FOUR FACTOR AUTHENTICATION SYSTEM TEXT-BASED GRAPHICAL PASSWORD SCHEME USINGMACHINE LEARNING

GNANAKUMAR

NS. Sudhakaran
IT-Finalyear
Jei Mathaajee Engineering College
Kanchipuram

K. Mageshwar
IT-Finalyear
Jei Mathaajee Engineering College
Kanchipuram

ABSTRACT: Captcha is a security mechanism designed to differ-entiate between computers and humans, and is used to defend against malicious bot programs. Text-based Captchas are the most widely deployed differentiation mechanism, and almost all text-based Captchas are single-layered. Numerous successful at-tacks on the single-layer text-based Captchas deployed by Google, Yahoo! and Amazon have been reported. In new two-layer Captcha scheme was deployed in 2015 by microsoft. This appears to be the first application of two-layer Captchas. It is therefore natural to ask a fundamental question: is the two-layer Captcha as secure as its designers expected? Intrigued by this question, we have for the first time systematically analyzed the security of the two-layer Captcha in this paper.

INTRODUCTION

APTCHA(Completely Automated Public Turing Test to CTell Computers and Humans Apart) is used to prevent automated registration, spam or malicious bot programs [1],

It automatically generates and evaluates a test, difficult for computers to solve, but easy for humans. If the success rate of solving a Captcha for humans reaches 90% or higher, and computer programs only achieve a success rate of less than 1%, the Captcha can be considered secure [3].

Current Captchas can be divided into three categories: text-based, image-based and audio-based. Text-based Captcha is usually based on English letters and Arabic numerals, and uses sophisticated distortion, rotation or noise interference to prevent the recognition of a machine. Compared to the latter two Captcha

categories, text-based Captcha is the most widely used scheme

This wide-spread usage is due to its obvious advantages [5], [6]: users' task is a text recognition problem which is intuitive to users worldwide; globally, people can recognize English letters and Arabic numerals, thus text-based Captcha has few localization issues; at last, text-based Captcha

Manuscript received May 11, 2016; revised October 11, 2016; December 13, 2016 and February 7, 2017; accepted March 1, 2017. Date of publication March 1, 2017; date of current version March 1, 2017. the National Natural Science Foundation of China (61472311) supported this work and the Fundamental Research Funds for the Central Universities.

The authors are with the Institute of Software Engineering, Xidian Univer-sity, Xi'an, Shaanxi, 710071, P.R.China. (e-mail: hchgao@xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.

was the earliest form of Captchas and people are more willing to undertake this challenge compared to other Captcha forms.

Security is the most significant concern for text-based Captchas. There have been numerous examples of Captcha failure, and many attacks have been proposed, noted in [7]=

Researchers have recently claimed that their simple generic attacks have broken a wide range of text-based Captchas in a single step [10], [11]. Although many text-based Captchas have been proven insecure, the most recent studies [12], [13] suggest that Captcha is still a useful security mechanism, and the security of text-based Captchas is currently a hot topic in the academic field.

With each failed Captcha scheme, Captcha designers accu-mulated experience. Then they designed better schemes, which aimed to improve both usability and security. These prior attempts also promoted a scientific understanding of Captcha's robustness. As [6] suggested, the robustness of text-based Captchas should rely on the difficulty of finding where each character is (segmentation), rather than what each character is (recognition). Current machine learning algorithms, such as Neural Networks or K-Nearest Neighbors (KNN), can easily recognize distorted or rotated single characters correctly

Therefore, text-based Captchas should be segmentation-resistant, and this principle has become the basis for designing text-based Captchas.

Based on these findings, Microsoft deployed a two-layer Captcha in 2015. It connected the sides of each character, across both top-to-bottom and right-to-left, in order to increase the difficulty of detecting where each character is. This was the first application of two-layer Captchas, and it appeared more secure than previous schemes. All attacks proposed were unsuccessful, including these powerful generic methods pro-posed in [10], [11]. It is therefore natural to ask a fundamental question: is the two-layer Captcha as secure as its designers expected? This open question precipitated our study.

In this paper, we examine the security of the two-layer Captcha. We attack this scheme by a series of processing steps. First, a novel

two-dimensional segmentation approach is proposed to separate a Captcha image along both vertical and horizontal directions, which helps create many single characters and is unlike traditional segmentation techniques. Second, we improved LeNet-5 [14], a radical Convolutional Neural Network (CNN) model, as our recognition engine. Our CNN model, as compared with the traditional template-based KNN, shows improved performance and effectiveness. Third, a Captcha generating imitator was designed which is able to automatically produce Captchas similar to the real-world ones deployed by Microsoft. We adopt it to generate a significant

amount of data for our training sets to train the CNN. Then, the trained CNN is used to recognize the Captchas deployed in the wild. Our attack has achieved a success rate of 44.6% with an average speed of 9.05 seconds on a standard desktop computer (with a 3.3 GHz Intel Core i3 CPU and 2GB RAM). Judging by both criteria commonly used in the Captcha community [5], [9], we have successfully broken the two-layer Captcha deployed by Microsoft.

This appears to be the first systematic analysis of two-layer Captchas. Our two-dimensional segmentation technique is innovative; it can be applied as a basis for other successful attacks. Additionally, our data preparation method, imitating the Captcha generation process to build training sets, and using it to train the recognition engine to recognize the objects in the real world, is the first attempt in the field of analyzing Captchas' robustness. Our experiments demonstrate that it decreases time spent on data preparation and reduces manual labor. Also, this approach can be adapted for other projects requiring a large amount of data to train machine learning systems, and for people working in Captcha design, our work provides guidelines for designing Captchas with better security and usability.

The remainder of the paper is organized into the following sections. A review of related work is provided in Section 2. Then, Section 3 introduces and analyzes Microsoft's two-layer Captcha. Following that, the whole attack process is presented in Section 4, and Section 5 details the experiment process and presents attack results. In Section 6, we talk about other design choices, the applicability and novelty of this work, and also provide guidelines for designing Captchas with better security and usability. Finally, Section 7 concludes the paper with a discussion of the implications of our work

EXISTING SYSTEM:

security primitives hard was based on mathematical problems . emerging and exciting security was used in hard AI problems for new paradigm, but has been underexplored

creating cryptographic primitives is Fundamental task in security which is based on hard mathematical problems that are computationally intractable.

- limited success has been achieved by this pradigm as compared with the cryptographic primitives based on hard math problems and their wide applications.
- Hard AI (Artificial Intelligence) problems used for security, initially proposed is an exciting new paradigm. Inviting puzzle is the most notable primitive which distinguishes human users from computers by presenting a challenge.

OUR WORK

New security primitive was presented by us which was based on hard AI problems, namely, a novel family of graphical password systems built on top of Puzzle technology,
which we call Puzzle as graphical passwords (CaPRP). CaPRP is both a Puzzle and a graphical password scheme.

security problems number was addressed by CaPRP combined altogether, such as online guessing attacks, relay attacks, and, if combined withdual-view technologies, shoulder-surfing attacks. Notably, a CaPRP password can be found only probabilistically by automatic online guessing attacks even if the password is in the search set. A novel approach was offered by CaPRP to address the well-known problem in image popular graphical password systems in hotspot image, such as PassPoints, that often leads to weak password choices. Panacea is not a CaPRP, but it also gifted a reasonable security and usability which was appeared to be suit well with few practical applications for developing online security.We present exemplary CaPRPs built on both text Puzzle and image-recognition Puzzle.

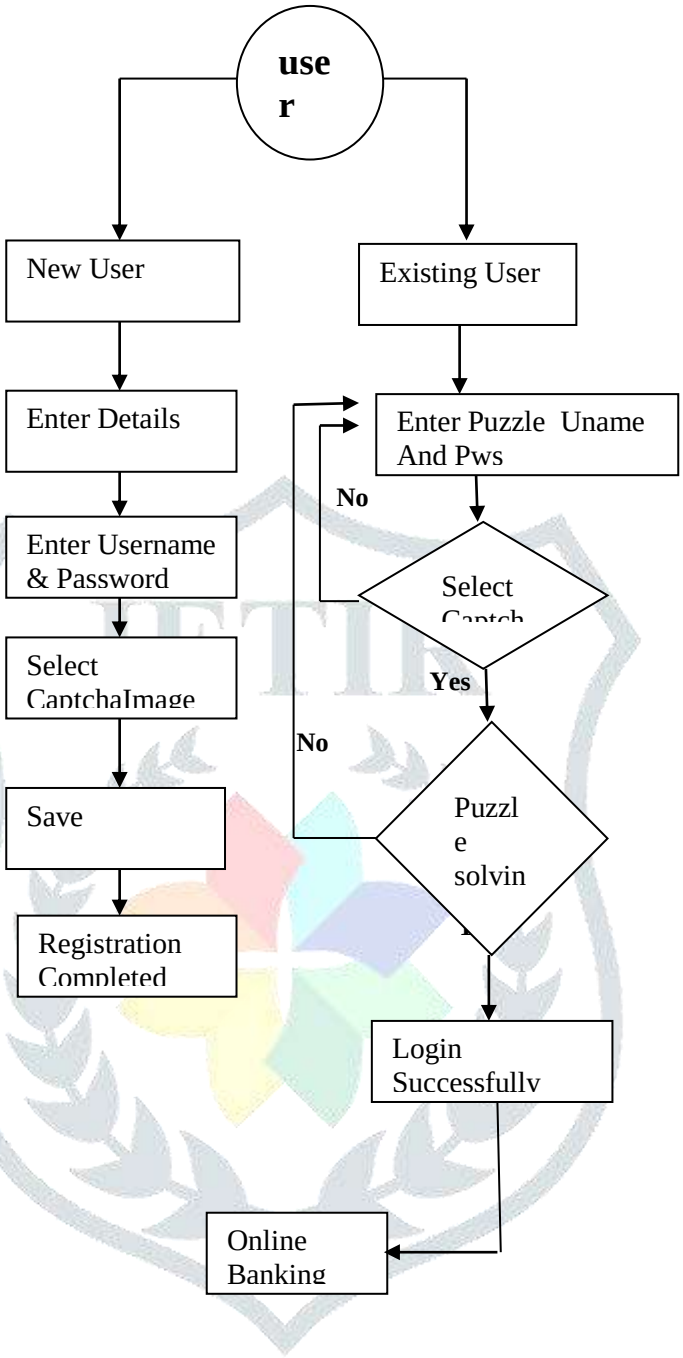
One of them is a text CaPRP where in a password is a characters sequence like a text password, but entered by clicking the right sequence of character on CaPRP images. CaPRP offers protection against online dictionary attacks on passwords, which have been for long time a major security threat for various online services. This threat was floodedmore and considered as a top cyber security risk. Defense against online dictionary attacks is a more subtle problem than it might appear.

It offers reasonable security and usability and appears to fit well with some practical applications for improving online security.
This threat is widespread and considered as a top cyber security risk.

Defense against online dictionary attacks is a more subtle problem thanit might appear.
Puzzle Login(top of Puzzle technology Using mathematical problems).

Image Puzzle Solving Using AES Algorithm.

SYSTEM ARCHITECTURE:



Net Beans 7.0.1

Programming languages such as Java, PHP, C/C++,and HTML5 was developed primarily in integrated development environment using Net Beans. It is also an application platform framework for Java desktop applications and others. The Net Beans IDE is written in Java and can run on Windows, OS X, Linux, Solaris and other platforms supporting a compatible JVM.

A set of modular software components was developed from an application that was allowed by a Net Beans called modules. Applications based on the Net Beans Platform (including the Net Beans IDE itself) can be extended by third party developers The Net Beans Team actively supports the product and seeks feature suggestions from the wider community.

Community testing and feedback was preceded by a time for every release.

Net Beans IDE provides first-class comprehensive support for the newest Java technologies and latest Java specification enhancements before other IDEs.

Popular languages such as JDK 8 previews, JDK 7, Java EE 7 including its related HTML5 enhancement and Java FX was supported buy the first free tool of IDE Net Beans .With its constantly improving Java Editor, many rich features and an extensive range of tools, templates and samples, Net Beans IDE sets the standard for developing with cutting edge technologies out of the box.text editor was much lesser than IDE. Lines, matches words and brackets, and highlights source code syntactically and semantically which was indents by Net Beans Editor. It also provides code templates, coding tips, and refactoring tools.

The editor supports many languages from Java, C/C++, XML and HTML, to PHP, Groovy, Java doc, JavaScript and JSP. Many other languages was supported by plug in of extensible feature of editor. Keeping a clear overview of large applications, with thousands of folders and files, and millions of lines of code, is a daunting task. Different views of your data,was provided by a IDE Net Beans from multiple project windows to helpful tools for setting up your applications and managing them efficiently, letting you drill down into your data quickly and easily, while giving you versioning tools via Subversion, Mercurial, and Git integration out of the box. Because of well-organized code the new developer understand the structure of your application whene they are joins . Net Beans provides static analysis tools, especially integration with the widely used Find Bugs tool, for identifying and fixing common problems in Java code. In addition,Breakpoints were placed in our source code with the help of Net Beans Debugger , add field watches, step through your code, run into methods, take snapshots and monitor execution as it occurs.

The Net Beans Profiler provides expert assistance for optimizing your application's speed and memory usage, and makes it easier to build reliable and scalable Java SE, Java FX and Java EE applications. Without looking into source code you can debug your interfaces using visual debugger for Java SE applications, which was contained by a Net Beans . Take GUI snapshots of your applications and click on user interface elements to jump back into the related source code. Net Beans IDE 7.0.1, which has full support for the official release of the Java platform.

MODULE EXPLANATIONS:

1.PUZZLE LOGIN

The security and usability problems in text-based Login Andpassword schemes have resulted in the development of Puzzle password schemes as a possible alternative.

We can visualize the sum 1+2+3+...+n as a triangle of character . Numbers which have such a pattern of character are called Triangle (or triangular) numbers, written T(n), the sum of the integers from 1 to n time Using Factorial base Login Puzzle Solving.

2.RANDOMCAPTCHA SELECTION

Humans and machines was separated by using test called CAPTCHA. "Completely Automated Turing test to tell Computers and Humans Apart" Simply called in terms of CAPTCHA. It is an image test or a simple mathematics problem which a human can read or solve, but a computer cannot do it. Hundreds of accounts was setup automatically by using a program to stop computer hackers with help of CAPTCHA, such as email accounts. It is named after mathematician.

Each and every individual chosen randomly and entirely by chance, Every individual has the same probability of being chosen during the sampling process at any stage and for the sample as any other subset of n individuals we could chosen the same probability subset of each individuals. This process and technique is called as simple random sampling, and should not be mismatched with systematic random sampling. A simple random sample is an unbiased surveying technique.

3.IMAGE PUZZLE SOLVING

we study how to prevent DoS/DDoSattackers from inflating their puzzle-solving capabilities. To this end, we introduce a new client puzzle referred to as software puzzle.we publish their puzzle algorithms in advance which is not same as the existing client puzzle schemes, a puzzle algorithm in the present software puzzle scheme is randomly generated only after a client request is received at the server side and the algorithmis generated such that:

For solving the problem the attacker was unable to implemented the puzzle in advance and the attacker needs considerable effort in translating a central processing unit puzzle software to its functionally equivalent GPU version such that the translation cannot be done in real time. Moreover, we show how to implement software puzzle in the generic server-browser model.

4.OTP GENERATION

For only one login session or transaction we use valid one-time password (OTP) is on a computer system or other digital device. OTPs avoid a number of shortcomings that are associated with traditional (static) password-based authentication; a number of implementations also incorporate two factor authentication by ensuring that the one-time password (OTP) needs access to something a person has (such as a small key ring fob device with the OTP calculator built into it, or a smartcard or specific cell phone) as well as something a person knows (such as a PIN).

5.ONLINE BANK

Online banking otherwise called as net banking, e-banking, or virtual banking. Is a electronic payment or e-payment is a system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website. The core banking system was typically connected to or be part of operated by a bank and is in contrast to branch banking that was the traditional way customers access banking services.

CONCLUSION

The software puzzle may be constructed upon on data puzzle it and it could be added with any existing server-side data puzzle scheme, and easily deployed as the present client puzzle schemes do .CAPTHCHA is widely research field act as internet rectifier to secure web applications by discern human from bots. CAPTCHA presented which will improve resistance of math calculus CAPTCHA. By use, Boolean operations and expressions instead of trigonometric and differential function which will help in reduce the complexity of CAPTCHA and help to achieve better usability and security as compared to math calculus CAPTCHA. Boolean CAPTCHAcan be easily use by educateduser. No need of technical skill, by using intellectual mind to solve this CAPTCHA and help to reduce time complexity.

REFERENCES

[1] L. Von Ahn, M. Blum, N. J. Hopper, and J. Langford, "Captcha: Using hard ai problems for security," in *Advances in CryptologyłEUROCRYPT 2003*. Springer, 2003, pp. 294–311.

[2] L. Von Ahn, M. Blum, and J. Langford, "Telling humans and computers apart automatically," *Communications of the ACM*, vol. 47, no. 2, pp. 56–60, 2004.

[3] E. Bursztein, M. Martin, and J. Mitchell, "Text-based captcha strengths and weaknesses," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011, pp. 125–138.

[4] K. Chellapilla, K. Larson, P. Y. Simard, and M. Czerwinski, "Building segmentation based human-friendly human interaction proofs (hips)," in *Human Interactive Proofs*. Springer, 2005, pp. 1–26.

[5] C. Kumar, L. Kevin, S. Patrice, Y., and C. Mary, "Computers beat hu-mans at single character recognition in reading based human interaction proofs (hips)," in *CEAS 2005 - Second Conference on Email and Anti-Spam, July 21-22, 2005, Stanford University, California, USA*, 2005.

[6] J. Yan and A. S. El Ahmad, "Breaking visual captchas with naive pattern recognition algorithms," in *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*. IEEE, 2007, pp. 279–291.

[7] E. Bursztein, M. Martin, and J. Mitchell, "The strengths and weaknesses of text-based captcha," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011, pp. 125–138.

[8] H. Gao, J. Yan, F. Cao, Z. Zhang, L. Lei, M. Tang, P. Zhang, X. Zhou,

[9] Wang, and J. Li, "A simple generic attack on text captchas," in *Proc. Network and Distributed System Security Symposium (NDSS). San Diego, USA*, 2016.

[10] E. Bursztein, A. Moscicki, C. Fabry, S. Bethard, J. C. Mitchell, and Jurafsky, "Easy does it: captchas usable is more ," in *Sigchi Conference on Human Factors in Computing Systems*, 2014, pp. 2637–2646.

[11] Y. Lecun, "Lenet-5, convolutional neural networks."

[12] M. Naor, "Verification of a human in the loop or identification via the turing test," 1996.

[13] D. Lillibridge, Mark, M. Abadi, K. Bharat, and Z. Broder, Andrei,"Method for selectively restricting access to computer systems," Feb. 27 2001, uS Patent 6,195,698.

[14] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: Breaking a visual captcha," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–134.

[15] P. Simard, "machine learning used to break visual human interaction proofs (hips," *Advances in neural information processing systems*, vol. 17, pp. 265–272, 2005.

[16] S. Hocevar, "Pwntcha-captcha decoder web site," 2007.

[17] H. Gao, W. Wang, J. Qi, X. Wang, X. Liu, and J. Yan, "The robustness of hollow captchas," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 1075–1086.

[18] C. Karthik and A. Recasens, Rajendran, "Breaking microsofts captcha," 2015.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] B. M. ONeil, "Neural network for recognition of handwritten digits,"2013.

[21] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convo-lutional neural networks applied to visual document analysis," in *null*. IEEE, 2003, p. 958.

[22] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*. Cite-seer, 1990.

[23] Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolu-tional neural networks," *arXiv preprint arXiv:1312.6082*, 2013.

[24] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[25] H. Gao, W. Wang, Y. Fan, J. Qi, and X. Liu, "The robustness of" connecting characters together" captchas." *J. Inf. Sci. Eng.*, vol. 30, no. 2, pp. 347–369, 2014.

[26] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *JOSA A*, vol. 4, no. 12, pp. 2379– 2394, 1987.

[27] V. Turchenko and A. Luczak, "Caffe: Convolutional architecture for fast feature embedding," *Eprint Arxiv*, pp. 675–678, 2014.

[28] M. Mohamed, N. Sachdeva, M. Georgescu, S. Gao, N. Saxena, C. Zhang, P. Kumaraguru, P. C. Van Oorschot, and W. B. Chen, "A three-way investigation of a game-captcha: automated attacks, relay attacks and usability," in *ACM Symposium on Information, Computer and Commu-nications Security*, 2014, pp. 195–206.