IMAGE AUTHENTICATION SCHEMES AGAINST KEY-LOGGER SPYWARE

M. N. Doja and Naveen Kumar
Department of Computer Engineering,
Jamia Millia Islamia, New Delhi, India.
Email: *ndoja@yahoo.com*,
naveenkumar@rediffmail.com

***Abstract*- Spywares has become major problem now days. This type of software may track user activities online and offline, provide targeted advertising and/or hold in other types of uninvited activities. Password collection by spywares is increasing at a shocking pace [1]. The problem of entering sensitive data, such as passwords, from an untrusted machine, is obviously undesirable, however roaming users generally have no other option. They are in no point to review the security status of Internet cafe or business center machines, and has no alternative to typing the password. We consider whether it is possible to enter data to confuse spyware assumed to be running on the machine in question. The difficulty of mounting a collusion attack on a single user's password makes the problem more tractable than it might appear. This problem of password security can be improved by biometric based authentication and graphical authentication, however availability and cost of biometric authentication is considerable problem. In this paper, we present an alternative user authentication based on Images that is resistant to keylogger spywares. We have design and implemented a method that uses a strengthened cryptographic hash function to compute fast and secure passwords for arbitrarily many accounts while requiring the user to memorize only few memorable points in the image. In addition to keylogger spywares our design is also highly resistant to brute force attacks and prone to Dictionary attack, allowing users to retrieve their passwords from any location so long as they can execute our program and remember a short secret. This combination of security and usability will attract users to adopt our scheme. This paper will be useful for information security researchers and practitioners who are interested in finding an alternative to spyware resistant user authentication.**

I. INTRODUCTION

Spywares has become serious threat to computer security. According to the pew Internet & American life project (PIP) survey 50% of Internet users see software programs like spyware as a serious threat to their online security [2]. The term spyware first came into use in 1995; today spyware is a serious and persistent problem that none of the known internet-security technologies like firewalls or anti-viruses can address fully. Anti-spyware technology was first introduced in 2000, but the surge of newer anti-spyware solutions continues even today, it give a clear indication of the commonness of Spyware as an ever growing problem.

There are different ways through which spyware can enter into a computer system for example as a software virus or as the result of installing a new application. It can impair the operation of computers, causing them to crash and interfering with the ability of consumers to use them.

Spyware programs often cause significant degradation in system performance. Spyware can even cause computers to crash. Microsoft reported that 50% of its customers' computer crashes are traceable to spyware [3]. Spyware may use so many system resources that users are no longer able to use their keyboard, mouse, and their cursors freeze. Spyware can impair the operation of computers, causing them to crash and interfering with the ability of consumers to use them. Spyware programs often cause significant degradation in system performance.

There was general agreement that spyware can assert control over the operation of computers in ways that substantially limit the ability of consumers to use their computers. For example, some spyware programs change users' browser setting, which is often referred to as "browser hijacking." Spyware may change the web page displayed when the browser first opens, i.e., the home page, and frustrate efforts to replace that home page with the user's original home page. Spyware may also insert links to its own websites into the user's "Bookmarks" or "Favorites" list. One spyware program, for instance, intercepts search queries sent to Google, a popular search engine, and then displays its own search results. The search results appear to be from Google but contain links to pornographic websites that would not have appeared with an actual Google search. Moreover spyware risks are not limited till the degradation and harming computer performance and operations. Spywares can also, steal the users personal information and do tracking of an users online activity. The most serious privacy risks arise when spyware installed on a computer includes password hijackers or keylogger.

A keylogger captures all keystrokes that the user types on the computer keyboard, including passwords, personal information entered into an online registration form (e.g., a mailing address or telephone number), financial information submitted as part of an online transaction, and the contents of emails or instant messages. One can have firewall installed in a computer, however normally firewalls are designed to block specific kinds of threats and look only at certain attributes of incoming transmissions, much like the post office looks only at the addresses or attributes on a letter, but does not look at, or attempt to evaluate, the letter's content. Some of the major spyware categories are adware, malware, keyloggers, browser helper objects, worms, trojans, password hijackers, E-mail

978-0-7695-3263-9/08 $25.00 © 2008 IEEE
DOI 10.1109/SNPD.2008.166

flooders, firewall killers, spoofers, hacking tools, dialers, tracking cookies, remote administration tools, backdoors and annoyance tools. We have focused mainly on the password hijackers and Keylogger spywares as these are the most insidious threats to a user's personal information. Passwords, credit card numbers, and other sensitive or personally identifying information are potentially exposed.

II. ALTERNATIVE USER AUTHENTICATION

Most of the applications are based on text based password entry for user authentication, however there are some other promising solutions for user authentication like image-based authentication and biometric authentication.

All user authentication schemes are based on three fundamental pieces of information: what you know, what you have, and who you are [4] which, also corresponds to token-based authentication, knowledge-based authentication and biometric authentication. For proving who they are, users can provide their name, email address, or a user ID. Since this information provides no assurance of identity, some system operators are beginning to employ biometrics (such as fingerprints, voice recognition, iris scans, or retinal scans) as methods of user identification. For proving what they have, users can produce service cards (i.e., ATM cards), physical keys, digital certificates, smart cards etc) [5]. For proving what they know, users can provide a password or pass phrase, or a personal identification number (PIN). This information is essentially a secret that is shared between the user and the system. Knowledge based techniques are the most widely used authentication techniques and include both alphanumeric and graphical authentication.

In graphical authentication, system use one or several images to authenticate a user rather than typing a password. Biometric authentication techniques are further categorised into Physiological and Behavioural based schemes, such as fingerprints, voice recognition, iris scan, or facial recognition are not yet widely adopted. The major drawback of this approach is that such systems can be expensive, and the identification process can be slow and often unreliable. However, this type of technique provides the highest level of security. Recently, efforts have been made towards graphical authentications schemes, which are resistant to password hijackers and keylogger spywares and prevent stealing of users password.

III. IMAGE/GRAPHICAL PASSWORD AUTHENTICATION

Alphanumeric passwords also have drawbacks, most notably in terms of memorability and security like hacking of password. This has led to innovations to improve these password schemes. The underlying idea is that, using images will lead to greater memorability and decrease the tendency to choose insecure passwords because human's ability of visual memory is much more powerful than the textual memory [6].

The first idea for Image/graphical passwords was explained by Blonder [7]. His approach was to let the user click, with a mouse or stylus, on a few selected regions in an image. If the correct regions were clicked-on the user is authenticated, else the user was discarded. According to Blonder graphical password scheme, only pre-processed images can be used; the click regions can only be chosen from certain pre-designed regions in the image. This implies that the users cannot provide images of their own for making passwords, and users cannot choose click places that are not among the pre-selected ones. Some similar schemes are being proposed like Passlogix [8] has developed a graphical password system where, users must click on various items in the image in the correct sequence in order to be authenticated. Invisible boundaries are defined for each item in order to detect whether an item is clicked by mouse.

One alternative scheme is proposed in [9] which accepts user input of alphanumeric password can be entered through virtual keyboard which accepts the input from mouse however this approach lead to the memorability problem and in effect harms the security. Another technique "PassPoint" system on the same line of graphical password has been developed by Wiedenbeck, et al. [10] extended Blonder's idea by eliminating the predefined boundaries and allowing arbitrary images to be used. As a result, a user can click on any place on an image (as opposed to some pre-defined areas) to create a password.

IV. AN ALTERNATIVE IMAGE AUTHENTICATION

Our design allows the use to choose any images by eliminating the predefined boundaries including the users own images, digital photos of landscapes, paintings, etc. Moreover, we let users choose any places that attract them as click regions; such places are easier to remember. However, allowing arbitrary click locations lead to a stability problem, which we had to overcome. The problem is that we cannot expect users to click always on exactly the same location Calculating a tolerance around each chosen pixel area can improve the range of user to select the same chosen point but to improve the security of the system, the selected password or pixels must be stored in the hash form instead of plain form, and hashing does not allow approximation: two passwords that are almost (but not entirely) identical will be hashed very differently. Hence in our approach we have partitioned the image into squares grid, which can be adjusted by the user e.g. 10x12, 20x20 grid squares etc.

We display the square grids to the user, this eliminate the possibility that a user may choose a click point that happens to be close to an edge of a partitioned square grid. In this system user would create his or her portfolio from the set of images that are generated by dividing the user's own image or any other image into number of grid squares and each of the squares would represent an independent image.

User may choose any image stored in the database or can select any image from his or her private database, further user would be asked to select few grid squares out of it by clicking at various portions of the image. Those grid squares are passed through secure one way hash function MD5 which will generate 128 bit fixed length unique output, this output will be stored in the password file and will act as a password for the

user in future. The hash should also depend on the secret key, which will be the click points.

Furthermore, the hash should not be easily forged or estimated without the knowledge of the key. Although they are very secure, these hash functions are not robust as they are very sensitive to every bit of the image data. This is undesirable and inconsistent with human visual perception [11]. As a result, we proposed the grid-based evaluation to compute the hash of the password clicks. This MD5 algorithm will take the pixel values of the image grid squares selected by the user as a input and will produce a 128 bit unique value which will be stored in the password file.

Once the user has chosen the image grid squares, those grid squares sequence would become his or her password for entering the system. When the user tries to log on next time, after entering the user name he or she will be presented with the same image that was used for password creation. User will be prompted to select the same grid square sequence, which was selected for password creation; if the user selects the same sequence then he or she is allowed to log on else not.

A. Algorithm of Proposed Image Password Scheme

This algorithm will take the pixel values of the image grid squares selected by the user as a input and will produce a 128 bit unique value which will be stored in the password file. The algorithm has following steps by which it generates the input values for MD5 by reading the grid squares selected by the user from the image.

a) Select an image. (Image with large number of memorable points should be preferred)

b) Divide the whole image into equal sized square grids as shown in figure 1 given below. The users according to their visibility can adjust the Grid size.

Fig. 1. Divide the whole image into equal sized square blocks.

c) As soon as the first click takes place on the image, start recording the grid pixels and the sequence in which user has selected the grid square images.

d) 8 bits each for Red, Green, Blue and one more component, which is luminance, is recorded for each pixel.

e) Link list is formed in which each node will have data for one grid square image.

f) Recording continues till the user clicks to finish entering password. The selected grid square images are represented with black-bordered squares as shown in figure 2 given below.

Fig. 2. Chosen image password indicated by red-bordered blocks

g) The whole link list is passed to the temporary buffer where padding and appending of length is done before sending it as an input for MD5.

h) Final password of the selected image grid squares is selected. Password file consists of the sequence of the values generated by MD5 without mentioning the user name.

i) This password will be compared to the list of password stored in the password file and if any matching is there login will be successful else not.

V. SECURITY ANALYSIS

There are many aspects of security analysis like evaluating a system against common password security issues, the main issues we have focused are brutal force attack problems which is directly dependent on password space, dictionary attack and Keylogger program that runs in the background, recording all the keystrokes. In this paper we have analysed our scheme against these security problems.

B. Keylogger Spywares

Keylogger software cannot be used to break image-based authentication. Also, mouse-tracking spyware will not be an effective tool against image-based authentication. As, mouse motion alone is not enough to break image-based authentication. Such information has to be correlated with application information, such as window position and size, as well as timing information.

C. Brute Force Attack

The main defense against brute force search is to have a sufficiently large password space. In Table: 1, we compared the password spaces of our scheme with that of alphanumeric passwords, for various parameter settings. The password space is the set of all passwords that are possible for a given

password scheme and for a given setting of parameters. For example, for alphanumeric passwords of length eight over a 64-character alphabet, the number of possible passwords is $64^8=2.8\times10^{14}$. In our image password scheme, as we tested it in our empirical study, with image size 451×331 and grid square size 20×20 (all measured in pixels) there are about 451×331/20×20=373 grid squares; hence, for passwords consisting of five click points, the password space has size $373^5=7.2\times10^{12}$. With the same settings, but with six click points, the password space has size $373^6=2.69\times10^{15}$. If in our graphical password scheme the image size is 1024×752, with grid square size still 20×20, and with passwords consisting of five clicks, the password space will have size 2.6×10^{16}. However, the size of the password space is not the only thing that matters. The usability and memorability of passwords are just as important.We see from these comparisons that for just five clicks, and for reasonably sized grid squares, graphical passwords have a larger password space than alphanumeric passwords. Moreover, one could easily increase the number of click points to six or decrease the grid square (e.g. from 20×20 to 14×14). Indeed, with around 10 click points our password space is comparable in size to a cryptographic key space. Thus, another contribution of this design is a large password space with a small number of clicks. For our experiment we used a smallish image of 451×331 pixels, because we needed room for buttons and text on the side of the image; thus we had to restrict the size of our password space, for the sake of the experiment. Even in that limited setting our password space is close in size to the space of random alphanumeric passwords of length eight, over a standard 64-character alphabet

TABLE I COMPARISON OF PASSWORD SPACE FOR ALPHANUMERIC PASSWORDS AND IMAGE PASSWORD SCHEME WITH DIFFERENT PARAMETERS

Password Scheme	Image size	Grid square size (pixels)	Alphabet size/ No. Squares	Length/No. Click points	Password space Size
Alphanumeric	N/A	N/A	64	8	2.8×10^{14}
Alphanumeric	N/A	N/A	72	8	7.2×10^{14}
Alphanumeric	N/A	N/A	96	8	7.2×10^{15}
Graphical	345×313	30×30	120	8	4.3×10^{16}
Graphical	451×331	20×20	373	5	7.2×10^{12}
Graphical	1024×752	20×20	1925	5	2.6×10^{16}
Graphical	1024×752	14×14	3928	5	9.3×10^{17}
Graphical (1/2 Screen used)	1024×752	14×14	1964	5	2.9×10^{16}

It is more difficult to carry out a brute force attack against graphical passwords than text-based passwords. The attack programs need to automatically generate accurate mouse motion to imitate human input, which is particularly difficult for image-based authentication. As password space size achieved by alphanumeric passwords of alphabet size (96) is 2.8×10^{14} in comparison to graphical passwords using image (1024x752) 9.3×10^{17}, which is better than the alphanumeric and attacker have to perform more computation to find the password. Overall it is found that image-based authentication is less vulnerable to brute force attacks than a text-based password.

D. Dictionary attacks

Since image-based authentication involve mouse input instead of keyboard input, it will be impractical to carry out dictionary attacks or automated dictionary attack will be much more complex than a text based dictionary attack. Dictionary attacks are infeasible, to a certain extent because of the large password space, but primarily as there is no pre-existing searchable dictionary used for graphical information. It is also not easy to devise automated attacks. Whereas we can recognize an image in less than a moment, computers spend a great amount of time processing millions of bytes of information. However, more research is needed in this area for developing such graphical dictionaries and testing the image-based passwords.

VI. SYSTEM EVALUATION

An evaluation was conducted on proposed image authentication system to check the visual perception and memorability of the user about the images. In this forty experienced computer users participated in the study. A single PC with a high-resolution nineteen-inch monitor was used in the experiment. Participants were randomly assigned to the graphical or alphanumeric condition.

Each individual participated in three sessions. The first session lasted about 35 minutes. First, the participants were given explanation of procedures of the experimentation then they chose a graphical password. Graphical password users had to select and enter five different points on the picture. They were told that they would have to remember the points and the order in which they were input. Alphanumeric users had to enter eight characters including at least one upper case letter, one special character and one digit. They were also told not to choose a password they had already used, excluding their personal names etc. This was done because it is best practice standard for alphanumeric passwords [12]. The system enforced that the participants re-enter the password until they chose a valid password.

A graphical password of eight or five points was used based on our analysis (Table I), which shows that in terms of security five click points provide a password space as

large as or larger than an alphanumeric password of eight characters. When the participant had created a valid password, the password was displayed as feedback to the participant before going on to the next phase. The evaluation was separated mainly into two phases leaning and retention phase.

In the learning phase the participants entered the password repeatedly until they achieved ten correct password inputs. They received binary feedback on the correctness of each password input. In the retention phase, password retention was measured three times: at the end of the first session (S1), one week later (S2), and four weeks later (S3). In these retention trials the participants had only to enter their passwords correctly one time. If the participant entered an incorrect password, the system instructed the participant to re-enter the password.

A. Learning Phase

After choosing a password, participants practiced their password in the learning phase. The criterion for success was 10 correct logins. The participants continued to input the password until the criterion was met. We measured the number of incorrect password submissions and the total time spent in practice. Table II shows the means and standard deviations.

TABLE II MEANS (STANDARD DEVIATIONS) OF NUMBER OF INCORRECT SUBMISSIONS AND TOTAL PRACTICE TIME IN THE LEARNING PHASE (ALPHANUMERIC N=20/GRAPHICAL N=20)

	Mode	Mean (SD)
Number of incorrect submissions	Alphanumeric	0.40 (0.68)
	Graphical	4.80 (7.16)
Total practice time (seconds)	Alphanumeric	66.08 (4.92)
	Graphical	171.89 (24.46)

In the learning phase the alphanumeric group took fewer trials to achieve 10 correct password inputs than did the graphical group. This is also reflected in significantly longer total times to input the graphical passwords. Seventy percent of the alphanumeric participants input the password 10 times without any errors, and all alphanumeric participants were able to achieve the criterion with a maximum of two incorrect password inputs.

The graphical group took more trials and had more variability. Forty percent of graphical participants achieved input of the password 10 times without any errors, and 70 percent achieved the criterion with a maximum of three incorrect password inputs. Surprisingly, the least successful twenty percent of the group made between 17 and 20 incorrect password inputs. However. It should be clearly noted that most graphical participants did not have serious problems in the learning phase.

B. Retention Phase

In the retention phase participants input the password longitudinally three times. In each retention trial the participants had to enter their password correctly only one time. The number of incorrect password submissions and time for the correct submission are shown in Table III.

TABLE III MEANS (STANDARD DEVIATIONS) OF NUMBER OF INCORRECT PASSWORD SUBMISSIONS AND TIME FOR THE CORRECT SUBMISSION (ALPHANUMERIC N=20/GRAPHICAL N=20)

	Mode	Mean S1 (SD)	Mean S2 (SD)	Mean S3 (SD)
Number of incorrect submissions	Alpha-numeric	0.25 (0.79)	2.20 (2.73)	1.75 (2.47)
	Graphical	1.55 (1.57)	2.75 (3.88)	1.50 (2.80)
Time for correct submission (seconds)	Alpha-numeric	5.23 (1.66)	9.42 (3.70)	9.24 (3.72)
	Graphical	8.78 (4.40)	24.25 (15.21)	19.38 (17.57)

In the retention phase, the accuracy of password inputs differed by trial for both groups. Since the S1 trial took place in the same session as the creation and learning phases, there were few bad inputs. In the S2 trial participants had more difficulty recalling their passwords, regardless of which group they were in. In the final S3 trial there appears to have been some consolidation of the passwords in memory because the incorrect inputs were lower than in S2 (though not significantly) in spite of the long time lapse. The lack of significant differences between the alphanumeric and graphical modes on the correctness of password inputs and lack of interactions between mode and trial, indicate that the main factor in correctness was password memory for both groups. It is encouraging that the graphical group was as to do as well or better than the graphical group, in their first experience with remembering image-based authentication.

VII. CONCLUSION

Generally computer systems access is based on the use of alphanumeric passwords. However, spread of spywares have created major threat on hacking the user sensitive information like password, credit card number, online banking account etc. The image based user authentication is highly resistant to keylogger spywares and difficult to hack. Also, our scheme has better password space over alphanumeric passwords and reduces the brute force attack of passwords. Similarly, image authentication has an advantage in password space over Blonder-style graphical passwords in term of retention of password. Image authentication that makes passwords more memorable and easier for people to use and, hence, it is more secure. Specifically, we proposed an image authentication, which uses hash function to store the password point on the images in the form of pixels. It is recommended to use well-featured images, as it will give more variation in the pixel values, moreover well featured images are quite easier to recognize. In addition, Image based authentication matches the capabilities and limitations of most handheld devices and provides a simple and intuitive way for users to authenticate. Besides user authentication, image passwords may also be used in other security applications where conventional passwords have been used traditionally.

REFERENCES

[1] White Paper, "Combating the Spyware menace: Solutions for the Enterprise", *London, United Kingdom*, http://www.omniquad.com/, Accessed January 2008.

[2] Susannah Fox, "Public Policy Spyware: The threat of unwanted software programs is changing the way people use the Internet", *Pew Internet and American Life Project*, July 2005, http://www.pewinternet.org/PPF/r/160/report_display.asp, Accessed January 2008.

[3] Tim Johnson, "Spyware is a Blended Threat: Your security demands a layered approach", *White paper*, September 2005, www.surfcontrol.com, Accessed January 2008.

[4] J. Thorpe, and P.C. Oorschot, "Towards secure design choices for implementing graphical passwords", *ACSAC '04: Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC'04)*, Washington, DC, USA, Vol. 3, pp. 664 – 666, 2004.

[5] I. Jermyn, A. Mayer, F. Monrose, M.K. Reiter, and A.D. Rubin, "The design and analysis of graphical passwords", *Proceedings of the Eighth USENIX Security Symposium*, pp. 1–14, 1999.

[6] D. Bensinger, "Human memory and the graphical password", http://www.activetechs.com/solutions/security/sso/bensinger.pdf. Accessed January 2008.

[7] Blonder, G.E., 1996. Graphical passwords. United States Patent 5559961.

[8] Passlogix,V-Go, www.passlogix.com, Accessed January 2008.

[9] M.D. Fleetwood, M.D. Byrne, P. Centgraf, K. Dudziak,B. Lin, and D. Mogilev, "An analysis of textentry in Palm OS: Graffiti and the Virtual Keyboard".*Proc. HFES 46th Annual Meeting , Santa Monica: HFES*, 2002, pp. 617-621.

[10] S. Wiedenbeck, J. Waters, J.C. Birget, A. Brodskiy, and N. Memon, "Authentication using graphical passwords: Effects of tolerance and image choice", *in Symposium on Usable Privacy and Security (SOUPS)*, at Carnegie-Mellon Univ., Pittsburgh, 6-8 July 2005.

[11] Rachna Dhamija and Adrian Perrig, "Déjà Vu: A User Study Using Images for Authentication", *9th Usenix Security Symposium*, August 2000.

[12] Robert Morris and Ken Thompson, "Password Security: A Case History", *Communications of the ACM*, 22(11), pp. 594-597, November 1979.