CAPTCHAS WITH A PURPOSE

AUTHOR: SUHAS AGGARWAL

PRESENTED BY: SRUSHTI DHOPE
INTRODUCTION

- CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
- Program that protects websites against bots by generating and grading tests that humans can pass but current computer programs cannot.

Why is it needed?
- Prevent degradation of the quality of service of a given system, whether due to abuse or resource expenditure.
- To protect systems vulnerable to e-mail spam.
- To stop automated posting to blogs, forums and wikis, whether as a result of commercial promotion, or harassment and vandalism.
CAPTCHAs With a Purpose
TIME BASED CAPTCHA

- Flash alphabets (one at a time) in different locations of the screen at small intervals. User should type the alphabets in the sequence in which they appear. The letters are distorted to make the task harder for computers.

- Display images in a sequence. Images appear for an instance and then disappear. User should give a caption to the images in the sequence in which they appeared.
TIME BASED CAPTCHA

- These CAPTCHAs expire, if they don’t receive a response in a particular time interval. Idea behind this is to make use of challenges which humans can solve quickly but computers take some time to solve.

- As user will have to quickly respond to the alphabets displayed as they appear and disappear, it can be a suitable means to measure certain reaction capabilities of human beings.

- If system is detected to be under abuse, time gaps can be increased significantly to make CAPTCHA solving time consuming and expensive.
Type The Letters You See Below Into the Box in the sequence they appear excluding repeated sequence

\[ a \]

Type The Letters You See Below Into the Box in the sequence they appear excluding repeated sequence

\[ y \]

Type The Letters You See Below Into the Box in the sequence they appear excluding repeated sequence

\[ t \]
A sentence is selected. Two random words are selected from the sentence and are swapped. A random alphabet is filled in each whitespace present in the sentence. Challenge is to write the sentence in the correct format.

Bad officials are elected by good citizens who do not vote
SENTENCE BASED CAPTCHA

Purpose: Educating people. Sentences can be general knowledge facts, thoughts of the day, recent headlines.

Sentence based CAPTCHA is a bit difficult for humans. It might also be possible to attack it using natural language processing techniques. Sentences making use of common sense facts may be more useful since machines don’t have access to this information but then this CAPTCHA will loose its practical utility.

A possible reversible attack can be to write a program which can guess correct words in the sentence and perform a search over the web and identify the correct sentence from the results obtained. A scheme which can be useful in this scenario is to omit out certain alphabets from the words and present a sentence composed of partially ‘eaten up’ words.
HUMAN EMOTION BASED CAPTCHA

- A statement or a graphic is displayed to the user. User has to type a string describing his emotion as an answer.
- Purpose: Can be used for conducting online polls on a very large scale. Sometimes when an important poll is being conducted, all answers given by the users will be marked as acceptable and their different reactions corresponding to an event can be obtained. This won’t affect the security of the CAPTCHA as such a poll will be conducted only sometimes and it won’t be possible to know when it is being conducted.
HUMAN EMOTION BASED CAPTCHA

Type emotion corresponding to the image in the box

Emotions: happy, thinking, horror, sad, shock

In the late '90s, Microsoft secretly developed its own version of Linux, but shelved it after quality control researchers deemed it "too stable".
Russell’s Circumplex Model of Emotions

- Measure for comparing emotions.
- The two principal axes are x: pleasure & y: excitement.
- For example, low excitement and displeasure correspond to ‘depression’ in the lower left quadrant.
- Emotions that are close to each other on the model are perceptually similar to humans, and vice versa. Opposite or most dissimilar emotions are diagonal to each other on the model.
- This model can be deployed in emotion-based CAPTCHA for identifying correct and incorrect answers when tags depicting emotion attached with the image are not sufficient, thus will improve the accuracy of CAPTCHA.
IDENTICAL EMOTIONS

- This game is an extension of the ESP game whose main purpose was to label images on the web. Identical emotions, instead of assigning description to the image, helps to assign emotions to the images, thus provides tags describing emotions depicted by the images.

- The game is played by two partners assigned randomly from among all the people playing the game. They are both given the same image and are required to type the ‘emotion’ the image depicts. Once both players have typed the same string, it can act as a good tag for the image, describing emotion conveyed by the image. We say that their emotions are 'identical' & they move on to the next image.
Time: 01:45
Score: 0
Passes: 0
Off-limits: My-labels

Time: 00:51
Score: 1
Passes: 0
Off-limits: My-labels
**Off-Limit Words**

- If an image has been labelled many times, system generates off limit words called ‘taboo words’. Players are then not allowed to use these words for labelling. This makes the game harder as the players now will have to think of another possible emotion which the image conveys. The first time an image is used in the game, it will have no taboo words. When it has been labelled by players in previous games, it will start having taboo words.

- An image passes through the system multiple times. When it has acquired many labels which have become off limit words, emotion becomes too difficult for players to guess & they begin to pass it each time. At this point, the image is removed from the database.
SOME MORE GAMES WITH A PURPOSE

- User are asked to enter multiple labels (maximum three) for a video or news article since they contain a lot of information & matching single common keywords can be a bit difficult. For matching, these keywords might not be same strings but can be synonyms. A subset of keywords can also be matched i.e. if more keywords match, the players get more points.

- Players can choose the category for their labels. They can communicate with each other to mutually agree on a category. Emotions can be one such category.
IMPLEMENTATION OF EMOTION BASED CAPTCHA IN AN ECOMMERCE APPLICATION: AN ONLINE SHOPPING PORTAL
SCENARIO BASED CAPTCHA

- Idea behind this CAPTCHA is to utilize the analytic and understanding capability of humans rather than merely recognizing objects.

  Monkey eating a banana
  Chimp eating

  A couple dancing
  A couple performing a ballroom dance

  A student studying at night
  A student studying for her exam
SECURITY OF CAPTCHAS

- As there are plenty of CAPTCHAs available today, a set of different CAPTCHA problems is maintained and a random CAPTCHA is selected from these CAPTCHAs before a login prompt is granted.
- This makes it very difficult to attack these CAPTCHAs. Even if a program is written to attack one type of CAPTCHA problem, after one unsuccessful try, a different type of CAPTCHA problem will be presented to the user and attacking program should be capable to solve that as well.
- CAPTCHAs such as the sentence based CAPTCHA may be a bit annoying for people to solve continuously, but in the above technique, a random CAPTCHA is presented each time, so people might encounter it only sometimes & therefore be willing to solve it to gain some useful knowledge.
- An image can be a sentence, the challenge can be figuring out a sentence, typing emotion corresponding to it, typing letters in the sentence, etc. So each CAPTCHA has a different solution which can be figured out by humans but is difficult for a computer.
SECURITY OF CAPTCHAS

- Token Bucket Scheme: Bots use small number of IP addresses that submit a very large number of incorrect responses, interspersed with a much smaller number of correct responses. This can be used to increase the security of CAPTCHAs.

- Initialise bucket with some capacity. When an incorrect attempt occurs decrease its capacity by one. Order of decrement can increase if number of successive incorrect attempts are large. When capacity becomes zero or less, even if a CAPTCHA guess is correct, treat it as incorrect. This implies that the CAPTCHA is locked.

- Example: Emotion based CAPTCHA - According to the formula of gambling, the number of trials $N$ necessary for an event of probability ‘$p$’ to occur with a degree of certainty $C$ is given by:

  \[ N = \frac{\log(1 - C)}{\log(1 - p)} \]

  Then, if we take 7 emotions, number of trials required to ensure that the CAPTCHA is solved once with a degree of certainty over 90% is given by: $\log \left( \frac{1 - 90}{100} \right) / \log \left( 1 - \frac{1}{7} \right) = 14$. 
After a large number of incorrect attempts, bucket value goes below zero and CAPTCHA is locked even though the answer is correct. The user will now have to log in to the shopping portal again to place an order.
INCREASING USABILITY OF CAPTCHAS

- Survey on people from 11th grade to PhD: Most of them found time & emotion based CAPTCHA much easier to solve than a sentence based CAPTCHA. Most of them, found emotion based CAPTCHA more fun compared to the others.

- Partial credit scheme can be used to make CAPTCHAs easier for people. For example, if in a sentence based CAPTCHA, he is able to figure out 5 out of 7 words, he is given partial credit and a second CAPTCHA is presented to him. If he is able to get that partially correct as well, he is believed to have solved the CAPTCHA challenge completely.
CONCLUSION & FUTURE WORK

- reCAPTCHA helps digitize books & is most widely used. CAPTCHAs with a purpose serve various other practical purposes & can be used in conjunction with each other to broaden the application domain of CAPTCHAs & make CAPTCHA solving more interesting.

- Keyword extraction from articles using statistical algorithms and natural language processing techniques to analyze data, extracting keywords that can be utilized to index content, generate tag clouds. The number of tags it generates is large thus compromising quality. For indexing purpose, if these tags /keywords can be filtered out/ranked based on human computation game, it will greatly enhance tag quality thus enhancing the quality of index.

- A game can be designed to obtain summaries/pictorial representation of articles which makes them easy and quick to understand.

- Developing games which can help in translating online portals such as shopping & travel portals to different languages.