Middle-East Journal of Scientific Research 24 (10): 3125-3130, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.3125.3130

Invisible Captcha Using Encrypted Virtual Keyboard

[1]S. Thangavelu, [2]T. Purusothaman and [3]G. Gowrison

[1]Research Scholar, M.S.University, Tirunelveli, India
[2]Asso. Prof / CSE, Government College of Technology, Coimbatore, India
[3]Asst. Prof / ECE, Institute of Road and Transport Technology, Erode, India

Abstract: Captcha is acronym for completely automated public Turing test to tell computers and humans apart, intended to discriminate humans from automated programs called as bots. Captcha differentiates humans and bots by setting a task that is easy for humans to complete but more difficult for bots to implement. Thus Captcha confirms the user on website, is a human and not a computer bot. However, most of the text based Captcha suffers from OCR software attacks, in which the bots acquire Captcha image from signup page of the website and reproduce in the text box for authentication. This can be prevented by the proposed Invisible Captcha approach. The Invisible Captcha is based on the concept 'What you see is not true', designed substantially to protect Captcha, from OCR attacks. The OCR programs are capable to read, only the visible Captcha on the Login page. But the proposed approach generates an invisible Captcha through encrypted virtual keyboard which is not visible to anyone. By cognition ability the humans only can complete this grading test and get authenticated. As a result Invisible Captcha enhances the security of web applications.

Key words: OCR · Virtual keyboard · Cognition · Security · Authentication · Bots

INTRODUCTION

Captcha is a Human interactive proof [1], used to ensure the security of the web applications which distinguishes humans and computer bots automatically. Captcha protects web sites from the unauthorized entry of the bots and prevents from doing any malicious activities in the Internet applications. The various types of Captcha are [2],

- Text based Captcha
- Image based Captcha
- Video based Captcha
- Audio based Captcha

Text Based Captcha: The text based Captchas [3] are very popular, widely used and accepted method by all section of users because of its simplicity and user friendly. The proposed invisible Captcha method is originated from the existing text based Captcha methods only. The various text based Captcha methods are shown in Fig. 1.

Gimpy is one of the reliable text based Captcha which gives a real challenge to the users. Randomly chosen words from dictionary are displayed with distortion and overlapping. The users need to differentiate and identify the words in subset and enter in the textbox for authentication. Ez-gimpy [4] is a simplified version of Gimpy previously used by Yahoo, in their signup page; it uses one single word with distortion. In Pessimal print Captcha [5] a low quality printed text image with certain degree of distortion is displayed to the user. The user needs to enter the correct text for authentication, which is hard for the bots. In Baffle text Captcha [6] words not present in the English Dictionary are produced and presented to the user with some tilting to prevent any form of attacks. Microsoft MSN Captcha [7] uses eight characters including numbers in dark blue foreground and grey color background as Captcha characters. Wrapping and ripple effect is created to confuse the bots.

Captcha prevent and protect the websites from unauthorized access by automated programs, spammers, spoofers and search engine crawlers [8]. The prime advantage of text based Captcha are easy to generate and evaluate but experiences OCR attacks owed to poor design and security overlooks [9]. The proposed approach focuses the OCR attacks on text based Captcha and suggest an alternate approach to enhance the security. The Captcha design [10] should involve the following parameters:

Corresponding Author: S. Thangavelu, Research Scholar, M.S.University, Tirunelveli, India.

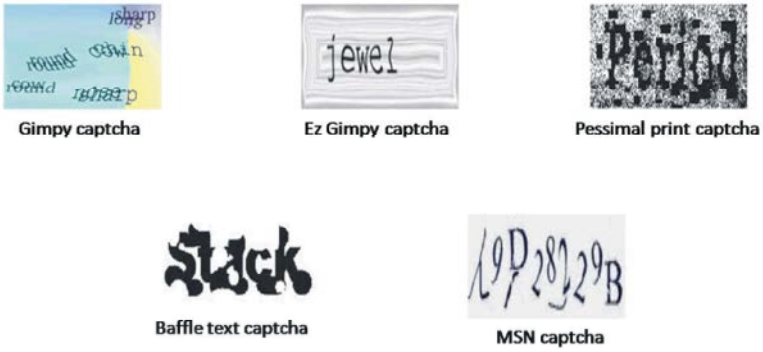

Fig. 1: Text Based Captcha methods

- Robustness: to resist the attack of malicious programs
- Usability: to increase the rate of passing the test by humans
- Scalability: measures the number of challenges that a Captcha scheme can generate without scarifying the robustness and usability

OCR Attacks: Optical Character Recognition [11] is an enhanced technology designed to convert the scanned images, PDF documents into editable text format. It is used to digitize lots and lots of documents within a short period of time. The process involves artificial intelligence, pattern recognition and computer vision. Hackers use sophisticated OCR to attack the Captcha images and perform malicious activities in the web services. The function of OCR involves [12],

- Image Acquisition: Captcha image is acquired from the Login page of a website by scanning the image.
- Segmentation: The process of separating the characters individually in the region of interest area is known as Segmentation. The OCR gets confused when superimposing or overlapping the characters each other. Right now, humans outperform computers in segmentation
- Pre processing: The noise introduced in the Captcha results the poor recognition of characters. The preprocessing involves smoothing and normalization which improves the character recognition.
- Feature Extraction: Features are the characteristics of individual characters in Captcha image. The extraction works based on the features present in a character such as intersections, lines open space etc and not the concrete character pattern.
- Recognition: This final process involves the identification of characters and numbers.

Most of the text based Captcha attacks by OCRs are known as blind attacks. The bots acquire the Captcha by scanning the image in the signup page of websites using OCR software. This can be prevented in the proposed Invisible Captcha approach, which involves human cognitive activities and encrypted virtual keyboard in Captcha design.

Related Work: Enormous amount of research has been in progress for the security of web applications, various research papers indexed the attacks on text based Captcha and the design of virtual keyboard.

Merrill Serrao *et al.* [13] exposed that most of the Captcha employed in popular websites has been broken frequently by software like Captcha sniper etc with the success rate of 28% to 100%. Aziz Barbar *et al.* [14] claimed that the most of the text based Captcha are broken with the success rate of Ez-Gimpy 92%, Gimpy 38%, Microsoft 60%, Others 49 to 100 % and suggested character image semantic approach as an alternate approach. By simple generic attack based on Log-Gabor filter method Haichang Gao *et al.* [15] attacked the text based Captcha engaged in the websites like Wikipedia, eBay etc with a maximum of 77% success rate. Jeff Yan *et al.* [16] conclude that the security of text based Captcha relies on segmentation resistance and proved that the attack on MSN Captcha by character segmentation techniques obtained success rate of more than 60%. Anjali Avinash Chandavale *et al.* [17] proved that the attack on text based Captcha is successful with the maximum rate of 97% by novel preprocessing, segmentation and recognition techniques. Oleg Starotenko *et al.* [18] proposed a segmentation process based on three color bar character encoding method and proved that the Google reCaptcha has been broken with the success rate of 82% to 95%.

Rajarajan *et al.* [19] proposed a new concept in virtual keyboard design. In which the keyboard are logically divided into four groups and the keys are randomized and not in any fixed positions. The keys can be shuffled clockwise, anticlockwise, crosswise etc. With hide key option to avoid shoulder surfing. Ankit Parekh *et al.* [20] proposed an anti screenshot keyboard, when cursor moves on a particular key, the entire keys in that row will be replaced by a special character so that the intruder software cannot make a screenshot of the key pressed. Andrea Bianchi *et al.* [21] proposed a haptic keyboard design which makes visually impossible for the observer to detect which keys are selected. A novel color keyboard has been introduced by Agarwal *et al.* [22] in which the alphabets and numerals are denoted in different colors. All the keys are shuffled each and every time after the user clicks a key. Hide Key option is provided to avoid any screenshots.

Nairit Adhikary *et al.* [23] propose a two factor authentication for the untrusted zone computers. A fabricated password and an USB device with Unique ID are used through the onscreen virtual keyboard for authentication in untrusted zones. Cihan Topal *et al.* [24] discusses about the efficiency analysis of virtual keyboard and proposed to fix the location of letters in order with a compact keyboard layout. High occurrence letters are located close together in the new layout for efficient operation.Soumalya Ghosh *et al.* [25] proposed an effective virtual keyboard design with size and space adaption. The keyboard is designed on the basis of key size and centre distance between keys to increase the performance of text entry. Radha Damodaram *et al.* [26] proposed a random virtual keyboard for banking applications. The design and display of the keyboard may be changed randomly depending upon date, time and client. The credentials are encrypted in client side itself to avoid any attacks and decrypted in the server side. Kumar chellapilla *et al.* [27] proposed segmentation based HIP that is easy for humans but difficult for computers. Also introduces Global Warp and Local warp image characters with thin and thick arcs to prevent OCR attacks on Captcha.

Proposed Method: Virtual keyboards [28] are technological breakthrough. They provide marvelous flexibility in design and use. It can be programmed to any language and special letters too. In the physical keyboard, the positions of keys are fixed but in virtual keyboard, the position of keys and their values are changed according to the requirement through software programs. Virtual keyboards are alternative to physical keyboard, used to enter the important credentials like user name, password in banking and other applications. In the proposed approach, virtual keyboard with encrypted keys are provided to the user for Captcha entry. The encryption technique for virtual keyboard is selected randomly by server each and every time for every user to ensure high security. A virtual keyboard using simple Caesar cipher algorithm [29] is experimented for the proposed approach. If 'n' value is 4 then every letter is shifted by 4 positions as indicated in Table 1.

If the virtual keyboard consist only 26 alphabets alone then the encryption function is.

$$e(c) = (c + k) \pmod{26}$$

where c represents the alphabet character, k is the key value that is number of character shifts. If the virtual keyboard is designed with 67 keys, A-Z, a-z, 0-9 and five function keys. Then the encryption function will be;

$$e(c) = (c + k) \pmod{67}$$

The server arbitrarily select any one of the available encryption technique such as Auto key, Play fair cipher, Mono alphabetic, Hill cipher [30] or different 'n' values for every user or virtual keyboard with random key codes, as shown in Table 2.

For 26 key random virtual keyboards the encryption function will be;

$$e (c) = Ran (26)$$

If the random virtual keyboard is designed 67 keys with A-Z, a-z and 0-9 with five function keys. Then the encryption function will be;

$$e (c) = Ran (67)$$

Experimentation: In the Experimentation stage, a website homepage login screen is displayed to the user. The server also generates and displays a random six character Captcha image in the user region. The screen shot of Login page is shown in Fig. 2. A look up table with IP address of the client, session ID, displayed Captcha and encrypted invisible Captcha are stored in server database for comparison and authentication as shown in Table 3.

Table 1: Simple Caesar (n=4)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

Table 2: Random virtual key assignment

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Q	W	E	R	T	Y	U	I	O	P	A	S	D	F	G	H	J	K	L	Z	X	C	V	B	N	M

Table 3: Lookup Table

IP Address	Session ID	Original Captcha displayed in user login screen	Encrypted Invisible Captcha in Lookup table for comparison and authentication
----	----	A5nKm8	E9rOq2

Fig. 2: Screen shot of Login page

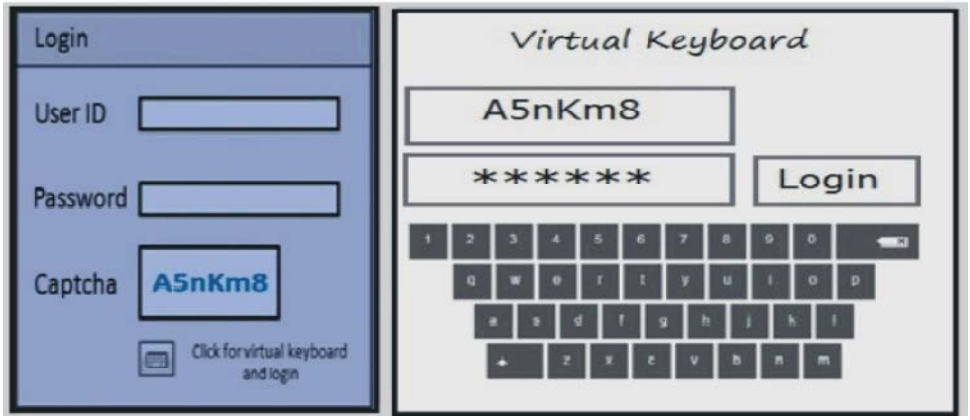

Fig. 3: Screen shot of Login page with virtual keyboard

Fig. 4: Concept of Encrypted virtual keyboard

A check box with virtual keyboard icon is provided in the signup page for selection and activation of virtual keyboard. By cognition ability, the user clicks the check box and a popup window appeared with virtual keyboard, input textbox and submit button. The user inputs the Captcha characters through virtual keyboard by mouse action and click 'LOGIN' button for authentication; the screen shot image is shown in Fig. 3.

In the signup page the Captcha 'A5nKm8' will be displayed to the user in the login screen. By cognition skill, the human users click on the virtual keyboard icon to activate the virtual keyboard and entered the Captcha characters by mouse action. The user is not aware of the encryption used in the keyboard or Lookup table in database. The encrypted Captcha is denoted by special characters only in the user region and sent to the server for comparison and authentication. The concept of encryption is shown in Figure 4.

Human cognitive abilities [31] are brain based skills. The humans need to carry out any task from simplest to more complex with the basic knowledge of learning, remembering and problem-solving activities. Perception, attention, motor skills, visual and spatial processing are the important cognitive ability that human possesses. The Invisible Captcha approach is devised with motor skill cognitive activities of human. The bots are unsuccessful in performing the cognition activities and hence unable to gain access of web applications.

CONCLUSION

The Invisible Captcha has been implemented by an exclusively designed encrypted virtual keyboard. Captcha entry through virtual keyboard itself is a secured method; but in addition the proposed approach provides more security by encrypted virtual keyboard. The Invisible Captcha substantiates a secured approach with the following:

- Cognition based click action on check box, to activate the virtual keyboard, proves the presence of human in the web site.
- The user need to activate the text box in the popup window for Captcha entry
- Entry of Captcha characters through encrypted virtual keyboard by mouse action ensures Human presence.
- Since encryption is involved, there is no need of twisted Captcha characters or background noise, which is purely optional.
- All the encryption process, comparison of Captcha characters is performed in server and it is not visible to anyone, therefore it will not create panic to the users.
- The bots may acquire the Captcha image using OCR software and by segmentation, preprocessing techniques it identifies the Captcha characters as 'A5nKm8' and reproduces the same in the textbox. But no text box is provided in the Login screen. Bots unable to perform click action on check box as well the mouse action on virtual keyboard.
- In all the cases the Captcha characters scanned by the OCR will not match with the Invisible Captcha 'E9rOq2' stored in lookup table. Therefore authentication gets failed.
- Thus, Invisible Captcha approach provides 100% enhanced security against OCR attacks.

REFERENCES

1. Kumar Chellapilla, Kevin Larson and Patrice Y. Simard, 2005. Building Segmentation Based Human-Friendly Human Interaction Proofs, Lecture Notes in Computer Science, Springer, pp: 1-26.
2. Kaur Kiranjot and Sunny Behal, 2014. Captcha and Its Techniques: A Review, International Journal of Computer Science and Information Technologies, 5(5).
3. Buvanesvari, R. and V. Prasath, 2015. A New Security Mechanism for Graphical Password Authentication using Combo Captcha in Video Technology, International Journal of Science and Research, 4(1).
4. Bilal Khan, Khaled Alghathbar and Muhammad Khurram Khan, 2013. Cyber Security Using Arabic CAPTCHA Scheme, The International Arab Journal of Information Technology, 10(1).
5. Coates, A.L., S. Bired and J. Fateman, 2003. Pessimal Print: A Reverse Turing Test, International Journal on Document Analysis and Recognition, 5(2): 158-163.
6. Chew, Monica and Henry S. Baird, 2003. Baffle Text: a Human Interactive Proof, In the Proceedings of the SPIE/IS&T Document Recognition & Retrieval Conference, USA, pp: 22-23.
7. Choudhary Sarika, Ritika Saroha and Yatan Dahiya, 2013. Understanding Captcha: Text and Audio Based Captcha with its Applications, International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).
8. Kulkarni, Sushama and Fadewar, 2013. Captcha Based Web Security: An Overview, International Journal of Advanced Research in Computer Science and Software Engineering, 3(11).
9. Xiao Ling-Zi and Zhang Yi-Chun, 2012. A Case Study of Text-Based CAPTCHA Attacks, In the proceedings of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, IEEE.

10. Tanvee Moin Mahmud, Mir Tafseer Nayeem and Md. Mahmudul Hasan Rafee, 2011. Move & Select: 2-Layer CAPTCHA Based on Cognitive Psychology for Securing Web Services, International Journal of Video & Image Processing and Network Security, 11(5).
11. Chandarana Jagruti and Mayank Kapadia, 2014. Optical Character Recognition, International Journal of Emerging Technology and Advanced Engineering, 4(5).
12. Mithe Ravina, Supriya Indalkar and Nilam Divekar, 2013. Optical Character Recognition, International Journal of Recent Technology and Engineering, 2(1).
13. Serrao Merrill, Shanu Salunke and Amrita Mathur, 2013, Cracking Captchas for Cash: A Review of CAPTCHA Crackers, International Journal of Engineering Research & Technology, 2(1).
14. Barbar, Aziz and Anis Ismail, 2015. Character Image Semantic-Based CAPTCHA, International Journal of Future Computer and Communication, 4(3).
15. Gao Haichang, Jeff yan and Zhengya Zhang, 2016. A simple generic attack on text Captchas, In the Proceedings of Network and Distributed System Security Symposium, San Diego, USA.
16. Jeff Yan and Salah El Ahmad, 2008. A Low cost Attack on a Microsoft, In the Proceedings of the ACM conference on computer and communications security, New York.
17. Anjali Avinash Chandavale and A. Sapkal, 2012. Security Analysis of CAPTCHA, In the Proceedings of International Conference, SNDS, India, Springer.
18. Oleg Starotenko, Claudiacruz perez and Fernando Uceda Ponga, 2015. Breaking text based Captchas with variable word and character orientation, Journal of Pattern Recognition, 48(4).
19. Rajarajan, Maheswari and Hemapriya, 2014.Shoulder Surfing resistant Virtual Keyboard for Internet Banking, World Applied sciences Journal, 31(7).
20. Parekh Ankit, Ajinkya pawar and Pratik munot, 2011. Secure Authentication using Antiscreenshot Keyboard, International Journal of Computer Science, 8(5).
21. Biranchi Andrea and Ian Oakley, 2010. The Secure haptic keyboard: a tactle password system, In the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp: 1089-1092.
22. Agarwal, M. Mehra and R. Pawar, 2011. Secure authentication using dynamic virtual keyboard layout, In the Proceedings of the International conference and workshop on emerging trends in Technology, ACM.
23. Adhikary Nairit, Rohit Shrivastava and Aswani Kumar, 2012. Battering key loggers and screen recording software by fabricating passwords, International journal of computer network and information security.
24. Cihan Topal, Burak Benligiray and Cuneyt Akinlar, 2012. On the efficiency issues of virtual keyboard design, In the IEEE International conference on Virtual Environments Computer Interfaces and Measurement Systems.
25. Ghosh Somalya, Sayan Sarear and Manoj kumar Sharma, 2010. Effective virtual keyboard design with size and space adaption, In the Proceedings of the 2010 IEEE students Technology symposium, kharagpur, India.
26. Damodaram Radha and M.L Valarmathi, 2010. Security Measures of Randvul keyboard, International Journal of Computer Science and Engg, 02(3): 619-625.
27. Kumar Chellapilla, Kevin Larson, Patrice Simard *et al.*, 2005, Designing Human friendly HIPs, in; Proceedings of the SIGCHI conference on Human factors in Computing Systems, pp:711-720, ACM, new York, USA.
28. Nikhil Koul, Pranav Nawathe and Pranav Tulpule, 2014. Virtual Keyboard, International Journal of Scientific & Engineering Research, 5(4).
29. Purnamaa Benni and Hetty Rohayani, 2015. A New Modified Caesar Cipher Cryptography Method With Legible Cipher text From A Message To Be Encrypted, Procedia Computer Science 59, Elsevier.
30. Rajput Yashpalsingh, Dnyaneshwar Naik and Charudatt Mane, 2014. An Improved Cryptographic Technique to Encrypt Text using double encryption, International Journal of Computer Applications, 86(6).
31. Belk Marios, Christos Fidas, Panagiotics *et al*., 2015, Do Human cognitive differences in Information Processing affect and performance of Captcha, International Journal of Human computer studies, , Elsevier, pp: 84.