Enhanced Password Processing Scheme Based on Visual Cryptography and OCR

[1]Shashikala.S.V, [2]Nandan.B.N [3]Praveen Kumar [4]Priyadarshini.B [5]Rakshitha.B.K

[1]Head of Department [2345]UG Students

Department of Computer and Engineering

BGS Institute of Technology, Mandya

Abstract-Traditional password conversion scheme for user authentication is to transform the passwords into hash values. These hash-based password schemes are comparatively simple and fast because those are based on text and famed cryptography. However, those can be exposed to cyber-attacks utilizing password by cracking tool or hash-cracking online sites. Attackers can thoroughly figure out an original password from hash value when that is relatively simple and plain. As a result, many hacking accidents have been happened predominantly in systems adopting those hash-based schemes. In this work, we suggest enhanced password processing scheme based on image using visual cryptography (VC). Different from the traditional scheme based on hash and text, our scheme transforms a user ID of text type to two images encrypted by VC. The user should make two images consisted of subpixels by random function with SEED which includes personal information. The server only has user's ID and one of the images instead of password. When the user logs in and sends another image, the server can extract ID by utilizing OCR (Optical Character Recognition). As a result, it can authenticate user by comparing extracted ID with the saved one. Our proposal has lower computation, prevents cyber-attack aimed at hash-cracking, and supports authentication not to expose personal information such as ID to attackers.

Keywords-hash, visual cryptography, image-based password scheme

I. INTRODUCTION

User authentication in general systems has proceeded basically through verification of the ID and password. In order to send and verify password, the system uses a hash-based password scheme that transforms original password to hash value by famed function[1]. The advantages are that it can be adapted in system without difficulty, and computational velocity of process is fast because a type of hash-based scheme is fundamentally based on text utilizing popular hash function such as MD5, SHA256. But it is vulnerable to attacks such as brute-force attack or dictionary-based attack plainly by password cracking tool or hash-cracking online sites. Assume that someone defines password "1qaz2wsx" in a system. If an attacker is aware of the hash value "

1c63129ae9db9c60c3e8aa94d3e00495", the value can be sufficiently cracked simply by free crack site like Figure 1. Even though the attacker doesn't know any information about hash function, he or she can easily guess which kind of hash function is adapted in the system. As the result, the attacker can cause secondary damage to the system.

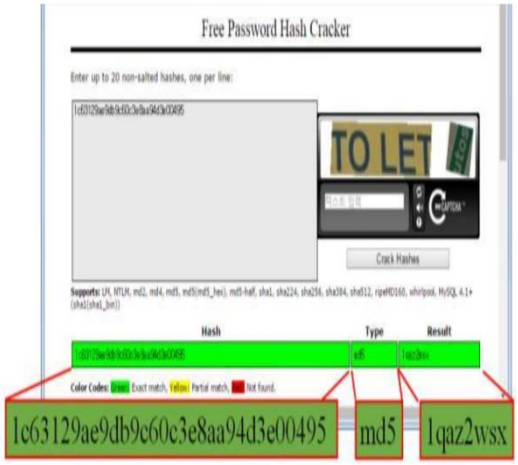

Figure 1. Result of cracked password in "crackstation.net"

Participants have the responsibility on this kind of attacks[2]. When a researcher inquired to many people about password-management behaviors, most people responded with the following five negative behaviors:[3]

choosing a computer password for the first time hardly changing a password letting someone else use own password handwriting their password next to the computer sharing a password with family, friends or coworkers.

Another major reason is using of easy passwords. The factors contributed in password security are password reuse, frequency of changing password, length, entropy level, and uniqueness with regard to password. Figure 2 shows how much people manage their passwords over satisfying the factors. First graph (a) indicates frequency for people to reuse own password, and 14 of 31 people (45.16%) answered often password reuse. Graph (b) shows how long password length is, and we can know most people have approximately 9 letter password. We can realize many people create password on their own information from graph (c) and rarely don't change password from (d).

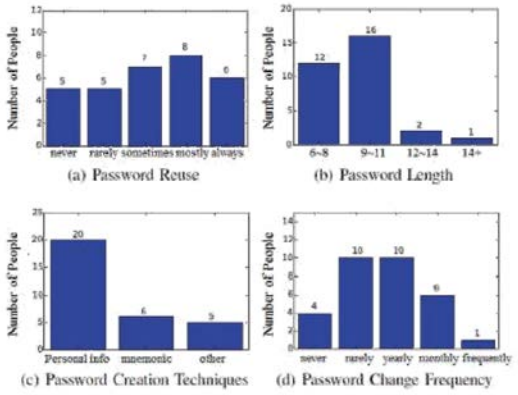

Figure 2. Survey result about password management [4]

Consequentially those behaviors become weak point and affect whole system. Many researchers have improved hash-based password scheme into the combination of password and some salts in hash function. However the salts prefixed to password cannot obstruct precomputed birthday attack to forge an unknown password[5]. In view different from text -based scheme, we suggest enhanced password scheme based on an image created by VC. The image implicitly involves password and ID. In order to verify password, proposed scheme checks ID through OCR[6]. The goal of our proposal is to prevent cyber-attack and support privacy of personal information.

II. RELATED WORK

A. Visual Cryptography

VC suggested by Naor and Shamir in 1994 is a type of image cryptography having few computation. It is to make two images derived from original image just by converting each pixel to pattern looking like noise or gray. The images are shared to others. If you again want to catch sight of original image, you gather and stack up the shared images then can see the image. Distinctively it has lower computational cost to encrypt than other cryptography. Decryption method does not even require any computation because it is dependent only on sight of human

In order to build the shared images, firstly you should prepare an original image including secret message “0129” such as picture (a) of Figure3. It must be exactly composed of white background and black letter. In fact, the research about VC has been extended to half-tone picture moreover color picture. But we are supposed to explain basic VC referred to this paper.

For encryption, you should prepare some patterns consisted of 4 subpixels arranged in a 2 x 2 array. The half of 4 subpixels is filled with black and the rest becomes transparent. It can make 6 pattern which is horizontal, vertical and diagonal. VC transforms per a pixel of original image to one of those. After VC-based image is made in full, the subpixels become as noise because shared image is combination of randomly collected patterns. The way to construct pixels of background and message in shared image should be different from each other.

If what you want to convert to a pattern is a pixel of background in original image, it should be following to background pixel matrix in Figure4. The pattern is randomly determined by one of the forms according to pattern no. For example, if you want to convert a pixel of background in original image to pattern no3, subpixels of the position in first shared image should accurately become left-vertical pattern. You also have to set definitely same subpixel at the precise position of second shared image. After all background pixels of the both shared images are determined according to this process, the background finally seems gray mixed with white and black.

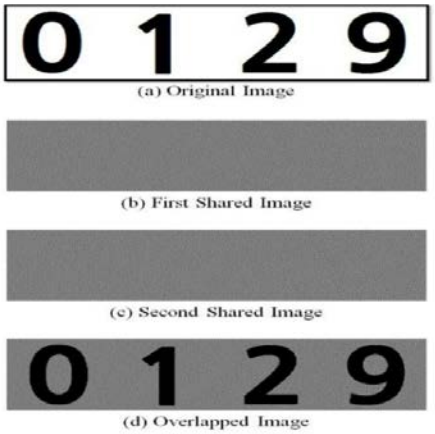

Figure 3. Pictures related with VC

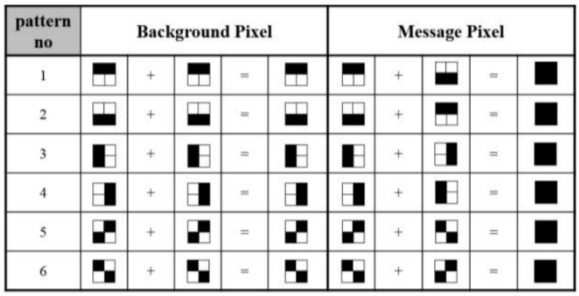

Figure 4. How to make pixel pattern in shared image

Likewise, the message part of shared image is produced in accordance with matrix of message pixel in Figure2. If you transform a modifying pixel of message in original image to pattern no 6, the subpixel of first shared image becomes left-diagonal pattern exactly at the position. Another shared image has to be defined as antitypical shape at the same position. The both shared images are correctly stacked up then the range of meaning message is black.

In conclusion, first shared image as picture (b) in Figure 3 seems gray. Presented on picture (c) in Figure 3, second shared image also appears similar with first shared image because the own pattern is defined by a pattern of first shared image. However the both shared images never

reveal secret message “0129” and any rule to build the shared image because black and transparent pixels are mixed in randomly. Exclusively when both shared images are stacked up, the view of human can confirm the message like picture (d) in Figure 3. If both images are not matched from start point to end point, or one of both images is distorted, you cannot view the message at all. The principle is to utilize higher contrast of character than background. Therefore VC has lower computation for encryption and needs not any computation for decryption.

B. OCR

OCR algorithm has been used in converting printed or written text into text to edit in machine possessed mainly in public offices such as banks, polices, hospitals, etc. People are able to recognize the text from the picture, but actually the brain performs process to interpret the picture read by eye. Facilitating this principle in machine, OCR is specified by few algorithms.

Basic OCR algorithm is template-matching method to add algebraically value to acquire a letter which is corresponded within the segments of input characters. It is implemented by calculating the total sum of the differences between sampled template and normalized original data. Other method to recognize various font is structure analysis approach that is no mathematical rule. The structure is composed of some component, and the components have features relations between the components. Therefore the method considers some logical relationship between the components such as pixels.

But during development of OCR method, several problems can be occurred as follows :

Rarely distinguish some characters for computers to understand. (example. Number one “1” and lower case L “l”)

Be more dark background or printed whole image than words For the reason, it has been researched more to recognize the letters. This paper is adapting one of various OCR algorithms to suggested mechanism. It is Tesseract that is developed by HP in 1984 at the start however now is possessed by Google. Because it is open source, many program utilizing Tesseract has been developed in online and is reported to be accuracy rate from 71% to 98%. Beforehand, we implemented some test on “FreeOCR” program and “newocr.com” site. Both programs are free programs based on Tesseract and are able to confirm whether the image mixed with numbers and characters is well recognized. The program calling “FreeOCR” perfectly reads, and distinguishes number and string from test image as Figure 5. But the site calling “newocr.com” does not exactly read as Figure 6 because it recognizes number “0” as capital “O” between capital letter “W” and small letter “r” at fault.

Actually “FineReader” as paid OCR engine at ABBYY has nearly 100% accuracy rate even if input image is attached with few noise like mixed color or blur background and letter[10]. However Tesseract algorithm is adapted in this paper for a reason to show sufficiently high accuracy rate and be open source.

Figure 5. Character Recognition in FreeOCR

Figure 6. Character Recognition in newocr.com

III. ENHANCED PASSWORD PROCESSING MECHANISM

We suppose that a server in general system identifies a user for user authentication. This section explains specifically the procedures between the user and the server though proposed password scheme based on VC and OCR. Simulation result is also provided.

A. Proposed Scheme

Before user authentication, the user has to register himself or herself to server system. Figure 7 presents the registration process.

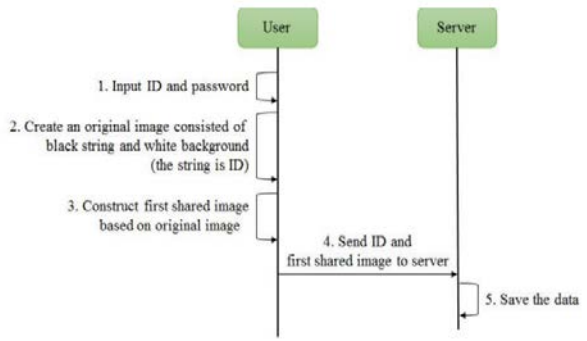

Figure 7. Initial registration process

First, user inputs the ID and password on device. The device starts to create an original image consisted of black letters implying ID and white background. The user may save the image in the device. The device constructs first shared image

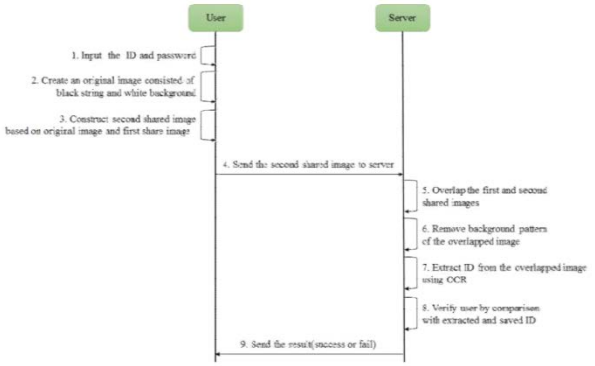

Figure 8. Process of enhanced password Scheme

adapting VC. The pattern to make up the first shared image is determined by pseudorandom generator with SEED which has password and ID as salts. After completing building the first shared image, the user sends ID of text type and the image instead of password to the server via security channel and destroys the image. If the server saves the data about user information, the initial registration process is finished. The server does not know the password at all because it is impossible for server to retrieve the password of the user from just one shared image.

Proposed password processing scheme is as follows:

1) The user inputs the ID and password.

2) The device of user creates an original image composed of black characters and white background. If the saved original image exists on user's device, it dose not have to create the original image again.

3) Although the device does not possess the first shared image, it can thoroughly construct second shared image refered to the original image and first shared image because the device already knows the SEED to make up the first shared image.

4) The user sends the second shared image only to the server.

5) The server overlaps the first shared image saved and the second shared image received.

6) The server should remove the background of the overlapped image as in Figure 3 (d), to gain original image.

7) ID is retrieved from the background-removed image by OCR .

8) The server confirms whether the extracted ID corresponds with saved ID, and determines success or fail.

9) The result is sent to the user.

B. System Implementation

We developed proposal scheme-based application for communication between user and server on internet. It is installed in the devices of user and server as in TALBE II. Suppose that user run on android such as Nexus 7, because this paper wants to show that proposed scheme can be adapted on a machine even with lower spec than general desktop. Operating system is installed on each machine. The device of user part uses Android 4.0. The server with Window7 has static IP(last number is 75) and 9002 port. We import basic java library as well as "java.io and java.net" for networking programming, "java.awt" to manage sockets for networking and "javax.imageio" to conduct images on VC. Especially the server has to import Tesseract API downloaded from Git in order to derive user's ID from stacked images after removing background pattern.

TABLE I. SYSTEM ENVIRONMENT

Part	*User*	*Server*
Development Type	Application	application
Device	Nexus 7	General desktop
OS	Android 4.0	Window 7
IP	~	xxx.xxx.xxx.75
Port	~	9002
Java version	JDK 7	JDK 7
Main Library	java.io.* java.net.* java.awt.* javax.imageio.* java.util.*	java.io.* java.net.* java.awt.* javax.imageio.* java.util.* Tesseract API

IV. EVALUATION

Our scheme has a few differences from traditional password-based scheme. The first is the adopting VC instead of text-based hash. The second is that the output value is user's ID even if input value is password and salt as in traditional scheme. The last is that user sends only one image involving the ID and password for authentication. Based on these features, our proposal has advantages as follows:

Lower computational cost preventing cyber-attack using vulnerable points of hash functions supporting privacy of users By applying the peculiarity of VC, suggested scheme also has identical peculiarity. VC requires little computation to create random pattern number per pixel for encryption. Random number generator has lower computation complexity than hash function because a

pseudorandom number is obtained just by repeating exclusive-or (XOR) operator with a shifted version[11].

Secondly, this scheme is able to prevent cyber-attack such as dictionary-attack and birthday-attack from the attackers aiming at cracking hash values. Even though the attacker extorts saved image, it is impossible for the attacker to acquire any information about original password or rule to array subpixels. Actually even as the shared image is expanded, it seems like mosaic as in Figure 9. Even if the attacker knows that the image is built by repeating some shapes with regard to subpixels, he or she can never understand the rule to match the shapes with pattern number and the rule to generate pseudorandom number. The dictionary for VC is not able to exist because shared image size is very diverse different from static hash size, and it is more difficult to search the information by image than by text.

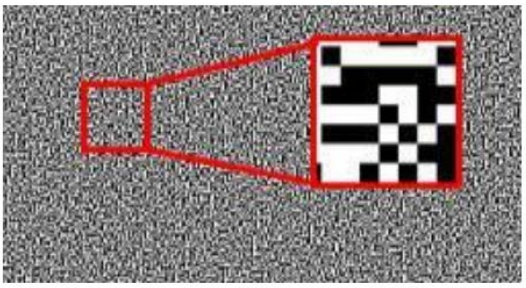

Figure 9. Expansion of shared image

Lastly, this scheme supports privacy for user. The server saves only one shared image instead of the password and receives another shared image not to expose ID from user. As a result, no information of user such as ID or password is revealed in each shared image discriminated from conventional scheme.

V. CONCLUSTION

Many people use the same or short length of passwords in multiple systems and are neglectful password management. Consequentially cyber-accidents are occurred often. We suggested a distinctive method different from conventional password scheme. It is based on encoded images by VC with a SEED number and OCR and more strong protection from cyber-attacks. We evaluated security aspect on attacks, computational cost and privacy. Our proposal is light weight and more secure in the aspect that hashed values of important information are not stored in the system.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2016R1A2B4015899).

REFERENCES

[1] Gaw, Shirley, and Edward W. Felten, "Password management strategies for online accounts," Proceedings of the second symposium on Usable privacy and security. ACM, 2006.

[2] Nguyen, Thi Thu Trang, and Quang Uy Nguyen, "An analysis of Persuasive Text Passwords, "Information and Computer Science (NICS), 2015 2nd National Foundation for Science and Technology Development Conference on. IEEE, 2015.

[3] Tam, Leona, Myron Glassman, and Mark Vandenwauver, "The psychology of password management: a tradeoff between security and convenience, "Behaviour & Information Technology 29.3 (2010): 233-244.

[4] Wang, Luren, Yue Li, and Kun Sun, "Amnesia: A Bilateral Generative Password Manager," 2016 IEEE 36th International Conference on Distributed Computing Systems

[5] Gauravaram, Praveen, "Security Analysis of salt|| password Hashes," Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on. IEEE, 2012.

[6] Dana Yang, Inshil Doh, Kijoon Chae, "Mutual Authentication based on Visual Cryptography and OCR for Secure IoT Service," Source of the Document 2016 6th International Workshop on Computer Science and Engineering, WCSE 2016, 2016, Pages 214-219

[7] M. Naor and A. Shamir, "Visual Cryptography," Advances in Cryptology EUROCRYPT94 LNCS, Vol. 950, pp. 1-12, 1995.

[8] Mori, Shunji, Ching Y. Suen, and Kazuhiko Yamamoto, "Historical review of OCR research and development," Proceedings of the IEEE 80.7 (1992): 1029-1058.

[9] Patel, Chirag, Atul Patel, and Dharmendra Patel, "Optical character recognition by open source OCR tool tesseract: A case study," International Journal of Computer Applications 55.10 (2012).

[10] Holley, Rose, "How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs," D-Lib Magazine 15.3/4 (2009).

[11] Marsaglia, George, "Xorshift rngs," Journal of Statistical Software 8.14 (2003): 1-6.