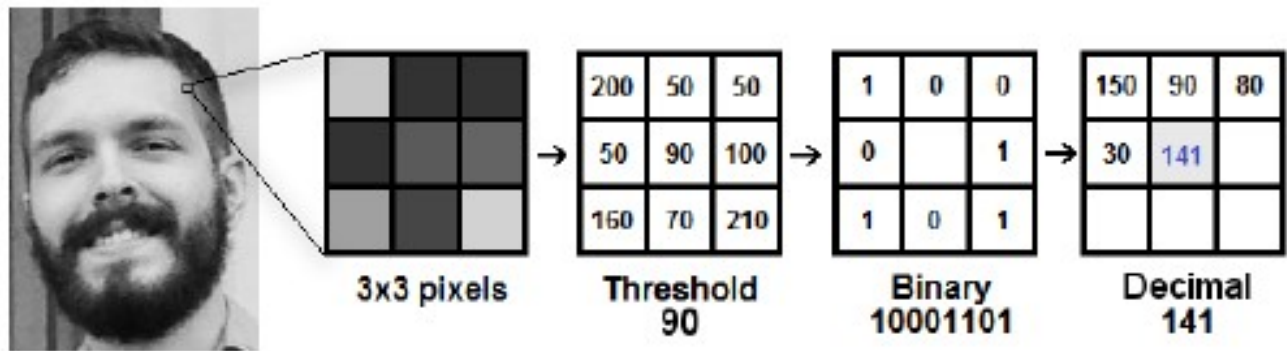


Figure 4: Local Binary Pattern Algorithm processing an image [14].

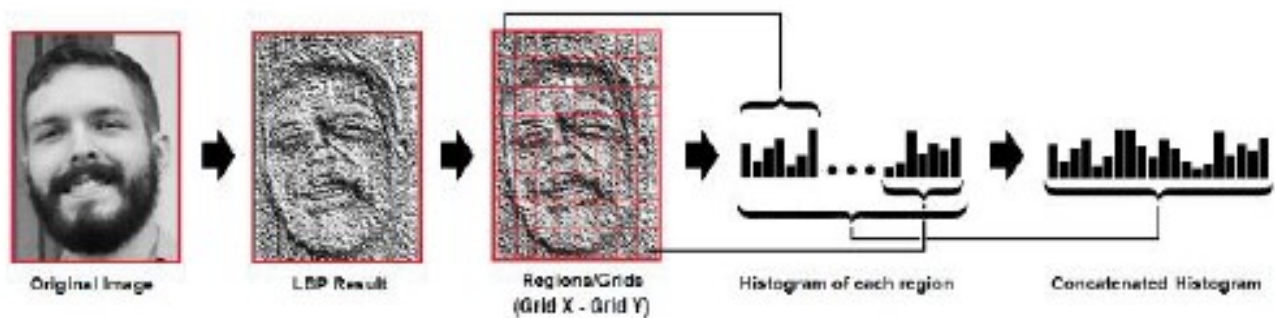


Figure 5: Extracting the histogram [14].

binary value from each position from the matrix line by line into a new binary value (example is 10001101). Then, we convert this binary value to a decimal value and set it to the central value of the matrix, which is actually a pixel from the original image. At the end of this procedure (LBP procedure), we have a new image, which represents better the characteristics of the original image.

The LBP procedure was expanded to use a different number of radius and neighbors, makes it Circular LBP.

3.3. Extracting the Histograms

Using the image generated in the last step, we can use the Grid X and Grid Y parameters to divide the image into multiple grids, as can be seen in figure:

Based on the Fig. 5, we can extract the histogram of each region as follows:

- As we have an image in grayscale, each histogram (from each grid) will contain only 256 positions (0~255) representing the occurrences of each pixel intensity.
- Then, we need to concatenate each histogram to create a new and bigger histogram. Supposing we have 8×8 grids, we will have $8 \times 8 \times 256 = 16,384$ positions in the final histogram. The final histogram represents the characteristics of the image's original image.

3.4. Performing the Face Recognition

In this step, we have an already trained algorithm. Each histogram created is used to represent each image from the training dataset. So, given an input image, we perform the steps again for this new image and creates a histogram, which represents the image. So to find the image that matches the input image we just need to compare two histograms and return the image with the closest histogram. We can use various approaches to compare the histograms (calculate the distance between two histograms), for example: Euclidean distance, chi-square, absolute value, etc. In this research, we used the Euclidean distance based on the following formula:

$$D = \sqrt{\sum_{i=1}^n (\text{hist}1_i - \text{hist}2_i)^2} \quad (7)$$

where D is the IDENTITY from the image with the closest histogram. The algorithm should also return the calculated distance, which can be used as a 'confidence' measurement. Lower confidences are better because it means the distance between the two histograms is closer. Use a threshold and the 'confidence' to automatically estimate if the algorithm has correctly recognized the image. Assume that the algorithm has successfully recognized if the confidence is lower than the threshold defined.

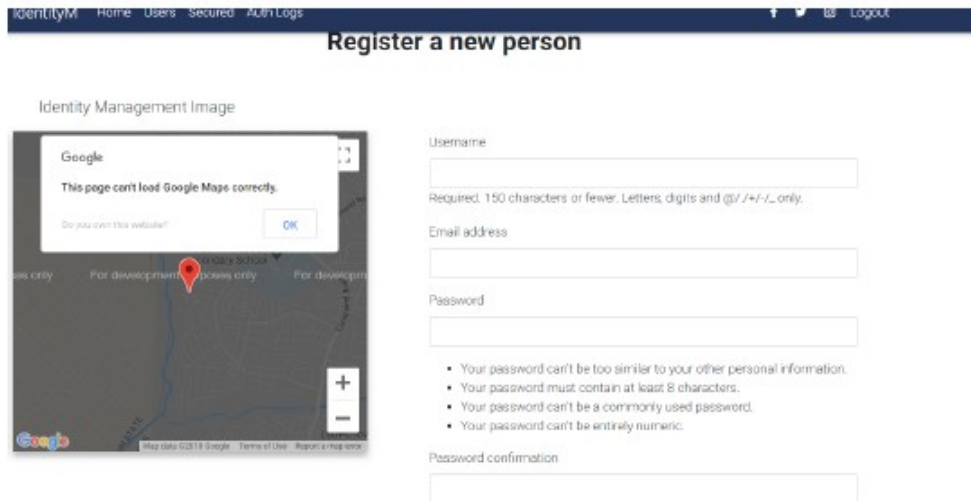


Figure 6: Web page showing registration page.

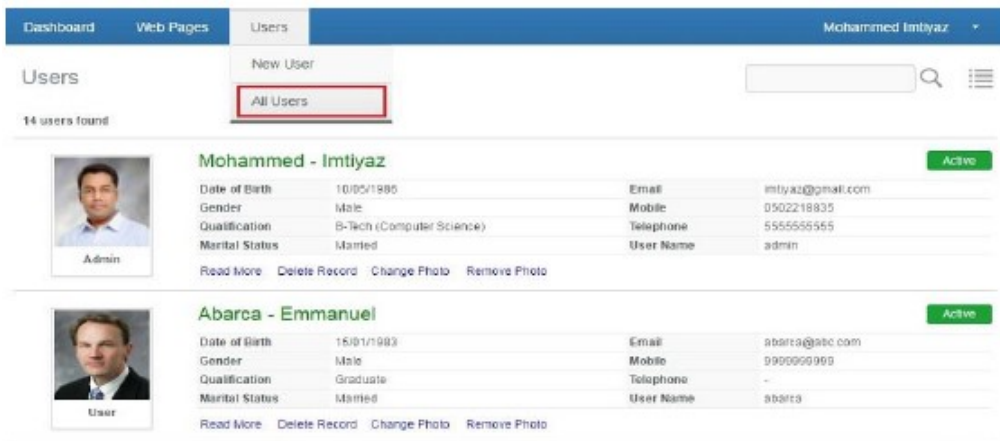


Figure 7: Web page showing already registered users and their credentials

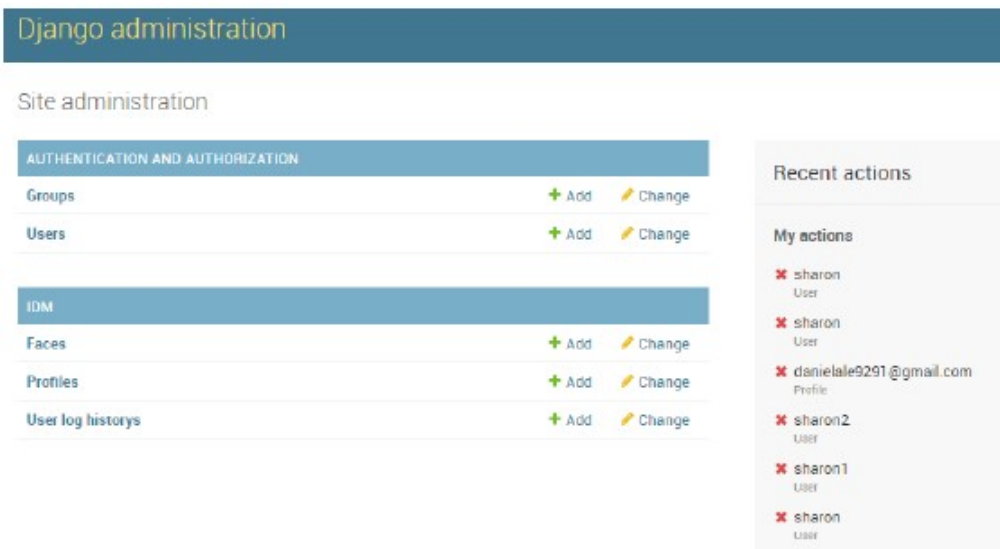


Figure 8: Web page showing admin page.

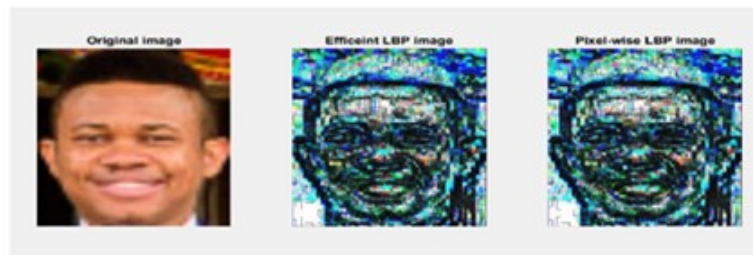


Figure 9: Local binary pattern result.

4. SYSTEM IMPLEMENTATION

The purpose of the implementation phase is to translate the model designed into source code. Each component of the design is implemented as a program module. The end product of this phase is a set of program module that has been individually tested. Each module is unit tested to determine the proper functioning of all the individual modules. It involved testing each module in isolation, as this is the most efficient way to debug the error identified at this stage. All program code are implemented using Python/Django.

4.1. System Testing

After the system had been completed and all components unit testing completed, the different modules or components are integrated in a planned manner. The integration was carried out incrementally over a number of steps. When all the modules have been successfully integrated and tested, complete system testing was carried out. The system testing helped ensured requirement and design conformance. The kind of system testing carried out is α -testing (i.e. testing performed by the development team).

4.2. Tools required

Below is a list of items that will be required to accomplish the goals and objectives of this project:

- i. Windows Operating System (at least Windows Professional)
- ii. Python/Django Development Kit
- iii. Python Web frame work platform
- iv. Json request call
- v. Relational Database Management Systems

4.3. Accuracy Chart

A simulation was run on the model for users ranging from 10 to 1000, and accuracy was found to remain as high as 98.01% as shown in Table 1.

The Identity Management System was designed and developed round-up using Python/Django. The system has two parts, the User Interface where the administrator can register users and manage users, as well as the WEB-API section which allows remote users and companies or mobile application to consume the resources through a JSON request call.

Table 1: Coordinates of well points.

Users	Accuracy (%)
10	100
100	99.96
200	99.45
300	99.24
400	99.1
500	98.95
600	98.86
700	98.56
800	98.43
900	98.21
1000	98.01

The welcome page appears when the application is launched. It contains the home, users, secured and audit logs. The system administrator can register new users (Fig. 6), in cases where the user has already been registered (Fig. 7), different roles can be assigned using the manage icons. Also, already registered users can login in using the log in icons.

The login activity logs allows admin to track different activities carried out by different users, this can be used to track security breaches. The admin can assign different roles to users, he can also view user log history, profiles and faces as shown in Fig. 8.

4.4. Face Recognition

The face recognition phase uses the LPBH algorithm, which makes it much easier to train individual faces as they are coming in and add to the feature space rather than re-training the entire database together again, which can cause a lot of computational issues. Figures 9 and 10 are examples of face images used for training, and three different variants of LPBH algorithms were considered in order to determine which will be the best, as well as the trade-offs of each.

The average time taken for training each image is about 40 seconds as seen in the Fig. 11. Figure 12 represents the histogram generated from the face after LPBH algorithm has been applied. For a single individual, the histograms from each image are concatenated and used for recognition. The more the images used, the more reliable the recognition process.

The sparse algorithm and tight algorithm variants shown below were also considered, and it can



Figure 10: Image processing using LBPH.

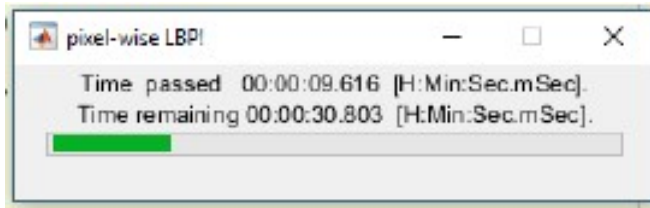


Figure 11: Image processing using LBPH.

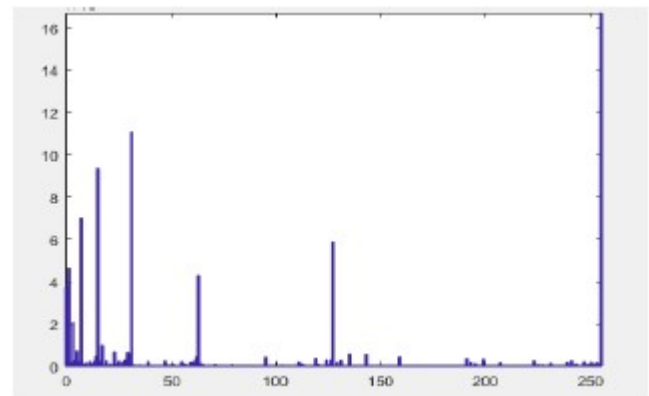
be shown that the tight algorithm is better than the other two since it attempts to use the significant part of the result while cutting off the insignificant section of the data. This is useful for low storage and fast implementations.

4.5. Cybercrime curbing by python IDM

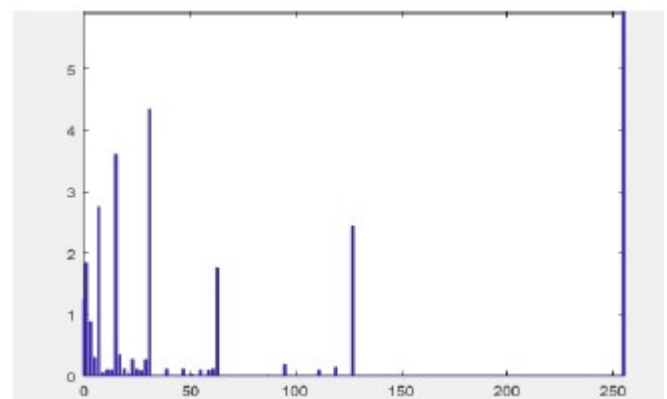
Cybercrime prevention can be achieved fairly quickly and in a cost-effective manner when armed with a little technical advice and common sense. Many cybercrime attacks can be avoided. Similar to target hardening for a residence or a business (examples are., lights, locks, and alarms), the more difficult it is for a cyber criminal to successfully attack a target, the more likely he or she is to leave it alone and move on to an easier target [14]. Python IDM addressed this by the use of a two factor authentication method which makes it difficult for the system to be hacked. Any enduring method of preventing cyber crime must reveal the identity of the perpetrators; this is achieved by the activity log in Python IDM. The activity log shows the time and date and also the number of times a user attempts to log in. In cases of security breaches, the system can detect the identity of the user that was used to log into the system and hence the user (criminal) can be apprehended.

4.6. Comparative Analysis

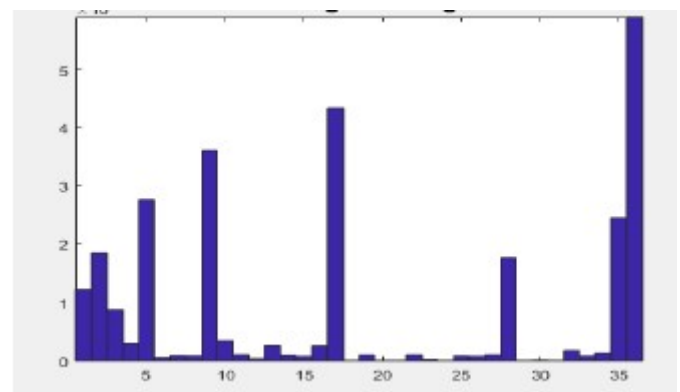
Successful Identity Management system increases the efficiency, security, access control and decreases the complexity, cost and many repetitive works. A comparative analysis of the identity management system developed has been carried out to establish the functionality of the work as discussed in literatures, the criteria that was chosen for this study are summarized in Table 2. The work was compared with LepidIdm and OracleIdm using the following metrics: Functionality, Security and Privacy.



(a) Regular LBP Histogram



(b) Simulation of pH with 10% increase in chloride



(c) Simulation of pH with 10% increase in bicarbonate

Figure 12: Histogram generated from the face after LPBH algorithm.

Table 2: Coordinates of well points.

System	Provisioning	Workflow	Password Reset	Single Signon	Directory Services	Replication	Application Development
PhytonIdM	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LepidIdM	Yes	Yes	Yes	Yes	Yes	No	No
OracleIdM	Yes	Yes	Yes	No	Yes	No	Yes

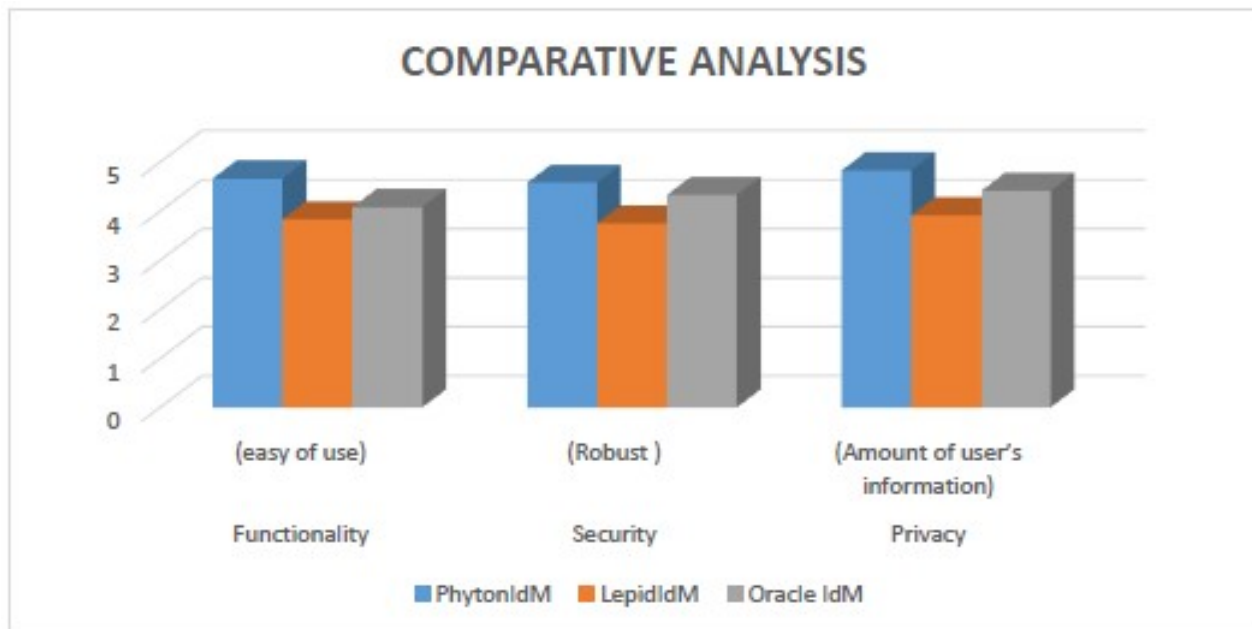


Figure 13: Comparative Analysis of developed IDM with two other IDM's.

After installation of the three IDMS on windows operating systems which was supported by three companies. Table 2 shows the comparative analysis based on the metrics used while the bar chart in Fig. 13 shows the graphical analysis. A unique weight of 5.0, 4.0, and 3.0 were assigned to very good, good and bad in each questionnaire to find the average result. The result shows that users find PythonIdM to have higher security, privacy and functionality.

5. CONCLUSION

With the tremendous expansion of the Internet during the last twenty years or so, more and more identities and credentials have been issued, making their management challenging, both for service providers and users. To address this menace, the need for the development of identity management solution to control cybercrime. Identity Management system has potential to provide a secure and collaborative environment.

In this research, a user identity management system called PythonIDM was developed. The system provided solution to the problem of Identity Theft with the help of privacy preserving multi-factor authentication. A two-factor authentication system was used in the implementation of the IDM solution that uses federated Identity Model. A trusted platform Module within the system has been developed to ensure strong integrity.

The proposed system is dynamic in nature and by integrating a face recognition system the system has a robust security. The solution can be rendered as a service to enhance extra security layer for applications since authentication takes place outside the application. This makes the customer free from the burden of installing and operating the application on own computer and also eliminates the dreadful load of software maintenance; continuing operation, safeguarding and support.

References

- [1] O. B. Longe and S. C. Chiemeké, "Cybercrime and Criminality in Nigeria: what roles is Internet Access Points Playing?" *European Journal of Social Sciences*, vol. 6, no. 4, pp. 132–139, 2008.
- [2] S. Brenner, *Law in an Era of Smart Technology*, ser. ISBN 978-0-195-33348-0. New York: Oxford University Press, Inc., 2007.
- [3] I. M. Erhabor, "Cybercrime and the youths," Department of Education, Ambrose Alli University, Ekpoma, Nigeria, 2008, pGDE Thesis.
- [4] O. O. Olanmi, "Computer Crimes and Counter Measures in the Nigerian Banking Sector," *Journal of Internet Banking and Commerce*, vol. 15, no. 1, pp. 1–10, 2010.
- [5] (2007) Categories of identity theft. Identity theft resource centre. [Online]. Available: www.idtheftcentre.org
- [6] J. Camp, "Digital identity," *IEEE Technology and Society Magazine*, vol. 23, no. 3, pp. 34–41, 2004.
- [7] M. S. Ferdous, A. Jøsang, K. Singh, and R. Borgaonkar, "Security usability of petname systems," in *Identity and Privacy in the Internet Age*, vol. 5838.

- Berlin/Heidelberg: Springer, 2009, pp. 44–59, lecture Notes in Computer Science.
- [8] S. I. A., “Examining the Effects and Challenges of Cybercrime and Cyber Security Within the Cyberspace of Sierra Leone,” *International Journal of Intelligent Information Systems*, vol. 7, no. 3, pp. 23–27, 2018.
- [9] Federal Trade Commission. (2005) National and state trends in fraud and identity theft from january to december, 2004. [Online]. Available: <http://www.consumer.gov/sentinel/pubs/Top10Fraud2004.pdf>
- [10] O. Udenze, “The effect of corruption on foreign direct investments in developing countries,” *The Park Place Economist*, vol. 22, no. 1, pp. 87–95, 2014.
- [11] M. S. Ferdous, “User-controlled identity management systems using mobile devices,” Ph.D. dissertation, School of Computing Science College of Science and Engineering University of Glasgow, United Kingdom, 2015.
- [12] M. A. Agana and H. C. Inyama, “Cybercrime detection and control using the cyber user identification model,” *International Journal of Computer Science and Information Technology & Security*, vol. 5, no. 5, pp. 354–368, 2015.
- [13] Hedayati, “An analysis of identity theft: Motives, related fraud, techniques and prevention,” *Journal of Law and Conflict Resolution*, vol. 4, no. 1, pp. 1–12, 2012.
- [14] M. K. Adu, B. K. Alese, and O. S. Adewale, “Mitigating cybercrime and online social network threats in nigeria,” in *Proceedings of the World Congress on Engineering and Computer Science 2014*, vol. 1, San Francisco, USA., October.