A High Level Security Mechanism for Internet Polls

Shahriar Mohammadi
Department of Industrial Engineering
K.N.Toosi University of Technology
Tehran, IRAN
smohammadi40@yahoo.com

Hossein Abbasimehr
Department of Industrial Engineering
K.N.Toosi University of Technology
Tehran, IRAN
ho.abbasimehr@gmail.com

Abstract— **Nowadays Internet Polls are becoming more and more important as they are being used on a large scale.**

Being a web application a poll can suffer from a variety of attack. Recently automated web tools are used to attack Internet Polls in order to alter the results via automatic voting. Several protection methods introduced to stop automatic voting including IP Locking, cookie-based, and Human Interaction Proof. In this article, existing methods for protection are examined and also we have proposed an innovative solution, which considerably increases the security of Internet polls and the reliability of their results. The proposed solution has two phases: the construction phase in which the poll content and CAPTCHA test are created and the utilization phase, in which vote counting operation is done. The CAPTCHA that is used in the proposed method is an image-based CAPTCHA. For the aim of voting, an Internet user has to drag his or her desired poll option expressed in form of movable text object and drop it on to the picture of an object indicated by the CAPTCHA. The proposed method is fair, robust, and resistant. One of the main advantages of the method is that it can be used by all ages.

Keywords- Internet Poll; web bot; Human interactive proof

I. INTRODUCTION

Internet Poll is a web application that can be defined as online survey of public opinion to acquire information about a specific topic. Nowadays Internet Polls are deployed in many websites for collecting information about people's preferences and opinions about several topics.

Being a web application, a poll can suffer from a variety of attack. Depending upon significance of topic; an appropriate security mechanism may be applied to secure Internet Polls from attacks. Attackers use automated web tools (bot) to overturn the result of Polls in minutes.

In this article existing methods for protection are examined, and also we have proposed an innovative solution which considerably increases the security of Internet polls and the reliability of their result. The rest of the paper is organized as follows: in section II, we discuss about the background of work. section III discusses about protection technique against massive voting. The contribution of paper is presented in section IV. Finally, conclusions are presented in section V.

II. BACKGROUND

A. Internet Poll

An Internet poll is a web application that allows Internet users to submit their opinions about specific topic. Normally, a poll system is a dynamic application, which takes advantage of the client/server paradigm. We can identify two main components, the client side, and the server side. The client side consists of a graphical interface, which shows to a voter the title message and possible choices of a poll. An Internet user can select a voting option and express its own preference by clicking on the "vote" button or see the poll results. The server side has to receive the data from the client, in order to complete the operations.

B. How Web Bots Perform Automatic Voting

A web bot can be defined as a computer program that executes a sequence of operations repeatedly, carrying out tasks for other programs or users without the need of human interaction [3]. Actions of bots can be driven by legitimate purposes or can rely on malicious plan Therefore; robots can accomplish two opposite goals: (1) help human beings in carrying out repetitive and time-consuming operations and (2) undertake hostile or illegal activities, becoming a serious threat to web application security [3]. However, web bots are more often referred to source of problems than useful programs. Ollmann [8] states that web bots can be quickly condensed into the following "generations":

- 1st Generation:
  This type of web bots automatically retrieves a set of predefined resources without trying to interpret the content of downloaded files.
- 2nd Generation: Depending upon the nature of the tool, it may just store the content locally (e.g. mirroring), it may inspect the retrieved content for key values(e.g. email addresses, developer comments, form variable, etc.), build up a dictionary of key words that could be used for later brute forcing attack, and so on.
- 3rd Generation: this type of web bots are capable of correctly interpreting client-side code such as JavaScript, VBScript, Java, or some other "just in time interpreted language".

Currently 1st and 2nd generations are rather common, but the 3rd are not widespread. In the context of Internet Polls, web bots analyzes the HTML page that contains the poll, they discover the poll options; and finally, they

978-1-4244-6893-5/$26.00 © 2010 IEEE

perform voting hundreds and thousands times. This task leads to the changing of poll result.

III. PROTECTON TECHNIQUE AGAINST MASSIVE VOTING

Basso and Bergadano [2] stated that currently there are three different categories of protection technique against massive voting they include:

1. Cookie-based methods
2. IP Locking methods
3. Human Interactive Proof methods

These methods have some advantages as well as disadvantages. From security point of view, we can classify aforementioned methods as weak, medium, or high level methods. We will discuss about these methods separately in the next subsections.

C. Cookie-based Methods

Cookie-based methods are based on cookies. A cookie is a text file that contains the information exchanged between client and server [10].

This information interchange between client and server is necessary for server to be able to verify that whether a client has already voted or not. However, the main problem about the cookies is that the user may disable cookies or deletes them therefore multiple voting becomes possible.

D. IP Locking Method

In this method the Server stores IP Addresses of voters; then in the next voting, server checks that whether an IP Address has already voted, If so that voting attempts are considered invalid. Although this method guarantees high level of security, but it has some disadvantages and this method does not allow to some Internet users to participate in the online polling.

In particular:

- People connected to the Internet through a dynamic IP address assigned by the ISP.
- People connected to the Internet from multi-user workstations.
- People connected to Internet from a LAN.
- People connected to Internet behind a Proxy. As in the NAT situation, different users have the same IP address, so they are all excluded from the voting but the first [3].

E. Human Interactive Proof

An effective solution to protect Internet Polls against web bots is HIP-based methods. Human Interaction Proofs (HIPs) or CAPTCHAs are systems that allow a computer to distinguish between another computer and a human [5]. CAPTCHA stands for “Completely Automated Public Turing Test to Tell Computers and Humans Apart”. CAPTCHAs must satisfy three basic properties. The CAPTCHA tests must be

- Easy for humans to pass;
- Easy for a tester machine to generate and grade;
- Hard for a software robot to pass; the only automaton that should be able to pass a CAPTCHA is the one generating the CAPTCHA [6].

Construction of HIPs of practical value is difficult. In practice, if one wants to block automated scripts, a challenge at which humans are about 90% successful and machines are 1% successful, may not be sufficient, especially when the cost of failure and repetition is low for the machine. Chellapilla and Larson [5] define sweet spot, in which the HIPs are easy for humans to recognize but difficult for hackers to crack. In reality, the existence of sweet spot is not guaranteed. The sweet spot is an area, unsolvable by today’s computers but solvable by humans. As the time moving, the sweet spot area becomes smaller, because computers are getting faster. Unfortunately, humans are unlikely to get better at solving HIPs in the same timeframe. There are different types of CAPTCHA; they are including Text-based, audio-based, and Image-based CAPTCHA.

In the purpose of protecting Online Polls against the web bots, we can use CAPTCHAs. Different types of CAPTCHA guarantee different levels of security. We will examine some type of CAPTCAHs in the following subsection.

1) Text-based CAPTCHAs

This category of CAPTCHAs is made of characters rendered to an image and distorted before being presented to the user. Solving the CAPTCHA requires identifying all characters in the correct order [5].These types of CAPTCHAs are most-used, perhaps the main reasons for their popularity is that they are easily understood by users without much instruction, and can be generated quickly.

Two of the most famous examples of Text-based CAPTCHA are EZ-Gimpy and Gimpy [1]. Yahoo! uses a simple version of this method, known as EZ-Gimpy (Fig. 1) .Recently, a more secure type of text-based HIP; called reCAPTCHA [9]. Other interesting CAPTCHAs are known such as ‘‘Baffle Text’’ and ‘‘Scatter Type’’ which are however relatively difficult to read by user due to their low legibility.

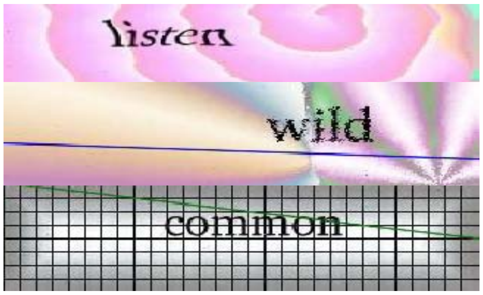


Figure1. Yahoo (EZ-Gimpy)

Breaking Text-based CAPTCHAs is not new, Mori and Malik[7] have successfully broken the EZGimpy (92% success) and Gimpy (33% success).This shows that for building a high-level security Internet Poll System, Text-based CAPTCHAs are not effective methods.

2) Image-based CAPTCHAs

CAPTCHAs belonging to this category require the user to solve a visual pattern recognition problem or understand concepts expressed by images. An example of such test was initially proposed by Blum and von Ahn [1] with the name ‘‘PIX’’ and it is now known as ‘‘ESP-Pix’’. The main problems of this type of CAPTCHA, which may lead the user to fail the test, are misspelling while writing the answer and synonyms of the correct answer [6].

The final version of Image-based CAPTCHA is ‘‘MosaHIP’’ which is devised by Basso and Sicco [3]. The term ‘‘MosaHIP’’ is an acronym for Mosaic-based Human Interactive Proof, which refers to the idea of a HIP based on a mosaic of pictures. MosaHIP have two different versions:

1- “Concept-based” MosaHIP test, which users are required to identify the picture indicated by the CAPTCHA and contained within the mosaic image.

2- ‘topmost’’ MosaHIP test, which asks the user to identify the image representing ‘‘something existing’’ and ‘‘laying upon other pictures’’, not overlapped by anything else.

Security analysis that Basso and Sicco [3] had performed show that both “concept-based” and “top most” MosaHIP CAPTCHAs are:

a) Resistance to segmentation through thresholding;

b) Resistance to segmentation through edge detection;

c) Resistance to shape matching;

d) Resistance to random guessing;

They performed usability test, which show almost every user was able to solve the concept-based version, with 98% of correctly passed tests. Regarding the topmost version, 80% of participants could correctly solve it. These results show that according to the CAPTCHA’s basic properties, the “Concept-based” version of MosaHIP is the best choice for deploying in Internet Poll, because it guarantees high level of security and usability.

IV. PROPOSED METHOD FOR SECURE ONLINE POLL SCHEMA

In this paper, a method for distinction between human user and web bot through dragging and dropping “moveable text object” on to the picture of an object among other object is presented.

In this proposed technique, we use “drag-and-drop” approach, which is used in [4]. The structure of suggested method is simple and as follows:

1) Creation phase

The creation phase consists of three stages. First, poll options in the form of movable text object are prepared. In addition, the poll question is prepared as well. This allows the user to drag and drop the desired choice. The output of this step is something similar to Fig. 2. Second, program Generates “Concept-based” MosaHIP for testing. In this stage, the algorithm, which is used for the creation of mosaic image in version of the “Concept-based” MosaHIP schema [3], is applied. The algorithm receives some images and processes them on the purpose of creating a mosaic image. The output mosaic image consists of fake and real subimages. By means of this process web bots cannot distinct between the real image and fake image but human user can.

The important feature of mosaic image is that subimages are pseudo-randomly positioned within the mosaic image and are also partially overlaying each other, so that separating them into subimages is not an easy task for a computer [3].

The database of pictures used in our proposed method contains a very large number of different images and a considerably high number of categories. This number of pictures can be provided in two ways: by extracting frames from different videos, and downloading pictures from Internet [12].

At the end of second stage the CAPTCHA is generated. Third, the generated CAPTCHA and poll options along with poll question are combined together. The result of this stage is schema similar to Fig. 3.

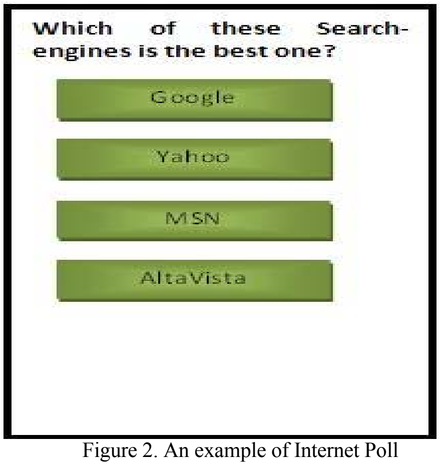
After finishing the three aforementioned stages, the proposed online poll is produced. The creation phase is completed in this point.

2) Utilization phase

This phase describes the operations taken in the time of using the poll by the Internet users. In this phase the user has to drag his or her desired choice expressed in form of movable text and drops it on to picture of an object indicated by the CAPTCHA.

This requires the user to recognize the picture of an object indicated by the CAPTCHA. When the user drops a choice on the indicated image, a vote related to that choice is submitted and correctly counted. Otherwise, a new test is presented to the user

For example, consider the following Internet Poll that its subject is about search engines popularity (Fig. 2). The poll wants user to select their preferred search engine.



Which of these Search-engines is the best one?

Figure 2. An example of Internet Poll

The poll options are including popular search engine such as Google, Yahoo, MSN, and AltaVista.

The poll content is combined with the already created Concept-based MosaHIP CAPTCHA. The created CAPTCHA requires user to drop his or her desired choice on the area of mosaic picture containing the images of footballs. As you see in the Fig. 3, for participating in the poll, the Internet User has to drag his or her choice and drop it on to the area of mosaic picture containing the image of footballs. By doing so, a vote related to that choice is submitted and correctly counted. Otherwise, a new test with different images will be presented to the Users. For example if user's choice is Google, he or she should drag the moveable text object that expressing Google, and drop it on to the area of mosaic picture containing image indicated in the CAPTCHA.

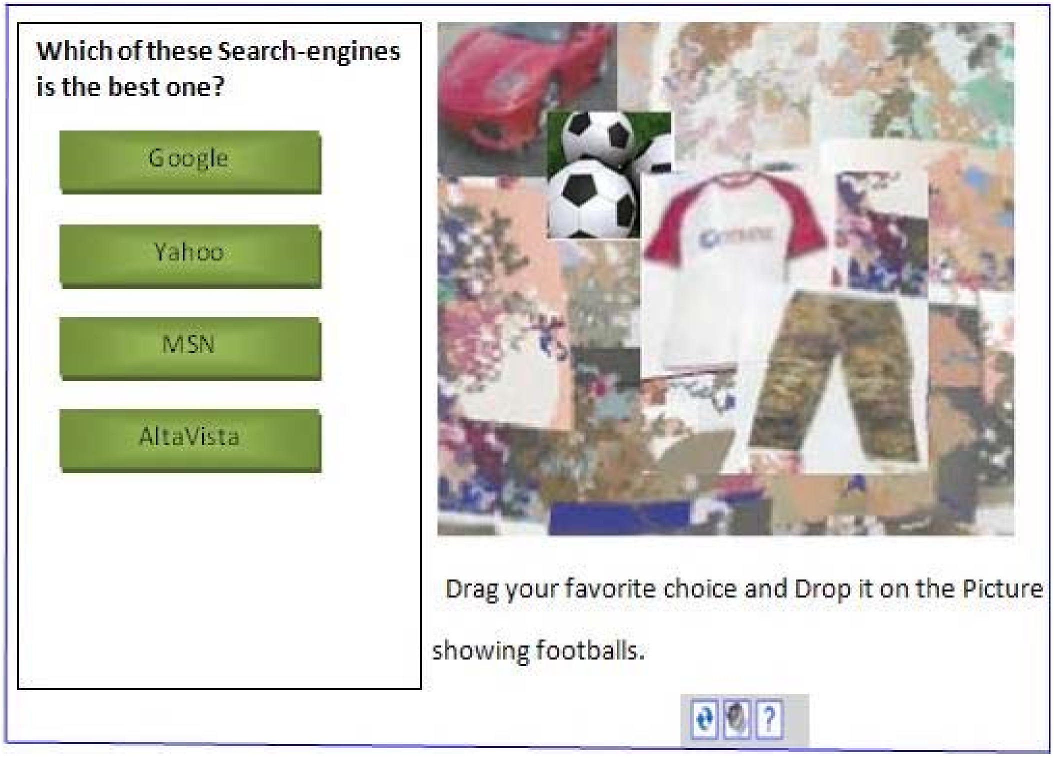
A. Making the Proposed Technique Usable for All Ages of People

We can embed an additional option to this technique. This option is a TTS (Text-To-Speech) system, in this situation Instead of showing question in the text format; the question is said using TTS System. The main advantage of this method is that it can be used by children who maybe illiterate and unable to read, since in this method the question is said using a TTS (Text-To-Speech) system. Although the main purpose of using TTS system is making the proposed method useable for all people even children, this option causes that the technique be more resistance because of recognizing the speech. In other words by using TTS system we will enhance both security and usability of our proposed technique.

B. Making Proposed Technique More Resistant

To make this technique more resistance against automatic voting, we can combine different protection techniques. The IP Locking technique can be used with the CAPTCHA, Therefore, if the numbers of subsequent failed attempts of Polling reach to a predefined threshold, then IP Locking can be used. The locking time can be exponentially increased with the number of subsequent failed attempts of polling. When the locking time expires, a new "Concept-based" MosaHIP test is presented to the user. IP Locking might cause a partial "Denial of Service". This might be a threat to fairness of technique.

A.



Which of these Search-engines is the best one?

Drag your favorite choice and Drop it on the Picture showing footballs.

Figure 3. The proposed technique

C. Advantages

Some important features of proposed method are as follow:

1) The method can be considered *Fair*, because everyone can participate in the poll. Using a visual CAPTCHA, a blind person, or one who is visually impaired, faces a lot of problems. However, these problems can be solved by using audio security challenge.

2) The method can be considered *Resistant*, since the "Concept-based" MosaHIP that is used in this technique guarantees high level of security.

3) The method can be considered *Robust* due to having strong CAPTCHA properties.

4) All ages of people can use this poll system.

5) Being based on a drag-and-drop approach, it alleviates the user from the discomfort of typing any text before submitting his or her vote, and allows a simple integration inside Internet portals and web pages.

V. CONCLUSION

In this paper, a secure online poll system is presented. Protection mechanisms used for stopping automatic voting and securing Internet polls are examined. We have proposed a two-phase technique for Internet poll, which is based on Image-based CAPTCHA. The first phase is named Creation phase and the second phase is named utilization phase.

In the creation phase, a CAPTCHA test is constructed. This CAPTCHA is based on mosaic image.

Since the constructed CAPTCHA consists of fake and real subimages, web bots cannot distinct between the real image and the fake one, but human user can. The important feature of mosaic image is that subimages are pseudorandomly positioned within the mosaic image and are also partially overlaying each other, so that separating them into subimages is not an easy task for a computer.

In this technique, An Internet User drags a poll's option and drops it on to area of mosaic picture containing indicated Image. We also have proposed to use TTS system. In this situation, instead of showing question in the text format; question is said using a TTS System, in this way the proposed method will be useable for children. We have suggested that IP Locking technique can be used along with proposed technique to make the proposed method more secure.

REFRENCES

[1] L. von Ahn, M. Blum, and J. Langford, "TellingHumans and Computers Apart Automatically,"Communications of the ACM, vol. 47, no. 2, pp. 57-60, February 2004.

[2] A. Basso, F. Bergadano, I. Coradazzi, P.D.Checco , Lightweight security for internet polls," EGCDMAS. INSTICC Press; 2004.

[3] A. Basso, S. Sicco, "Preventing massive automated access to web resources," computers & security 2 8 (2009) 17 4 – 188. Elsevier; 2008.

[4] Basso, M. Miraglia," Avoiding massive automated voting in internet polls," STM2007. Electron. Notes Theor. Comput. Sci.2008; vol. 197(2). Elsevier

[5] K .Chellapilla, K .Larson, P.Y. Simard, M. Czerwinski. "Building segmentation based human-friendly human interaction proofs (HIPs)." In: Baird HS, Lopresti DP, editors. HIP. Lecture notes in computer science, vol. 3517. Springer; 2005.

[6] M. Chew, J.D. Tygar, "Image recognition CAPTCHAs," In: Proceedings of the seventh international Information Security Conference (ISC 2004). Springer; 2004.

[7] G. Mori and J. Malik, "Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA, "Proceedings of the IEEE CS Society Conference on Computer Vision and Pattern Recognition, Madison, pp.134-141, 2003.

[8] G.Ollmann , "Stopping automated attack tools," Whitepaper – NGS software insight security research,http://www.ngssoftware.com/papers/StoppingAutomated AttackTools.pdf; 2005.

[9] reCAPTCHA: stop spam, read books. Department of Computer Science, Carnegie Mellon University, http://www.recaptcha. net/; 2007.

[10] A.S. Tanenbaum , "Computer Networks", 4th Edition, Prentice- Hall, 2003.

[11] The CAPTCHA project: completely automatic public Turing test to tell computers and humans apart. Department of Computer Science, Carnegie Mellon University, http://www.captcha.net;2000.

[12] K. Yanai, M. Shindo, and K. Noshita, "A fast image gathering system from the World-Wide Web using a PC cluster," Image and Vision Computing, Vol. 22, Issue 1, pp. 59-71, January 2004.