CAPTCHA

Presented By
PANKAJ JAIN [06ESVCS030]
Definition
Background
Types
Constructing CAPTCHAs
Conclusion
Bibliography
WHAT IS CAPTCHA?

I. Completely Automated Public Turing test to tell Computers and Humans Apart.

II. Created in 2000 for Yahoo to prevent automated e-mail account registration, by Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford, Carnegie Mellon University.
Introduction

- A program that can tell whether its user is a human or a computer.
- It uses a type of challenge-response test to determine that the response is not generated by a computer.
- Generic CAPTCHAs distort letters and numbers. Humans can read the distorted and noisy text.
**BACKGROUND**

1. In November 1999, slashdot.com released a poll to vote for the best CS College in the US.

2. Students from the CMU and the MIT created bots that repeatedly voted for their respective colleges.

3. This incident created the urge to use CAPTCHAs for such online polls to ensure that only human users are able to take part in the polls.
SO, CAPTCHA IS...

- A program that can generate and grade tests that:
  - Most humans can pass.
  - Current computer programs cannot pass.
TYPES OF CAPTCHAS:

- Text Based
  - Gimpy
  - Ez-Gimpy

- Audio Based
  - MSN Passport Service

- Graphics Based
  - Bongo
  - PIX
1.) Simple, normal language questions:

- What is sum of three and thirty-five?
- If today is Saturday, what is day after tomorrow?
- Which of mango, table, water is a fruit?

2.) Very effective, needs a large question bank.

3.) Cognitively challenged users find it hard.
1. Gimpy Based-
   i. Designed by Yahoo and CMU.
   ii. Picks up 10 random words from dictionary and distorts, fills with noise
   iii. User has to recognize at least 3 words
   iv. If user is correct, he is admitted
2. Ez-Gimpy Based-

i. This is a simplified version of the Gimpy CAPTCHA, adopted by Yahoo in their signup page.

ii. Ez – Gimpy randomly picks a single word from a dictionary and applies distortion to the text.

iii. The user is then asked to identify the text correctly.
Text BASED: CAPTCHA

3. MSN Captcha:
   i. Provided for Microsoft’s MSN services
   ii. Use 8 characters
   iii. Warping is used to distort
   iv. Very strong implementation, hasn’t been broken
   v. It is segmentation-resistant
GRAPHICS BASED CAPTCHA

Graphic CAPTCHAs are challenges that involve pictures or objects that have some sort of similarity that the users have to guess.

1. Bongo-

BONGO asks the user to solve a visual pattern recognition problem. It displays two series of blocks, the left and the right. The blocks in the left series differ from those in the right, and the user must find the characteristic that sets them apart.
Graphic CAPTCHAs are challenges that involve pictures or objects that have some sort of similarity that the users have to guess.

2.PIX-

I. Uses a large database of labelled images.

II. It shows a set of images, user has to recognize the common feature among those.

III. E.g., Pick the common characteristic among the following four pictures----

"Aeroplane"
AUDIO BASED CAPTCHA

i. Consist of downloadable audio clip
ii. User listens and enters the spoken word
iii. Helps visually disabled users
iv. Below is the Google’s audio enabled CAPTCHA
v. Not popular
**CONSTRUCTing CAPTCHA**

Things to keep in mind:

a. Don’t store CAPTCHA solution in Web page’s metadata.

b. A CAPTCHA is no good if it doesn't distort.

c. Need a large database of different CAPTCHA questions.

d. Avoid repetition of questions.
Constructing CAPTCHA

- Embeddable CAPTCHAs:
  a. Available freely, just embedd code into Web page’s HTML code, from e.g., www.captcha.net.
  b. No maintenance.

- Custom CAPTCHAs:
  a. Fits to the theme of the page.
  b. Better protected from spammers.
  c. Can be written in any language—Perl, .NET, ASP, JavaScript.
APPLICATIONS

1. Protect online polls

2. Prevent Web registration abuse, protect passwords from brute-force attack

3. Prevent comment spam and spam emails

4. E-Ticketing, prevent scalping
CONCLUSION/SUMMARY

i. CAPTCHAs are an effective way to counter bots and reduce spam.

ii. They serve dual purpose—help advance AI knowledge.

iii. Applications are varied—from stopping bots to character recognition & pattern matching.

iv. Some issues with current implementations represent challenges for future improvements.
BIBLIOGRAPHY

2. reCAPTCHA (2007)
3. CAPTCHA (2007)
5. Using “captchas” to digitize books (2007)
6. „For Certain Tasks, the Cortex Still Beats the CPU“ (2007)