Innovative CAPTCHA-Based Authentication Technique for Cloud Data Protection

Shouket Ahmad Kouchay
Research Scholar Computer Science

Abstract—Cloud computing is an innovative standard to provide services over the web. Cloud computing has integrated into our life and work every day and way. But data security in cloud computing is a serious concern. Authentication is the main technique for securing the information. CAPTCHA is a verification test to guarantee that the response is only produced by humans and not by a computer bot. We propose a Double Check CAPTCHA for Cloud computing, where a user has to identify a specific picture and type its name in the text box and click on any specific portion of a picture instead of clicking on the submit button. It verifies the user twice, first time for identifying the concerned picture and second time to click on the right place for submitting the relevant word. Even if the bot gives the right response they can't click on the suggested portion of a picture for submission because human intervention is needed for the right answer. The proposed scheme is beneficial in protecting Cloud computing resources as they block the automatic bot programs in the Cloud server.

I. INTRODUCTION

Cloud computing provides a lot of benefits particularly in pervasive services where users can access computing services through the Internet. But robust authentication for data security is vital for cloud service adoption. [1]. There are a lot of security and data protection vulnerabilities in authentication techniques such as eavesdropping, replay, exhaustive and dictionary attacks, etc. The cloud services have been adopted by business companies and institutions to empower their customers, from e-banking, e-businesses, eLearning portals, electronic mails, and e-shopping, to other community-based applications such as for weblogs and Web forums and many more. But despite using this cloud technology, the information exchange in cloud services has become vulnerable in terms of security issues. Many of these web services are being consistently attacked by automated programs called bots. The threats from program bots are posing various types of attacks. CAPTCHA is the main element for user authentication for the cloud computing environment. CAPTCHAs are very beneficial to avoid Denial of Service and brute force attacks. The test is also useful to avoid unwanted comments on the blog and ultimately restricting advertising spam by intended computer programs. CAPTCHA is a standard Internet security method being used to prevent such attacks.

But as one CAPTCHA is developed, the spammers are there trying to attack it. Even big companies are not safe from such security vulnerabilities. According to news reports the three big web email service providers, Microsoft's Hotmail, Yahoo Mail, and Google's Gmail [2], also have been targeted due to security issues. Moreover, some CAPTCHAs are not user-friendly. Many users find existing character-recognition based CAPTCHAs difficult and frustrating. So there is a need to develop a robust, secure and user-friendly CAPTCHA to make the Cloud services and online database more secure and spam-free.

In this paper we have discussed various captcha schemes, security issues in cloud computing and have developed a secure user authentication technique Captcha for fighting the bots in cloud computing. The paper presents a CAPTCHA algorithm, which may require extra struggle from the CAPTCHA breaking algorithm by joining various distortion methods all at once. Such different resolutions could show a vital part in different cloud security techniques. Some related works are briefly described in section 2, the proposed security technique as well as the process is presented in section 3 and finally, in section 4 the conclusion of this work is reported.

II. RELATED WORK

Cloud Computing is a new resolution in IT engineering that decreases computing costs through virtualization. Hence, focusing on cloud storage security improvement is very important. Data cloud storage comes with risks and security threats. The organization's data is susceptible to security risks when it is connected to the Internet. Currently, virtual world criminals can access cloud data storage. Individual cloud data storage, vital information, documents, and files are usually linked to third party software, which opens doors to vulnerability data breach. What makes cloud storage vulnerable is that it stores individual or organization's data, which can be accessed by different sources linked to the cloud network. It is obligatory to supply a secure connection over several network resources. [3].

Numerous efforts have been made by the research community to propose and explore new security techniques in cloud security [11].
The authors [4] have reviewed important advances in biometric field. Graphical passwords are substitute to text-based passwords that can be solved easily and susceptible to attacks like shoulder-surfing attack.

Flooding attacks cause both direct and indirect denial of service (DoS). Usually, when a cloud finds a lot of requests for a particular server, it accounts for additional computing power to that service to handle all the requests. However, in the real situation, this would provide an advantage to the “hacker”. After that, the hacker only needs to focus on his flooding attack on a single cloud server so that he can gain access to cloud account services. This is service is known as direct ' Denial of Service' because the hacker focuses on a particular service to get it down. [5].

would provide an advantage to the “hacker”. After that, the hacker only needs to focus on his flooding attack on a single server so that he can gain access to cloud account services. This is service is known as direct ' Denial of Service' because the hacker focuses on a particular service to get it down. [3]. [1]. Kevin Curran, Sean Carlin, and Mervyn Adams, “Security issues in cloud computing”, Elixir Network Engg.38 (2011), pp.4069-4072, August 2011. [3]. Wu, B., Chen, J., Wu, J. and Cardei, M. (2007) A survey of attacks and countermeasures in mobile ad hoc networks. Signals and Communication Technology, Part II, 103-135.

A way to tell apart a human from a computer by a text is known as Turing Test. When a computer program can generate such tests and evaluate the result, it is known as CAPTCHA “Completely Automated Public Test to Tell Computers and Humans Apart[1].”

The concept of a CAPTCHA is motivated by real-world problems faced by internet companies such as Yahoo, Hotmail, Google. These companies offer free email accounts intended for human use only. However, they found that many online vendors were using computer programs known as bots to sign up for thousands of email accounts, from which they could send out массеs of spams, junk mails, etc. By using CAPTCHA in their services the user is required to solve it to create the email. Through the CAPTCHA these companies are now able to stop the program bots in entering their system[2].

The automated Turing tests was first proposed by MoniNoar[3]. CAPTCHA term was coined in 2000 at Carnegie Mellon University. The text base CAPTCHAs were used[4].

It is widely accepted that a good CAPTCHA must be both robust and usable. The robustness of a CAPTCHA is its strength in resisting adversarial attacks, and this has attracted considerable attention in the research community [1][5][6].

As reported recently, humans are being used to solve CAPTCHAs, either in a well organized manner commercially or by the use of games and other methods whereby humans are unaware that their responses are being used for malicious purposes.

These attempts make it futile to make harder AI problems, because a CAPTCHA should be solvable virtually by all humans, regardless by their intention. So, CAPTCHAs are and will remain deployed until an alternate scheme of human verification becomes practical.

Here are some more CAPTCHA methods. Gimpy method, Pessimal Print Method, Text-to-Speech method [8]. ARTIFACIAL [7] in this method of captcha the chances of human failure is moderate because of artificial distortions and algorithms exist to break artificially distorted pictures are the main weaknesses of artificial.

Multi Modal CAPTCHA [8]: MMC method of CAPTCHA application employs two current schemes Text-Based and Image-Based with one verification step. The word tags are being displayed on image A user has to identify the correct name of the over the image among the set of text labels on it, to pass the test.

Any CAPTCHA has two main demands: first is be comfortable for human to answer and (b) be very hard for a computer script to solve. Those to demands appear to contradict with each other. In the reality where the OCR, image recognition and Machine Learning techniques are well studied it is very hard to design a good CAPTCHA that will still be solvable by humans.

Since CAPTCHA was presented by Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langfordin in 2003 [9] lots of different mechanisms were presented to the public. Most of them were broken after just few years (in the good case) like the GIMPY CAPTCHA [10], Microsoft CAPTCHA [11], ASIRA and even the most common those days – reCaptcha.

The CAPTCHAs can generally be categorized into three groups: Text based CAPTCHAs, Image based CAPTCHAs and Sound based CAPTCHAs [12] The CAPTCHA research is still taking its first steps and look more like art rather then science. This is the main reason that most of them are been quickly broken. Even though this is a way to tell apart a human from a computer by a test is known as Turing Test. When a computer program can generate such tests and evaluate the result, it is known as CAPTCHA “Completely Automated Public test to Tell Computers and Humans Apart”) [6]. The automated Turing tests were first proposed by MoniNoar [7], CAPTCHA term was coined in 2000 at Carnegie Mellon University. The textbase CAPTCHA were used [6]. It is extensively acknowledged that a good CAPTCHA must be both secure and usable.
The strength of a CAPTCHA is its toughness in fighting bots and lot of research has been done in this area [3][7][9].

Gimpy method, Pessimal Print Method, Text-to-Speech method [10]. ARTIFACIAL [17] in this method of captcha the chances of human failure is moderate because of artificial distortions and algorithms exist to break artificially distorted pictures are the main weaknesses of ARTiFACIAL. Multi-Modal CAPTCHA [11]: MMC method of CAPTCHA application employs two current schemes Text-Based and Image-Based with one verification step. The word tags are being displayed on image A user has to identify the correct name of the over the image among the set of text labels on it, to pass the test.

CAPTCHA security technique could be used to prevent the bot attacks in cloud services. User has to type captcha code like OTP.

Any good CAPTCHA has two main demands: first is to be comfortable for the human to answer and be very tough for a bot to response.

The authors [12] in 2003 have presented lots of different captcha mechanisms that were broken in a short period like GIMPY CAPTCHA [12]. Microsoft CAPTCHA [15]. ASIRA and even the most common those days – reCaptcha. The CAPTCHAs can generally be categorized into three groups: Text-based CAPTCHAs. Image-based CAPTCHAs and Sound based CAPTCHAs [16].

Although people are taking part in developing good CAPTCHAs for the sake of security, while others are breaking them for the sake of improving the Machine Learning and OCR study [17]: Any program that solves the CAPTCHA test can be used to solve a hard unsolved AI problem. Most of the attacks found on CAPTCHAs are been presented to the companies using them before handling them to the public. This is done to allow those companies to fix their schemes before any malicious party can take advantage of them.

The Text-Based CAPTCHAs are designed to show distorted text on some background with techniques against segmentation. In the Microsoft MSN CAPTCHA, the designers tried to protect only against segmentation. This was done by adding arcs (some of the background color, some intersecting of letters color and some just around). The CAPTCHA was broken by Jeff Yan and Ahmad Salah El Ahmad [18].

Some of the CAPTCHAs like GIMPY and EZ-GIMPY use words to form a dictionary. This approach allows the CAPTCHA designer to use longer challenges since people are good in the reconstruction of words (even when they don't understand some of the letters), we see another example of the vulnerability when Slashdot CAPTCHA was broken with higher than 24% success rate. The Text-based CAPTCHAs are the most intuitive and common CAPTCHAs and at least they will stay like this in the incoming years. It is hard to design a strong CAPTCHA. Most of them are already broken while the common reCaptcha seems to hold, the question is for how long. Before designing any text-based CAPTCHA it is crucial to read all the known attacks, but even this will give you only protection against known attacks while new ones are discovered every year.

One of the problems with Text-Based CATCHAs and Image-based CATCHAs is accessibility. Some persons have some hearing or vision difficulty, so they can't solve those CAPTCHAs and since a lot of cloud and web services demand the user to solve the puzzle to use the service. The solution is to provide services with Audio CAPTCHAs [19]. Various authors [20] have suggested a solution to single-level authentication and security issues by using single password verification.

III. PROPOSED SECURITY TECHNIQUE

A secured user authentication technique for cloud computing has been proposed.

The Proposed captcha overcomes the drawbacks of the previous methods. Humans can pass the verification test easily. Even elderly people, children or people suffering from Dyslexia can easily identify the image and text labels.

Our Proposed Scheme implementation uses two existing design techniques.

- Text-Based
- Image-Based

These techniques are mutually exclusive. The model of this proposed CAPTCHA is different from other captcha such as MM captcha [11]. The proposed method verifies the user twice, first time for identifying the specific picture and second time for click on the right place for submission, it is one of the innovative technique to present the Double-check captcha as shown in below Figure: -
The proposed CAPTCHA can have implemented by using hundreds of images (birds, cloths, vegetables, etc.) from the current search engines like yahoo, Google, Bing, etc. The data would be freely available to all through these search engines. So, as to keep the 'Public' part of the CAPTCHAs. A database of a large set of images and text labels would be created. The user could easily recognize the image in the proposed scheme as the images taken are animals, clothes, furniture, fruits, etc.

In this proposed technique computer bots need to be able to pass the following four steps:

1. To identify the concerned picture
2. To find the name of the concerned picture on the screen
3. To enter the name of the picture in the text box
4. To click on the right place for submission

It is a tough task for the computer bot to recognize these operations in an accurate sequence, even if bot identifies the right answer they cannot click on the suggested position of a picture for submission or vice versa because for submitting the correct answer human intervention is necessary. It gives a relatively strong challenge for AI experts to design a scheme that could break the proposed captcha.

Some of the features of our method are listed below:

- Exclusive use of both Image and text labels/tags
- The image is scattered with some text labels.
- Second-time user verification by clicking on the right place for submission (Unique).

Thus Double Check captcha blocks efficiently block the bots scripts from flooding the site with scams, fraudulent profiles, registrations.

An innovative method is proposed for differentiating between human and bot programs. In this method, a user has to identify the picture and type its name in the text box and click on any suggested portions of the picture instead of clicking on the submit button. For example, the image of a shirt is being displayed to the user and the user is asked to click on the pocket to submit. It verifies the user twice, first time for identifying the concerned picture and second time to click on the right place for submitting the relevant word. If the user responds with the right text and click on the right place, it would be considered a human user, not a bot program. We proposed both image and text-based techniques being used simultaneously for making secure.

The algorithm of the proposed system is user-friendly and efficient. The process is explained in the following algorithm. The description of some notations are: - ‘A’ is the actual name of the image, ‘t’ is any text label, J is the image name, ‘i’ is the subscript variable, ‘val’ is an integer variable used for human verification. S= submit and cRP= click on the right position.

**Algorithm:**

```plaintext
a) Read ti,
b) val = 1;
c) Foreach val to 3
   d) Foreach ti to 4,
   e) If ti is equal to J && S=cRP
      F ← ti;
      Goto g;
   Else
      i ← i + 1;
      End Foreach
   val ← val + 1
   End Foreach
f) Goto h; (BotProgram)
g) Write F;
h) Stop.
```
Steps:

A. An image is being displayed (clothes etc.).
B. Four text-labels are randomly shown over the image.
C. Only one text word is relevant to the background image.
D. One has to identify the relevant word to enter in the given text box.
E. A user has to click on the right place for submitting the relevant label name.
F. Recognizing the correct text label is the first verification.
G. Submit click on the right place is the second verification.

A user has to recognize the image in max. three attempts to identify the text label and submit click on right place. If he fails to recognize the image, text label, and submit click position, the proposed system denies any more attempts and treats it as a program Bot.

A program flow diagram of Double Check CAPTCHA - a cloud security technique is Figure 2:

Figure 2. Flow diagram of Proposed Security Technique
IV. CONCLUSION

Single level authentication in the cloud computing environment can raise severe problems, primarily with sensitive data. Authentication is one of the main significant tests in the security of cloud computing. The captcha can be used as a pre-authentication to give robust security in cloud services.

We have developed a new technique where image reorganization and recognizing the right portion of an image for submitting the correct answer with a challenge that is very easy to answer by human users but is difficult for automated programs/bots. This research presents a Double Check CAPTCHA where a user has to identify a specific picture and type its name in the text box and click on any specific portion of the picture instead of clicking on the submit button. It verifies the user twice, first time for identifying the concerned picture and second time for click on the right place for submitting the relevant word. Even if bot knows the right answer they can’t click on the suggested portion of a picture for submission because for submitting the correct answer human intervention is necessary. Double-check captcha provides easiness with security. Other enhancements of this approach can be done after addressing a larger group of Cloud users and based on their feedback.

REFERENCES


