CAPTCHAs- Literary review

Origin of CAPTCHAs

- It was first mentioned in 1996 paper by Moni Naor.
- The idea was taken from the Cryptography where sender send a challenge to reciever to authenticate it’s identity.
- Many ideas were given like- recognising male or female, checking facial expression -happy or sad,filling in words or disambiguation(what does it refer to?).
- It was first developed by Alta Vista to prevent “bots”.

Captchas-using hard AI probelms for security.

-Initially, it was thought that Captchas can be used for security because there had not been any AI problem which could have broken a captcha.
- A CAPTCHA is a cryptographic protocol whose underlying hardness assumption is based on AI problem.
- It was being said that if AI can break Captcha, then it will solve a hard problems.

Why CAPTCHAs

-Online polls
-Free email services
-Search engine Bots
-Spams
-Prevention of dictionary attack

Why character recognition in Captchas

- Recognition of objects requires prior knowledge of the scene and different regions have different names of an object. It would not have been universal.
- character has unique data set and all of it is present on keyboard clearly. No confusion would be there.
- Also, no prior knowledge is required by the user.
- characters were designed by humans for humans and humans have been trained at the task since childhood.
- character-based HIPs can be generated quickly.

Different types of CAPTCHAs

How Captchas are build:

Building an actual reading-based CAPTCHA requires one to make several independent choices:

a) Character set: The character set to be used in the HIP.

b) Affine transformations: Translation, rotation, and scaling of characters

c) Adversarial clutter: Random arcs, lines, or other simple geometric shapes that intersect with the characters and themselves

d) Image warp: elastic deformations of the HIP Image at different scales

e) Background and foreground textures: These textures are used to form a colored HIP image from a bi-level or grayscale HIP mask generated by using a) through d)

Work done so far in breaking Captchas

- Kumar Chellapilla, Patrice Y. Simard Using Machine Learning to Break Visual Captchas-
- They suggested segmentation and then recognition and then check their results . They tried on different Captchas of Mailblocks, MSN , Ticketmaster, Yahoo, Yahoo v2

Anjali Avinash Chandavale and A. Sapka . Security Analysis of CAPTCHA .

In this paper they dicussed attacking the CAPTCHA as the three step process:
1)Preprocessing
2)Segmenting
3)Recognizing

Out of this segmenting is the most difficult part.

[3] Building Segmentation Based Human-friendly Human Interaction Proofs (HIPs)

- Comparing computer and human performance, we see that computers are better than humans at recognition (when segmentation is solved)
- Segmentation is intrinsically difficult for both computers and humans because:
    -Segmentation is computationally expensive

Decided workflow for project:

1)**Preprocessing**
2)Segmenation
3)Training the data set
4)Testing

References:

- Kumar Chellapilla, Patrice Y. Simard Using Machine Learning to Break VisualHuman Interaction Proofs (HIPs) Microsoft Research, one microsoft way, WA 98052 -2005
- Anjali Avinash Chandavale and A. Sapka . Security Analysis of CAPTCHA, 2012
- Kumar Chellapilla, Kevin Larson, Patrice Y. Simard, and Mary Czerwinski-Building Segmentation Based Human-friendly Human Interaction Proofs (HIPs) Microsoft Research, One Microsoft Way, Redmond, WA, USA 98052 ,2005