Research of the Chinese CAPTCHA System Based on AJAX

DONG Chen
School of Computer Science and Information Technology
Zhejiang Wanli University
Ningbo, Zhejiang 315100
P. R. CHINA
http://www.zwu.edu.cn

Abstract: - In this paper, we firstly introduce that CAPTCHA technology development and the various type of reading CAPTCHA, sound CAPTCHA, graph CAPTCHA at present. The second, we build one kind of the Chinese CAPTCHA model. The third, a new Chinese CAPTCHA system based on AJAX is proposed, which proceed to generate character characteristic based random problem and answer. On its performance, we show that the advantages of the Chinese CAPTCHA system due to its improved security and the reliability of the technology using it.

Key-Words: CAPTCHA, AJAX, system, network, attacker, random, XML

1 Introduction
With the development of information technology and increasing frequency of information exchange, the research on information security engineering is becoming more and more important. The attacker damage to the normal operation of the network order and threat to network security. The network is that people exchanges information and the sharing resource platform mutually, but it is full of large amount of rubbish information. The attacker attack the use of public network services through the network automatically, achieve their personal ends. For example, published the contents of garbage increased burden on the web site administrator; to spread propaganda and fishing sites on the internet for fraud; on-line up to vote, contrary to the principle of fairness; crazy web services account registered so that the slow pace of the server; violent crack user passwords so that the user information leak, and so on[1,2,3,4,5].

An attacker can have a network because of damage to the same site in response to the attacks and the legitimate users to submit the request. If the site users to be able to determine the legality of not allowing an attacker to obtain the requested services, will be able to stop the attacker's automatic attack. Artificial intelligence expert referred to as a CAPTCHA (for Completely Automated Turing Test To Tell Computers and Humans Apart) to verify the security mechanism, it is difficult to solve the problem of machines to distinguish between the computer and person by setting a person it is easy to solve. Such a mechanism will help site to determine the legality of uses: Computer program (computer or robot) or human, thus ensuring the stability of the site and users of information security, so that the normal operation of the network. The term CAPTCHA was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford of Carnegie Mellon University. At the time, they developed the first CAPTCHA to be used by Yahoo. When you visit the Yahoo station and apply for a mailbox free of charge, you will be demanded to distinguish one as well as photograph having curved characters in the register web pages. If your entering is correct, you will success to open website else point out that making mistakes[6,7,8]. Not only Yahoo, such as Hotmail, PayPal still has website providing the mailbox free of charge in many having this artwork-rization characters identification system. It is simple safe mechanism which will judge "person or computer ".

CAPTCHA technology is used in multi-send technology. In order to improve the CAPTCHA, this paper presents improvement of CAPTCHA. This paper is organized as follows. In section 2, In order to give a more comprehensive and precise understanding of the problem concerned, we review the concept and classification of CAPTCHA. In section 3, we propose the thinking of the Chinese CAPTCHA system, which proceed to generate character characteristic-based random problem and answer. In section 4, the Chinese CAPTCHA system solutioned and testing. In section 5, we analysis the test result which from section 4. Finally in section 6, we concludes the paper and verify the effectiveness of our proposed solutions through analysis and computer simulation.
2 Overview of CAPTCHA
How to make a secure and apt CAPTCHA is the essential of the CAPTCHA in the practice. More and more people study the CAPTCHA technology at present, and more and more websites use the CAPTCHA technology. CAPTCHA has the various type of reading CAPTCHA, sound CAPTCHA, graph CAPTCHA, their defect and advantage having each mainly[9,10,11,12].

2.1 Reading CAPTCHA
If a consumer required that goes to CAPTCHA in distinguishing a photograph, it is reading CAPTCHA (Fig. 1, Fig. 2).

![Fig. 1 The CAPTCHA of Yahoo](image)

![Fig. 2 Simple defect gimpy series](image)

The character of the photographs has carried out warping, deformation. Reality has three kinds mainly in application: Simple defect gimpy series (Fig. 2), the worst printed pessimal print (Fig. 3) and mix up the characters baffle text (Fig. 4).

![Fig. 3 Pessimal print CAPTCHA](image)

For avoiding a "robot" entering online chat, the room spreads rubbish MSG|WTG, gimpy series CAPTCHA is a company of Yahoo in 2000 year. Entrust Carnegie Mellon University to be that the same stuff that they design out defense mechanism. It requires that the consumer mounts warping by the fact that rightness imports a photograph, the deformable character passes a testing[13,14].

Pessimal Print CAPTCHA is that morals (Henry Baird) is in what 2000 year designs that by Paluoaoduo research centre (PARC) Bell. It requires that the consumer tests by the fact that the degenerate on importing a photograph, blurred correct individual word passes.

Baffle Text CAPTCHA is that morals and mound (Monica Chew) of University of California Berkeley branch school develop in cooperation in 2003 years by Bell. By the same token, it also requires that the consumer a degeneration passes on correct entering photograph, and or add the subtraction arithmetic to pass a testing the character handling the day after tomorrow's.

![Fig. 4 Mix up the characters CAPTCHA baffle text](image)

2.2 Sound CAPTCHA
This way is to emit a sound from the computer, requires that a user carries out sound identification. It broadcast the English letter and the Arabic numeral generally. This has been resolved difficulty having obstacle on the color and optesthesia. But, this kind of CAPTCHA still has no way to avoid upper barrier of language, problem is to use strange language describe that user is in the cards having no way to test smoothly by the fact that this is similar to CAPTCHA's[15,16,17].

2.3 Graphics CAPTCHA
Kind of CAPTCHA is able to demonstrate one here till several photographs, requires that a user answers the pertinent problem. Simpler practice is to make slight warping, to ask about a user afterwards with the photograph having some specially appointed scenery[18]: What photograph is this? Or several photographs having the common element is set free...
together, ask about a user: Where these photograph common point are in? Although what the computer is after all among could hardly distinguish a photo, this way is able to receive language barrier. For example, when system requires that a user imports English, language is a very deep chasm to the user who knows English not very. Besides, the same thing has firmly believed that the result also may have diversity in everyone's subjectivity. Somebody regards as a beefsteak, somebody also regards as meet. So the graphics CAPTCHA has been designed that in system is able to increase by a time play style selections only let a user choose standard answers among them generally, but the computer gets 1/30 by test chance right away if system only has thirty mark answers. This will reduce the CAPTCHA security by a wide margin.

2. 4 The Application of CAPTCHA

2. 4. 1 Vote Online
Www.slashdot.com carries out a "dangerous" investigation on September, 1999: "Which computer learned enthusiastically by is that the graduate is best?" Cast a ballot systematically at that time being able to remove the ticket number coming from identical IP address such as the great majority system. But the CMU student has used one cast a ballot procedure cheat system, the result CMU ticket number increases by very fast. the MIT student have made self casting a ballot also immediately procedure in a second day. This casting a ballot thereupon on line has become competition of two university. Their ticket number has all exceeded 20,000, but other school has all been less than a thousand. Similar circumstance has chosen middle through public appraisal also appearing in film actor of our country. By the same token on other various it may appear in casting a ballot. Can you believe such result?Thus need to use confirmation only when person ability casts a ballot. Therefore we use CAPTCHA to resolve the problem.

2. 4. 2 Free E-mail Service
It is easy to hit by the inner logon several thousands mails free of charge account number in one second by network robot procedure, And then, the number making use of these accounts announces the rubbish mail, invades and occupies the network bandwidth, wastes hardware resource. Register unless being that one people distinguishes the day after tomorrow before the computer, can make use of CAPTCHA to deter these robot procedure from gaining the account number casually otherwise. CAPTCHA also can make contribution out in the field of rubbish mail, worm virus mail identification. The regulation can be built: Unless being the mail that one people dispatches on the computer, you accept just now, otherwise refuse right away. Www.spamarrest.com studies this aspect marketplace applies in course of[19,20].

2. 4. 3 Search Engine Procedure
At present, this is equal to saying to web page robot procedure "do not need an index, please" when cover a page as with a net if needing to decline the index by search engine, need to do mark in HTML. It is feasible to inspect with serving business to search engine of important company. But if being going to ensure that web page is not needed right away by the index completely, make use of CAPTCHA. It is very important.

2. 4. 4 Shopping Intelligence Agency
When meeting unfinished products promotes the sales of some website. with the pole, low price sells. That these shopping intelligence acts for a software is able to often scan these website, builds the article gauge at a low price, rushes to purchase the sales promotion product even. This is extremely harmful to promoting the sales of the company Er Yan. They also require that the people who judges being hiding self behind computer during the past a CAPTCHA system is still computer.

2. 5 Major Advantage and Defet of CAPTCHA

2. 5. 1 Major Advantage of CAPTCHA
1) There exist deformation, affine alternation, Shrink and Grow in character.
2) The disturbance there is existed in gleam in the image.
3) The need there existing the above 6 characters in sort is distinguished.
The three characteristics have all increased degree of difficulty for identification. The confirmation code recognition usually needs the following three steps.

2. 5. 2 Major Defet of CAPTCHA
Because in this kind of CAPTCHA request operation simple, flair can provide by CAPTCHA is comparatively low.

3 Chinese CAPTCHA System
3.1 Current Situation

At present, many of the platform design can not be separated from CAPTCHA. CAPTCHA is the principle: every time the page to submit information, the system will automatically generate a random string of numbers or symbols (the CAPTCHA) only in designated areas enter the correct code to verify those to submit information to be successful. So now a lot of the registration page, the user will be asked to submit information when they enter the CAPTCHA there are also a number of malicious programs, continue to account for the specific password attempts, in the absence of adequate measures, the password is very likely to be out exhaustive. Therefore, Internet banking and other important login page, usually to verify the use of encryption technologies CAPTCHA to prevent a hacker on a particular registered user to use violence to break a specific procedure carried out continuously to prevent or try to log on to a continuous network of information submitted to the use of a network access methods.

CAPTCHA is applied to in the homeland. Especially owing to the version CAPTCHA, system application sees much most. Many station of in the homeland has also adopt CAPTCHA system. It is used to distinguish human being be still machine during the period of the system applying to be put into use. A lot of systems use figure CAPTCHA (Fig.5) or letter CAPTCHA (Fig.6).

CAPTCHA is at present middle in realizing Chinese, already can adopt various choice to mount integrated output in different platform. Use the programming language having the image output function such as ASP.NET, JSP, PHP. We can adopt separation system certainly, having gone so long as relevance web calling the CAPTCHA server serves.

It is generally Chinese CAPTCHA style in common, system will chang a lot of unlike the form and the color Chinese character as follows (Fig.7)

4 Chinese CAPTCHA Solution

4.1 The Applications Interactively of Traditional Web

Users fill out the form field and click the submit button in general traditional web applications. Then the entire form is sent to the server, the server will be transmitted to deal with it form the script (usually PHP or Java, may be the process of CGI or similar language), the script then sent back to complete the implementation of the new page. The page may have been filled with new forms of data, or it may be confirmation page, or have some of the original input data options page. The script on the server or the new procedure and return to form must wait. Into a blank screen, and then wait until the server to return data to re-draw. Fig.8 shows the course.
4.2 The New Interactive Web Application

Ajax full-called "Asynchronous JavaScript and XML", It is compound from the asynchronous communication technology by JavaScript, XHTML, W3C DOM, the core is JavaScript and XMLHttpRequest. XHTML make CSS Standardization, W3C DOM Analysis from the server-side XML information and respond to the dynamic display of information, XMLHttpRequest send asynchronous request and receive response. JavaScript achieve all logic in the browser.

AJAX solves both of these problems. AJAX works by sending an XMLHTTP request instead of an HTTP request. XMLHTTP requests are asynchronous, so while the XMLHTTP request is being processed behind the scenes, the user can continue interacting with the web application. And when the XMLHTTP response is received, the Internet Explorer DOM can be used to repaint only the area of the web page that holds the new data, rather than having to repaint the entire page. Fig.9 shows the course.

![Diagram](image_url)

Fig.9 The traditional web application.

4.3 Chinese CAPTCHA Step

In order to more suitable China, we consult the CAPTCHA system that characters distinguishes abroad owing to English, build one kind of right Chinese CAPTCHA model. The following is that this applies a step from character collection (if CAPTCHA choosing 4 separate words in GB2312) distinguishes a scheme's:

1) Users log on to the web site, a request to start the registration process.
2) Randomly select a number of figures, such as 4, at the same time they will be stored in the server session state.
3) Random selection show figures.
4) In the server memory in this generation of 4-digit map.
5) Randomly selected image transformation algorithm.
6) Random conditioning transform parameters. Memory image of the implementation of the transformation.
7) Users log on to the web site, a request to start the registration process.
8) Memory image of the implementation of the transformation.
9) Random selection algorithm and generates the background image.
10) Superimposed digital image and background image.
11) Output to the user registration page, including synthetic image.
12) Web users to enter information and images to fill in recognition of Chinese characters, submitted to the server.
13) Server to determine the number of users identified with the digital preservation of Session.

In this program, CAPTCHA test procedures and registration page on the same server running close together.

4.4 Analysis of System

At present, we adopt specifically for average Chinese character recognition "Statistical Pattern Recognition" and "Structural Pattern Recognition", still these the method combination way. But the Chinese CAPTCHA system on the contrary. It has fully utilized these method defects and graphic mixed transformation. It have enhanced identification degree of difficulty. Fig.10 shows the model.
4.5 Chinese CAPTCHA Algorithm

First, known number aggregation is defined:


2) Chinese character set CharacterSet = {MATERIAL, COLOR, SIZE, FON-TSTYLE, POSITION...};


4) Chinese style set FontStyleSet = {FontStyle. Strikeout, Font-Style. Underline, ...}

Besides, there are still if relative location, characteristics such as character size, character value size assemble, list here right away no one by one.

For example, that the chinese alphabetic string length is N on generated photograph. Upper content of photograph is composed of character characteristic-based random problem and alphabetic string: PictureContent=QUESTION + ANSWER.

And then, the random chooses the character, characteristic and characteristic value:

1) From carrying out N in Alphabet electing N time randomly, the character composes alphabetic string: stringArray={ Si | Si ∈ CharacterSet, i=1...N };

2) Color electing N from second random carrying out N in color set when kind of color is every character display: colorArray={Ci | Ci ∈ ColorSet, i=1...N};

3) Size that time of random elects from the N being in progress in FontSizeSet when N character style value big or small is every character display: fontSizeArray = {FSi | FSi ∈ Font-SizeSet, i=1...N};

4) Form that time of random elects from the n being in progress in FontStyleSet when the n character form is every character display

fontStyleArray = {FSti | FSti ∈ FontStyleSet, i=1...N}.

Then we use the form (Table. 1) explain the relation between Chinese character and the Chinese character characteristic.

<table>
<thead>
<tr>
<th>Chinese character</th>
<th>characterValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>Size</td>
</tr>
<tr>
<td>1</td>
<td>C1</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
</tr>
<tr>
<td>n</td>
<td>Cn</td>
</tr>
</tbody>
</table>

C: Color variable
S: Size variable
F: Font style variable

Then system proceed to generate character characteristic-based random problem and answer.

1) First, choose one or many characteristic and small advantages random characteristic value in the characterset, character∈CharacterSet, characterValue=X, X ∈ characterArray:

2) Generate the random problem, for example: question=”请找出图片中”+ characterValue +”的汉字”

3) At last system will generate question and answer random. Then it will mate character value is equal or no equal or not, and left of be the initiation order, it will be composed of answer alphabetic string answer.

Final CAPTCHA photograph display holds a page in the customer surface. Question and stringarray as content, it is demonstrated on the photograph and use the image module or the artwork kind warehouse.
4.6 Main Procedure

4.6.1 Designe Program
Therefore, we designed a program based on AJAX. The following are figures from such a select number of CAPTCHA number of steps to identify the application program.

4.6.2 Declaration to Protect the Resources
In the application documents to describe the deployment of (web.xml), respectively, on the <web-resource-collection>, <auth-constraint>, <user-data-constraint> element in the definition, namely the realization of the declaration to protect the resources, and the roles and transfer protocol. And to increase <login-config> element in the definition of user submitted after the jump page.

4.6.3 Make a CAPTCHA Image
1) Generation program
```php
$im=imagecreate(70,35)
$gray=ImageColorAllocate($im,200,200,200);
//For the regional distribution of a color image.
imagefill($im,68,30,$gray);
//Regional $gray image as a background color for filling.
while(($authnum=rand()%10000)<1000);
//Have a design in line with the requirements of the random number.
imagestring($im,5,10,3,$authnum,$black);
//To the level of function will be randomly generated value in painting the image in the designated location.
for($i=0;$i<200;$i++)
//In the image to add extra pixels to avoid scanning through the graphical approach to the random number to obtain information.
{$randcolor=ImageColorallocate($im,rand(0,255),rand(0,255),rand(0,255));
//A randomly generated color pixels.
imagesetpixel($im,rand()%70,rand()%30,$randcolor);
//In the image of the region on the location of the random drawing pixels on this point.
}
ImagePNG($im);
//PNG image format to be exported to the form page display.
ImageDestroy($im);
//Once completed picture output would destroy the image, the release of its image associated with memory.
```

2) Ajax graphics mode to achieve the procedure code:
```javascript
(1)Ajax engine in the user's browser to achieve:
<script language="javascript">
var httprequest=false;
httprequest=new ActiveXObject("Microsoft.XMLHTTP");
//IE-based browsers have a XMLHttpRequest object, the Ajax engine that is the core of the object.
document.getElementById("codeimg").innerHTML ="<img src="image.php"height="35"width="70">";
//Positioning the graphics display area code.
function loginRequest(url){
//This function is a function of the XMLHttpRequest object to the adoption of asynchronous data to the server-side authentication procedures to determine.
url="checkcode.php?checkcd=\+checkcd;";
//Location of the server-side validation process.
loginReq.open("post",url,true);
//Through the post to send a request to the server.
loginReq.onreadystatechange=processLoginResponse;
//After a complete authentication server processLoginResponse function will be to complete.
}
function processLoginResponse(){
//This function to deal with authentication server after the completion of the follow-up to deal with.
var res=loginReq.responseText;
//XMLHttpRequest object to the back-end server to verify the results given the value of variable res.
if(res==1){
window.alert("CAPTCHA error!");
//To determine an error code and prompts the user.
document.getElementById("codeimg").innerHTML ="<img src="image.php"height="35"width="70">";
//Positioning the new code of the graphics display area will have a new code of graphics which users enter a new code.
}
else
window.alert("CAPTCHA correct!");
//Code to determine the correct forms and information submitted to the server storage.
}
</script>
```

(2) The graphics display area code
```html
<span id="codeimg"></span>
```
//Span tag graphic elements on the ground to generate the code generated graphics procedure provides a display space, Ajax engine through the
elements of the id to verify the location graphics display area of the code.

4.6.4 The Main Server-side Code to Determine Procedures
$q\text{query} =$ "select count(*) from checkcode where chk_code='$\text{checked}'";
// The user to input the code for information, the server side of the database to find and return to the eligible number of items of data.
$count=$mysql_result($result,0,0);
// Implementation of the above statements and the results given in the variable $\text{count}$. 
if($\text{count}==0) 
\text{res}=1;
// $count = 0$ that did not meet the requirements of the project data that the user enters the wrong code.
else 
\$\text{res}=0;
// Otherwise, show that the user enters the correct code.
\text{echo}$\text{res};
// The results will be conveying this message to XMLHttpRequest object, and to bring back to the Ajax engine.

4.7 Solution
We can get several kinds the following types after accomplish system.

1) Size change CAPTCHA: This pattern is that the random appears from change Chinese character size(Fig. 11).

2) Size and color chang: This pattern is that the random appears from change Chinese character size and change color at the same time (Fig. 12).

5 Analysis of the Chinese CAPTCHA

5.1 Analysis of Security
The computer can not the character automation be understood for problem and the answer since current artificial Intelligence understands the difficult problem there existing in aspect in semanteme. The problem answer is also random, for example the random characteristic, random characteristic value according etc. So the attacker can not build questions and answers coming to break a solution the warehouse. It has reliable security.

5.2 Analysis of Using Kind Easily
The kind of Chinese CAPTCHA is using kind easily. It is easily and clearly, Chinese character’s background is disturbed or color is changed. It will be difficult to declassification to computer, but it is easy to human being to identification, unless to judging that color person having obstacle have no way to be completed.
5.3 Analysis of the Characteristic Chooses Tactics
How to choose the Chinese character characteristic value is the key designing Chinese CAPTCHA system successfully. Human being ability of character identification and semanteme is different from computer. It will help us to look thing oversimplifies and structure-rization and make the visual information change to meaning group. Characteristics of perceptual constancy provide human being with one kind of the basis judging object's, it is unable to change the judgement to object because of external environmental factor casually. Perceptual constancy in common include of size constancy, color constancy, shape constancy. Besides, material is hard to identification to computer. So we can guard against machine recognition coming thinking with material quality. Therefore, we have chosen easy person eye perception such as character color, size, background’s, and the computer has not exchanged the perception characteristic.

6 Conclusion
This paper introduces a technique that allows writing lightweight test automation to verify the functionality of AJAX Web application. The technique easily generalizes to arbit rarily complex applications and works for any AJAX enabled application, regardless of the implementation technology.

The test automation system is lightweight and is designed specifically to extend easily. Notice that the test method is not fully automated: it has to manually click on the Run Test button to launch the automation. Another extension to the AJAX test automation is automating the process of saving test scenario results. Both of them need to further research and design.

However, we have just studied the small part of it. There are many actual problems to need the study and solving. To study better method of CAPTCHA is our future work.

References:
