Article can be accessed online at http://www.publishingindia.com

Remote User Authentication using Optical Character Recognition

C. Sreekanth Reddy*, C. Shobha Bindu**

Abstract

User authentication (Zhao, 2007) is an important aspect in information security to protect privacy of the users. In order to do so, users have been using a traditional way of entering the password through keyboard. The main disadvantage of these traditional passwords (Vamvakas et al., 2008) is related to memorability of the users, where users are not able to remember those passwords. To reduce the memory burden on users, graphical passwords are introduced which uses graphics and images. This paper proposes a new Re-call Based Graphical password scheme which remotely authenticates the users using Optical Character Recognition (OCR) (Rodrı´guez-Serrano & Perronnin, 2012; Wikipedia; Bandyopadhyay, 2008). This paper also aims to avoid the Shoulder Surfing attack which is a major security threat to Graphical Password schemes.

Keywords: OCR, Remote User Authentication, Graphical Passwords, Text Passwords, Shoulder Surfing Attack

1. Introduction

In modern times, passwords are used to access the protected computers, mobile phones, auto teller machines (ATM), and many more. A typical computer user require passwords for many purposes like to access computer accounts, retrieving emails from servers, accessing the files stored in databases, networks, websites and etc.,

Major requirements of any authentication system are:

- Passwords must be easy to remember.
- Passwords must be secured.

The fact that traditional text passwords being difficult to memorize often leads the users to write them down or even save them in a file which may compromise the security of the passwords.

Graphical passwords are introduced as an alternative to text passwords by Blonder (1996) motivated by the fact that humans can remember images better than text. The Graphical passwords are divided into two categories given as follows:

1.1. Recognition Based Techniques

In this category, users will choose images, icons from a collection of predetermined sets of images. In authentication process, the user needs to recognise their registration choice among a set of images.

1.2. Review of Recall-Based Techniques

In this category, user needs to reproduce their passwords without giving any reminder, hints or gestures.

Passdoodle is a graphical password comprised of handwritten texts or designs, usually drawn on a touch screen or a canvas. In their paper, Jermyn *et al.* proved that doodles are harder to crack due to much larger number of password space than text passwords. Figure 1 is sample of Passdoodle password.

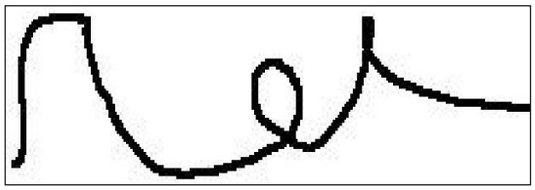
Weakness: With reference to their paper, they found that people could remember complete doodle image as an

* Student, Department of Computer Science Engineering, Jawaharlal, Nehru Technological University, Anantapur, Andhra Pradesh, India. Email: 23.oct.1988@gmail.com

** Associate Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Anantapur, Andhra Pradesh, India.

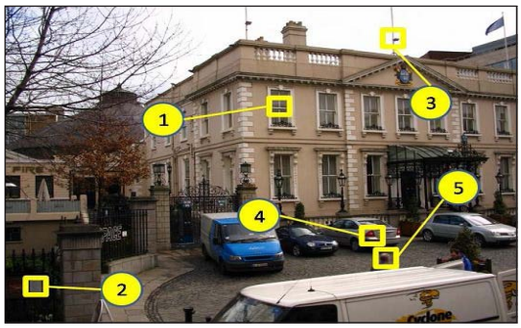
alphanumeric password, but they are less likely to recall the order in which they draw a doodle than the resulting image.

Figure 1: An Example of Passdoodle



Users were fascinated by the doodles drawn by the other users, and they frequently enter the Login details of other users to see a different set of doodles from their own.

Figure 2: Sample of PassPoint Password



PassPoint: In 2005, PassPoints were created in order to overcome the limitation of the Blonder Technique which is limitation of images. The image can be any natural image or a painting but at the same time it must be good enough in order to have many possible click points. On the other hand, the image is not secret and has no role other than helping the user to remember the click points. Another source of flexibility is that there is no need for predefined click regions with marked boundaries like Blonder technique. The user chooses several points on image in a particular order (Wiedenbecka *et al.*, 2005). Figure 2 shows an example of PassPoint password.

Weakness: Users in PassPoint system are able to create a valid password very easily and quickly, but they had difficulty learning their passwords than alphanumeric passwords, taking more trails and more time to complete the practice, on the other hand the login time is longer than traditional passwords.

Due to the weaknesses of the above techniques, the proposed scheme is aimed to avoid those weaknesses with effective login time.

2. Proposed Scheme

The proposed scheme extends the concepts proposed by Reddy & Bindu, (2013) as the base for providing authentication to the user remotely. The OCR is chosen in the scheme due to its high speed computational time by which the proposed scheme will have an advantage of login time over other schemes which uses graphic engines

The major disadvantage of the scheme proposed by Reddy & Bindu, (2013) is that even though that scheme uses an invisible rectangle to select the word at the time of login phase, the path of the selection is visible to the adversary by whom there is a chance of compromising the selected word or the password. In order to avoid that and to make the proposed scheme completely resistant to shoulder surfing the scheme uses a double click mouse for selecting the password at the time of login, which means user has to select the password by clicking both the mouse buttons pressed then only the word is selected otherwise the word will not be selected. By this even if the adversary was able to see the path he cannot be able to login.

As any other remote user authentication schemes, the proposed scheme also has three phases.

2.1. Registration Phase

This phase is executed only once, when the user first register himself with the server.

1. In this phase, user has to enter his/her username and then has to load a textual image of his/her choice, both the user name and the image are stored remotely using Oracle DB.
2. Then user has to select a word of his/her choice from that textual image and that word will be trimmed using pre-processing stage of OCR and then the rest of the OCR process is done on that trimmed image by getting the converted text as output and that text is hashed and is stored at server as password under that username.
3. So, completely the username, the image and the hashed password are saved in DB remotely as shown in Figure 3.

Figure 3: Remote Connections in Registration Phase

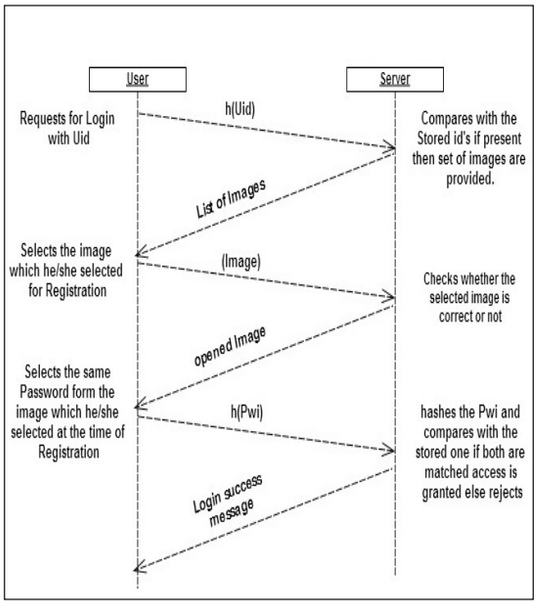
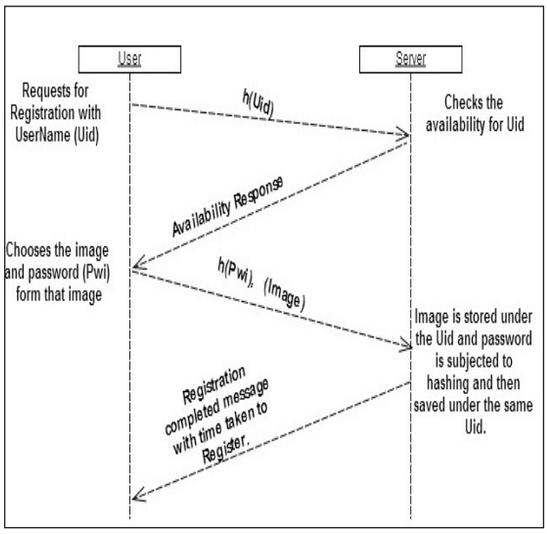


Figure 4: Login Connection



2.2. Login Phase

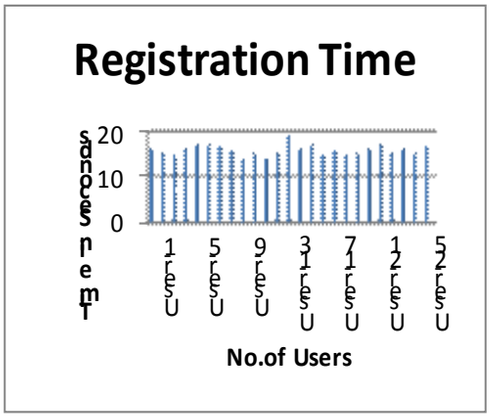
1. In the login phase, the user has to enter his/her username and the request is sent to the server for authentication.
2. If the username exists, then the server displays a list of predetermined images to the user from which he/she has to select the same image which the user selected at the time of registration if the user fails to select the same image then the authentication process fails, otherwise the image is retrieved from the Remote DB and then displayed to the user.
3. The user has to select exactly the same word which he/she selected at the time of Registration phase.
4. The user has to keep in mind that he has to press down both the mouse buttons while selecting the word otherwise the word will not be selected and the user will not be authenticated.
5. Afterwards the selected word is again converted to text using OCR and then hashed and then compared with the remotely stored password. If both the passwords are matched then the user is authenticated, else the user is prompted to login again from the beginning. The connections are shown in Figure 4.

2.3. Password Change Phase

This phase is invoked if the user wants to change his/her password. The user has to be authenticated by the system following the steps in login phase, and then the user can change his/her password.

1. The user makes a request to the server by clicking on the change password button then the server responds by providing a file chooser for selecting his/her textual image.
2. The selected image is explicitly provided to the user for password selection.
3. Then user can select a different password by double click event and that sends to Server.

Figure 5: Password Change Phase



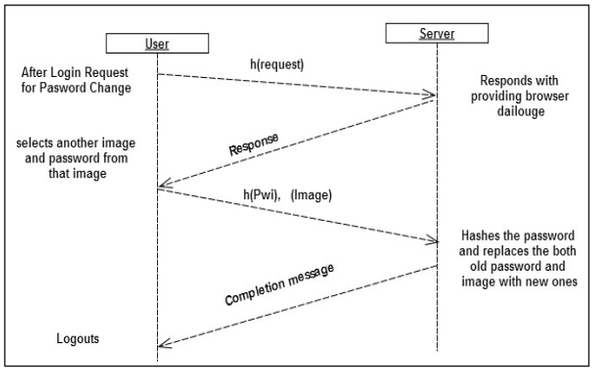
4. The remote server hashes the password and then replaces the old password with the new one with confirmation message.

3. Usability Analysis

25 novice users from the university campus, who were unfamiliar with the proposed scheme, were considered to study the usability related implications of the proposed scheme.

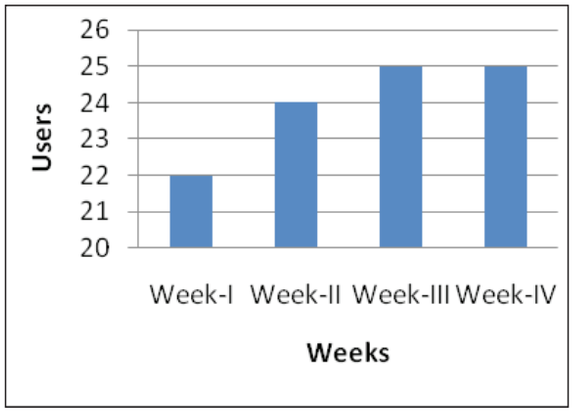
The time taken to register with the scheme by 25 users is depicted in Figure 6.

Figure 6: Registration Time for 25 Users



Testing is carried out with 25 users in a weekly manner for 4 weeks. Figure 7 shows the number of users who could successfully login after their immediate Registration.

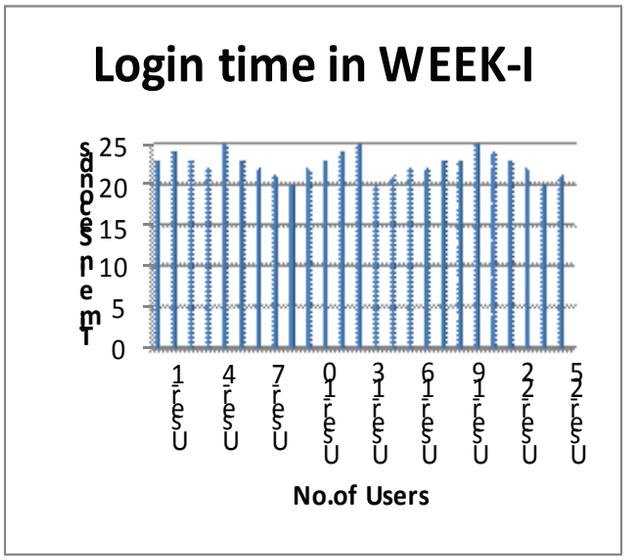
Figure 7: No. of Successful Logins



From the table (Figure 7), it is clear that users can remember their selected word effectively. Further, it is also noticed that after sufficient exposure to the proposed scheme users are able to lgin even more effectively.

The next and most important aspect of the usability is the login time. The proposed scheme has an advantage of that over many other schemes due to the fast and effective processing capabilities of OCR. The Figure 8 shows the time taken by 25 users to login in the first week.

Figure 8: Login Times of 25 Users in Week I



After a week, the users are asked to login again, when we observed the login time is reduced compared to week I. Figure 9 shows the login time of 25 users in 2nd Week.

Figure 9: Login Times of 25 Users in Week II

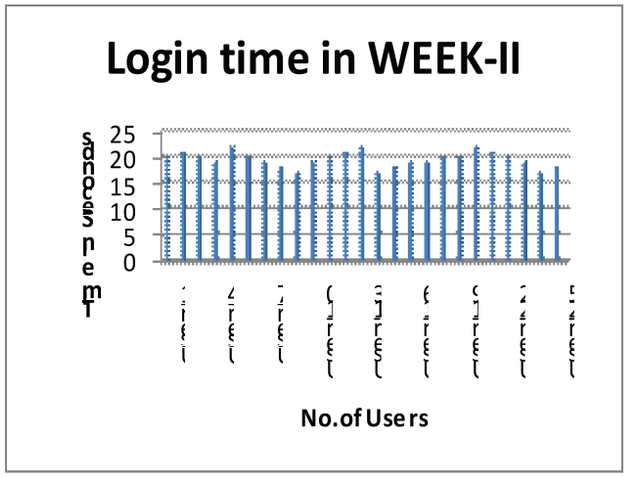


Figure 10: Login Times of 25 Users in Week III

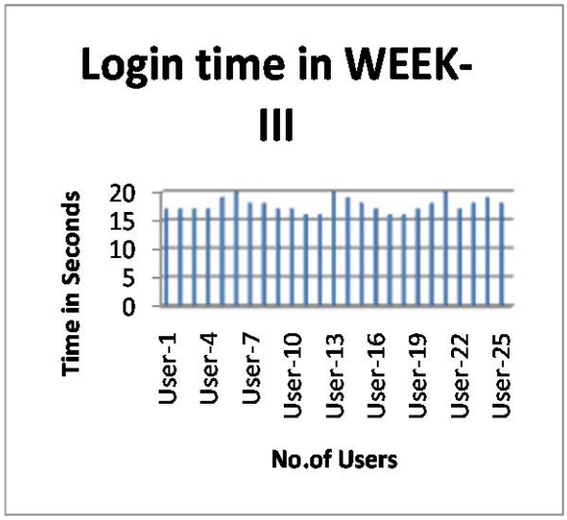


Figure 10 shows the login time of 25 users in 3rd Week.

Figure 11: Login Times of 25 Users in Week-IV

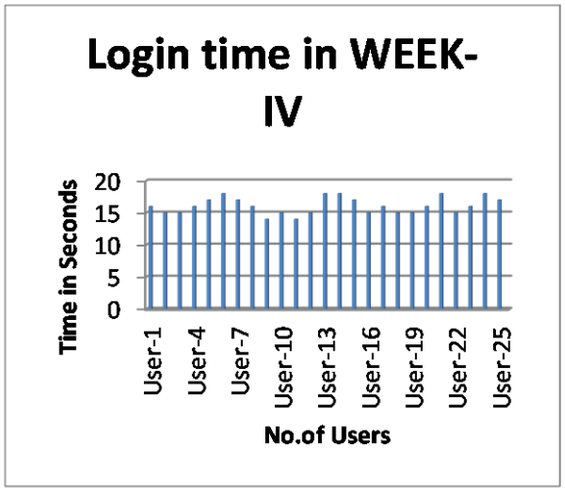


Figure 8 shows the login time after 3weeks. As the users learn the system the time taken to login is diminishing**.**

Table 1: Average Login Time

Week	*Avg.Login Time (OCR)*	*Avg.Login Time (PCCP)*
I	22.52	24.32
	19.52	26
III	17.68	27
IV	16.08	30.12

After the users are trained well about the proposed scheme, the time taken to login is considerably decreased to minimum as shown in figure 11 and the average time taken to login is also low as depicted in Table 1 compared to other Graphical password systems like PCCP (Chiasson *et al*., 2012) where they use Hotspots which increases the login time to 30 seconds.

3.1. System Usability Scale (SUS)

The System Usability Scale (SUS) provides a "quick and dirty", reliable tool for measuring the usability. It consists of a 10 item questionnaire with five response options for respondents; from strongly agree to strongly disagree.

3.1.1. The System Usability Scale

When a SUS is used, participants are asked to score the following 10 items with one of five responses that range from Strongly Agree to strongly disagree:

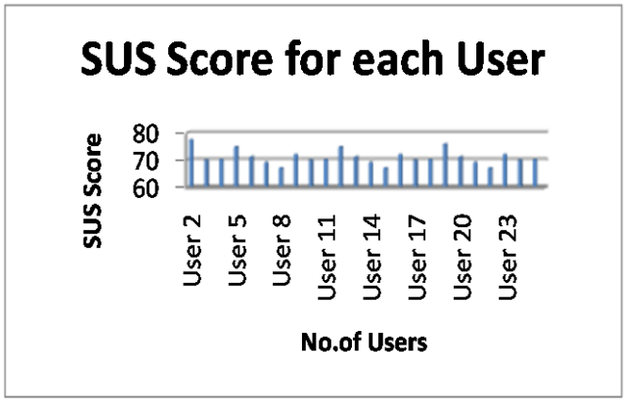
1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

3.1.2. Interpreting Scores

Interpreting scoring can be complex. The participant's scores for each question are converted to a new number, added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. Though the scores are 0-100, these are not percentages and should be considered only in terms of their percentile ranking.

The proposed scheme is tested for SUS with help of 25 novice users using Morae tool which is usability testing software that uses the above SUS process for calculating the Usability of the proposed scheme. Using Morae tool, the user experience of 25 users is recorded and the SUS Score is calculated for each user, Figure 10 depicts the same.

Figure 12: SUS Score for 25 Users



The SUS score indicates that how well a system can be used by the user without any external help. The average SUS score of the proposed scheme is 70.98 which is pretty good and indicates that the scheme is well used without any effort by the user.

4. Security Analysis

4.1. Brute Force Attack

As the password space increases, the chances of guessing or cracking the password will be decreased.

In the proposed scheme if the selected image contains 100 words and if user chooses a word which is equal to 4 letters then the password space will be $100C_4$. If the user selected a word which is equal to 8 letters then password space will become $100C_8$ which is not possible to crack using Brute force attack [10]. The same is depicted in Table 2. By default we consider images with 100 words.

Table 2: Password Space of the Proposed Scheme

Scheme	*Password length*	*Password Space*
Proposed scheme	4	100C4
	8	100C8

4.2. Shoulder Surfing

Another major security threat to any graphical password scheme is Shoulder Surfing [9] which means the extent to which someone can look over the shoulder of a person entering his/ her password and guess his/ her password.

The proposed scheme avoids shoulder surfing in two ways one by using an invisible rectangle for selecting the word at the time of login and another by confusing the adversary using double click events, which means user has to press down both the mouse buttons while selecting the password or else the word will not be selected which intern leads to authentication failure. Even if the adversary can observe the path drawing he will not be able to select the word.

5. Conclusion

The proposed scheme is not only used to restrict unauthorised users, but also used to protect databases from various attacks due to the absence of string input from users. The scheme proposed by Sreekanth *et al.* (2013) has a disadvantage of path visibility while selecting the password at the time of login, so the proposed scheme overcomes that problem using double click event and also provides authentication to the user remotely, all the user data is stored in a remote server. The proposed scheme also has an advantage of number of success full logins. Login time on the other hand is very low due to the fast computing speed of OCR. The login times of 25 users are considered in weekly manner for 4 weeks, after testing the time taken to login in diminishing week by week. The usability is tested using Morae tool and the SUS score is calculated for each user. The result of the SUS calculation shows that the proposed scheme can be well used by the user without any help or support from the external source.

References

Agarwal, G. (2010). Security analysis of graphical passwords over the alphanumeric passwords. *International Journal of Pure and Applied Sciences and Technology*, 1(2), 60-66.

Bandyopadhyay, S. K. (2008). *User authentication by Secured Graphical Password Implementation*. 7th Asia-Pacific Symposium on Information and Telecommunication Technologies, (pp. 7-12).

Blonder, G. E. (1996). U.S. Patent No. 5559961.

Chiasson, S., Stobert, E., & Forget, A. (2012). *Persuasive Cued Click-points: Design, Implementation, and Evaluation of a Knowledge-Based Authentication Mechanism*. IEEE Transactions on Dependable and Secure Computing, March/April, 9(2), 222-235.

Marco-Gisbert, H., & Ripoll, I. (2013). *Preventing Brute Force Attacks against Stack Canary Protection on Networking Servers*. 12th IEEE International Symposium on Network Computing and Applications (NCA) (pp. 22-24).

Optical Character Recognition. Retrieved from http://en.wikipedia.org/ wiki/ Optical_ character_ recognition.

Reddy, C. S. & Bindu, C. S. (2013). Effective password authentication system using optical character recognition. *International Journal of Advanced Research in Computer Science and Software Engineering*, November, 3(11), 1049-1055.

Renaud, K. (2009). On user involvement in production of images used in visual authentication. *Journal of Visual Languages and Computing*, 20(1), 1-15.

Rodrı´guez-Serrano, J. A., & Perronnin, F. (2012). *A Model-Based Sequence Similarity in Application to Handwritten Word Spotting*. IEEE Transactions on Pattern Analysis and Machine Intelligence, November, 34(11), 2108-2120.

Vamvakas, G., Gatos, B., Stamatopoulos, N., & Perantonis, S. J. (2008). *A Complete Optical Character Recognition Methodology for Historical Documents*. 8th International Workshop on Document Analysis Systems (DAS'08), (pp. 525-532).

Varenhorst, C. (2004). *Passdoodles: A Lightweight Authentication Method*. Massachusetts Institute of Technology, Research Science Institute, July 27, 2004.

Wiedenbecka, S., Watersa, J., Birgetb, J. C., Brodskiyc, A., & Memon, N. (2005). *Design and longitudinal evaluation of a graphical password system.* Academic Press.

Zhao, H. (2007). *A Scalable Shoulder-Surfing Resistant Textual-Graphical Password Authentication Scheme.* 21st International Conference on Advanced Information Networking and Applications Workshops 2, 467-472.