A new graphical password scheme against spyware by using CAPTCHA

Haichang Gao, Xiyang Liu, Sidong Wang
Software Engineering Institute
Xidian University, Shaanxi 710071, P.R.China
{hchgao, xyliu}@xidian.edu.cn

Ruyi Dai
School of Computer
Northwestern Ploytechnical University
Shaanxi 710072, P.R.China

1. INTRODUCTION

The security and usability problems inherent in text-based password schemes have resulted in the development of graphical password schemes as a possible alternative. However, most of the current graphical password schemes are vulnerable to spyware which is a program that gathers information about a computer's use and relays that information back to a third party [1].

To date, there have been some schemes which have made contributions to the development of graphical password in term of spyware resistance [2, 3]. Using a challenge-response protocol, they have an advantage in that they are resistant to replay attacks. Namely, even the third party who observes a successful login session cannot perform a replay attack. Though they have a positive effect on protecting users' password, they are not yet sufficient to stop attackers from harvesting passwords.

In this paper, CAPTCHA is used in a graphical password scheme to resist spyware. A CAPTCHA (Completely Automated Public Turing tests to tell Computers and Humans Apart) is a program that generates and grades tests that are human solvable, but are beyond the capabilities of current computer programs [4]. CAPTCHA uses open algorithms based on hard AI problems, and has been discussed in text-based password schemes to resist dictionary attack [5]. Innovatively, we explore CAPTCHA in the context of graphical passwords to provide better protection against spyware. As long as the underlying open AI problems are not solved, CAPTCHA is a promising way to resist spyware attack in graphical password schemes.

Based on this key idea, we have proposed a new graphical password scheme using CAPTCHA, designed to be strongly resistant to spyware attack, either by purely automated software or via human participation. A preliminary user study indicates that our scheme needs to improve in terms of login time and memorability.

2. THE PROPOSED SCHEME[1]

In our proposed scheme, users are allowed to select their own graphical password images (pass-images). To be authenticated, the user only needs to distinguish his/her pass-images from decoy images and then enter certain parts of the CAPTCHAs string below the pass-images. The CAPTCHA is an image of distorted string randomly generated by system. In the example shown in Figure 1, the three images in circle are the pass-images, the positions selected by user is (4, 6, 8), (1, 2, 4), and (3,5). The CAPTCHAs for each pass-image is 'oygewdsy', 'gcsmcwcz', and 'xgsvyeqg', thus the input sequences for each pass-image respectively are 'edy', 'gcm', and 'sy', and user can input any one combination of the three sequences to authenticate.

Figure 1: A part of screen for the proposed scheme

3. SECURITY ANALYSIS

3.1 Capability to withstand spyware

The security provided by our scheme relies on the property of CAPTCHA that it is hard for machines to recognize. Although the spyware can gather the information of user's every login containing the strings the user entered and the images shown on screen, the spyware cannot recognize CAPTCHAs, so it cannot ensure which CAPTCHAs correspond to the entered string and expose the pass-images.

User's every inputting will reveal a small amount of information about the password. For example, if a CAPTCHA does not contain any character obtained from the entered string, the corresponding image can be excluded. Thus, when the characters forming strings are randomly selected from a database of size of A,

Symposium On Usable Privacy and Security (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.

the possibility that there is no character which belongs to both of two strings is:

$$\frac{\sum_{i=1}^{M}((A-i)^{L}\cdot C_{A}^{i}\cdot a_{i})}{A^{L}\cdot A^{M}} \quad (1)$$

Here, $a_i = i^M - C_i^1 a_1 - C_i^2 a_2 - \cdots - C_i^{i-1} a_{i-1}$ ($a_1 = 1^M = 1$) means the number of all combinations of string which contains just *i* different characters, while *L* and *M* mean the lengths of the two strings. When *A*=26, *L*=8 and *M*= 8, the possibility is about 8.5%. At this situation, to exclude the decoy images as much as possible, dozens of user's logins are needed. Therefore, when the parameters are increased, dozens or hundreds of user's logins are needed and analyzed to get correct pass-images and pass-positions. Collecting information from a user's logins may take a period of time, and taking a screenshot greatly increases the spyware's resource consumption, hence the risk of discovering the spyware will increase and harvesting of passwords becomes more difficult.

Our proposed scheme is strongly spyware resistant to purely automated attacks. Furthermore, human intervention in spyware is not a realistic means of attack as it is exceptionally time consuming to break the scheme.

3.2 Password Space

Assuming users are equally likely to pick any element as their password. According to the definition in [1], the raw size is an upper bound on the information content of the distribution that users choose in practice.

We compute the size $S(L,N)$ of password space of total entered length equal to *L* when there are *N* images displayed. In our scheme, the number of pass-images is required to be more than three for security. Thus, *S* is defined in terms of $O(K,L,N)$, the number of passwords with number of pass-images equal to *K*:

$$S(L,N)=\sum_{K=3}^{L} C_N^K \cdot O(K,L,N) \quad (2)$$

According to the theorem of the partition of positive integer, the generating function of sequence of partition numbers is $\left(\frac{x}{1-x}\right)^K$, and there are $C(L-1,K-1)$ different partition situations. Assume the number of pass-positions for one pass-image is *n*, $O(K,L,N)$ can be defined in terms of *n* by:

$$O(K,L,N)=\sum_{i=1}^{C_{L-1}^{K-1}}\left(\prod_{q=1}^{K} C_M^{n(q,K,L,N)_i}\right) \quad (3)$$

Here, *M* is the length of CAPTCHAs used in our scheme.

Putting the pieces together, we can compute the size of the password space. The results for the password space are given in Table 1 (*N*=50 and *M*=8).

The data in Table 1 are encouraging. However, that is the raw size of our password space, in practice not all passwords are equally likely to be chosen by users. However, the effect of user choice on the practical space has not been considered.

Table 1. Password spaces of Length *L*

L	4	6	8	10
$\log_2$(#)	30.0	42.3	53.8	64.5

4. EXPERIMENT AND RESULTS

The mean login time of a preliminary experiment is 18.6 seconds which is acceptable for most participants. The results show that there is a significant difference in time to respond to a challenge ($F(19,197)=1.97$, $p<0.05$). As the CAPTCHA images are randomly located, the time for recognition is different.

Insecure user behaviors were noted in the experiment, creating risks for our scheme. First, the passwords selected by user often accord with a particular trend. Second, there is always a significant time gap when entering characters belonging to two different pass-images.

5. CONCLUSION

In this paper, we have presented a new approach to protect user's password against spyware attack. Our main contribution is that we introduce CAPTCHA into the realm of graphical passwords. From the security viewpoint, this exploration is expected to advance the development of graphical passwords. While the design of CAPTCHA is an interdisciplinary topic and the current collective understanding of this topic is still in its infancy, we do not claim that our scheme is definitely feasible immediately. But, as long as the state-of-art-algorithms cannot solve the hard AI problems, it is probable to construct a graphical password scheme with CAPTCHA that is strongly resistant to spyware.

The results of our experiments show that the login time and memorability is not ideal, which indicates an area for further research. Additionally, narrowing the time gap in the entering process and reduction of the impact of user's choice trend on security, provide areas for future research.

6. REFERENCES

[1] I. Jermyn, A. Mayer, F. Monrose, M.K. Reiter and A.D. Rubin. The design and analysis of graphical passwords. In Proceedings of the 8th USENIX security symposium, 1999.

[2] D. Weinshall. Cognitive Authentication Schemes Safe Against Spyware. In Symposium on Security and Privacy, 2006

[3] D. Hong, S. Man, B. Hawes, and M. Mathews. A graphical password scheme strongly resistant to spyware. In Proceedings of International conference on security and management. Las Vergas, NV, 2004.

[4] L. von Ahn, M. Blum, and J. Langford. Telling Humans and Computer Apart Automatically. Communications of the ACM, 2004, 47(2), pp.57-60.

[5] B. Pinkas, and T. Sander. Securing passwords against dictionary attacks. In Proceedings of the ACM Computer and Security Conference, 2002, pp.161–170.