


CAPTCHA


Roberto Gómez Cárdenas
rogomez@itesm.mx
<http://homepage.cem.itesm.mx/rogomez>

Lámina 1 Roberto Gómez C.



CAPTCHA

- Completely Automated Public Turing test to tell Computers and Humans Apart
- Invertir la prueba de Turing
 - Prueba de Turing: ¿cómo deducir si se esta interactuando con una computadora?






Lámina 2 Roberto Gómez C.



Precursores

- Manuscrito no publicado de Moni Naor, menciona por primera vez la automatización de la prueba de Turing en 1997.
 - Ninguna propuesta o formalización.
- Patente de Alta Vista en 1998 fue el primer ejemplo de uso de imágenes distorsionadas de texto para detectar robots.
 - Tiempo después fue roto por OCR

Lámina 3

Roberto Gómez C.



Definición

- En el 2000, fue formalizada por
 - Luis von Ahn, Manuel Blum & Nicholas J. Hopper de Carnegie Mellon;
 - John Langford de IBM
- “A CAPTCHA is a cryptographic protocol whose underlying hardness assumption is based on an AI problem.”
- Sitio web;
 - www.captcha.net

Lámina 4

Roberto Gómez C.



Enfoques

- Texto (ASCII/Unicode)
- Imágenes
- Voz
- Animaciones
- 3-D
- Combinaciones

Lámina 5

Roberto Gómez C.



Voz

- Vocalizar en voces sintetizadas o grabadas.
- Reconocimiento de voz vs rango de falla del usuario.
- Utilizado con CAPTCH visual para incrementar accesibilidad
 - Puede ayudar a los atacantes a adivinar de forma correcta.

Lámina 6

Roberto Gómez C.



CAPTCHA Animado

- Se puede usar Flash, MPEG y GIF animado.
- Generalmente combinado con voz.
- La debilidad del CAPTCHA de imagen esta presente.
- Usualmente fácil de romper debido a la información extra para su análisis.
- Mucho carga para el procesador y tráfico.
- No práctico en muchos casos.
- Ejemplos en: <http://captcha.net>

Lámina 7

Roberto Gómez C.




ASCII/Unicode

- Cambiar el texto para que no se entienda
 - SPAM = \$P4M
- Caracteres a palabras
 - rogoomez@itesm.mx = rogoomez at itesm punto mx
- Usar entidades caracteres URL/HTML
 - COPY = ¢0Ρ¥
 - COPY = %430P%59
- Mejor que nada, pero facil de romper
- Técnicamente no es CAPTCHA

Lámina 8

Roberto Gómez C.




CAPTCHAS basados en texto


- Gimpy, ez-gimpy
 - Tomar una palabra o palabras de un diccionario pequeño.
 - Distorsionarlas y añadir ruido y background
- Gimpy-r
 - Tomar letras aleatorias
 - Distorsionarlas, añadir ruido y background
- Simard's HIP
 - Tomar letras al azar y números
 - Distorsionarlas y añadir arcos


Lámina 9

Roberto Gómez C.





Ejemplos







advisior













Lámina 10

Roberto Gómez C.



Tres dimensiones

- tEABAG_3D
 - <http://www.ocr-research.org.ua/index.php?action=teabag>
- El reto de presenta en una imagen de tres dimensiones.
- Más difícil de romper que el de imágenes de dos dimensiones.
- Más recursos en el servidor
 - Carga de procesamiento alta.
- Puede combinarse con otros métodos.

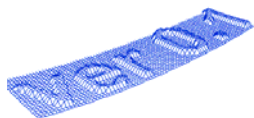
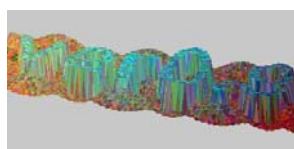



Lámina 11
Roberto Gómez C.



Dos originales

Prove you are human by reading this resistor:



00±/- 5%

0	0	0	5
3	3	3	5
4	4	4	5
5	5	5	5
6	6	6	5
7	7	7	5
8	8	8	5
9	9	9	5

Match the sliders on the left to each color band on the resistor.

[Click Here for a new resistor image.](#)

New to electronics? [Click here](#) to learn how to read resistor values.

Trace all fish



no CAPTCHA





Lámina 12
Roberto Gómez C.



Atacando a los CAPTCHA

- Ataque de hombre en medio
 - Copia el CAPTCHA de la víctima
 - Ponerlo en el website del atacante
 - Rederigir la respuesta a la víctima.
- CAPTCHA factory
 - <http://taint.org/2008/03/05/122732a.html>
- Reutilizar el identificador de sesión
 - http://www.puremango.co.uk/cm_breaking_captcha_115.php

Lámina 13

Roberto Gómez C.




CAPTCHA y PHP

- Uso de las funciones principales del lenguaje PHP.
- A través de la librería GD es posible generar imágenes (PNG) sin necesidad de diseñar filtros auxiliándose de un software de diseño.
- También se usan fonts de caracteres como el formato TIF.

Lámina 14


Roberto Gómez C.



Generando una imagen

```
<?php  
  
    header('Content-type: image/png');  
    $img = imagecreatetruecolor(100, 50);  
  
    imagepng($img);  
    imagedestroy($img);  
  
?>
```


Lámina 15 Roberto Gómez C.



Pasos

- Declarar una sesión.
 - Indispensable para “memorizar” el código escrito en la imagen.
- La prueba de la captura.
- El formulario HTML que manda llamar a un archivo externo que lleva a cabo el procesamiento.


Lámina 16 Roberto Gómez C.



Preparando el texto

```
<?php
function cadena()
{
    $md5 = md5(microtime() * mktime());
    $chaine = substr($md5,0,5);
    return $cadena;
}
$_SESSION['text']=cadena();
?>
```

Lámina 17 Roberto Gómez C.



El fondo

- El color puede ser de la elección que se desee o puede no haber,

```
<?php
//fondo negro
$fond = imagecolorallocate($img,0,0,0);
imagerectangle($img,3,3,40,40,$fond);
?>
```

Lámina 18 Roberto Gómez C.



Efecto sobre el fondo

- El efecto será de rectángulos que se superponen

```
<?php
for($cube=0;$cube<25;$cube++)
{
    $fondCol = imagecolorallocate($img, rand(0,130)+1,
                                rand(0,130), rand(0,130));

    $x=rand(0,$largeur);
    $y=rand(0,$hauteur);
    $w=rand(5,$largeur/2);
    $h=rand(5,$hauteur/2);
    imagefilledrectangle($img,$x,$y,$x+$w,$y+$h,$fondCol);
    imagecolordeallocate($img,$fondCol);
}
?>
```

Lámina 19

Roberto Gómez C.



Trazado de líneas

- Es una etapa intermedia.
- Trazar líneas entre el fondo de la imagen y el texto propuesto.
- Una línea será de un cierto color y las otras dos serán de colores diferentes.

```
<?php
    $col=imagecolorallocate($img, 50,128,250);
    imageline($img,10,10,400,150, $col);

    $col=imagecolorallocate($img, 128,250,30);
    imageline($img,0,0,39,29, $col);
    imageline($img,40,0,84,59, $col);}
?>
```

Lámina 20

Roberto Gómez C.



Trazado de elipses

- Las elipses forman parte de las etapas intermedias y también se les conoce como círculos.
- Puede ser generadas al azar.
- En este caso solo se genera una.

```
<?php
    $col=imagecolorallocate($img, 250,199,rand(128,255));
    imagearc($img, 20, 25, 60, 60, 0, 360, $col);
?>
```

Lámina 21

Roberto Gómez C.



Efecto sobre el texto

- La parte mas compleja del CAPTCH ya que diferentes criterios se pueden tomar en cuenta, sobretodo si se usan funciones propias de generación de valores aleatorios.

```
<?php
    $angle=mt_rand(10, 40);
    $font="arial.ttf";
    $x=rand(10,35);
    for($c=0;$c<5;$c++) {

        $size=mt_rand(13, 24);
        $text=$_SESSION['text'][$c];
        $y=30+rand(0,18);

        $color=imagecolorallocate
            ($img,rand(100,254),rand(100,254),rand(100,254));
        imagettfttext($img, $size, $angle, $x+15*$c, $y, $color, $font, $text);
    }
?>
```

Lámina 22

Roberto Gómez C.



CAPTCHA

Roberto Gómez Cárdenas
rogomez@itesm.mx
<http://homepage.cem.itesm.mx/rogomez>