

Issues for Chinese CAPTCHA

Ke Yang (yangke@cs.cmu.edu)

ABSTRACT

According to Accenture, Chinese will become the #1 Web Language by 2007. It is important to have the Chinese version of CAPTCHA by then. We discuss some special issues about Chinese: Chinese characters are denser, and allow much more deformation. These features may allow for more interesting ideas/constructions of CAPTCHA.

Some ideas might also be useful for general CAPTCHA.

Below are some issues:

1) Chinese is denser: there are many short sentences of complex meanings and many idioms. --- an example is "Shou3-Zhu1-Dai4-Tu4", which means "Stand by a stump waiting for more suicidal rabbits", which roughly means "daydreaming the impossible things".

The idea is: can we pack a semantically meaningful sentence into the challenge?

examples:

CHALLENGE = CAPTCHA("to be or not to be")

RESPONSE should be "shakespeare" or "hamlet"

So that there isn't a one-to-one correspondance from the text of the challenge to the response.

2) Calligraphy.

It seems Chinese Calligraphy allows for more deformation than English.

Change of approach?

English CAPTCHA: start from words --> add noise

Chinese CAPTCHA: start from sentences --> calligraphite

3) Simplified/Traditional Chinese.

Mainland: simplified, most texts are read from left to right, and then from top-down.

Hong-Kong and Taiwan: Traditional, many texts are read from top-down, and then from right to left.

The difference is quite significant.

Inputing Chinese is also a problem: not a very satisfied chinese input method IMHO.

It might make sense to do multiple choice instead of asking for a response written (typed) in Chinese.