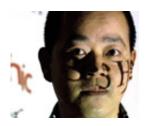# Touch typists could help stop spammers in their tracks

July 15 2009

(PhysOrg.com) -- Computer scientists at Newcastle University are about to give office workers a perfect excuse to play games: it's all in the name of research. Dr Jeff Yan, together with his PhD student Su-Yang Yu, has created 'Magic Bullet' as an effective solution to a problem which no known computer algorithm can yet solve.

This simple computer game turns a tedious manual labelling task into a form of light entertainment and could soon help companies improve their chances of tackling online spammers.

CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) is widely used by commercial websites such as Google and Yahoo to defend against malicious Internet bots which spread junk emails or grab thousands of free email accounts.

A common approach to testing its robustness is to try and attack or break

the scheme. This involves acquiring a set of labelled samples, but as computers find it difficult to recognise distorted test or images, this task usually falls to human researchers.

"Manually labelling samples is tedious and expensive," explained Dr Yan, who led the research. "For the first time, this simple game turns it into a fun experience with a serious application as it also achieves a labelling accuracy of as high as 98 per cent."

Spammers can make a lot of money out of using computer programs that can automatically bypass a heavily used CAPTCHA such as those used by Google, Microsoft and Yahoo. Therefore it is important for researchers to understand and improve the robustness of the system in order to stay one step ahead.

To fully evaluate the robustness of a CAPTCHA scheme at least 10,000 segments usually have to be labelled - a task which cannot be automated.

Magic Bullet is a dual-purpose online shooting game that can be played just for fun but also contributes to solving a real problem.

Players are randomly pitched against each other, with two in each team. Teams or players cannot communicate with each other and security techniques are used to ensure they are geographically apart to reduce the likelihood of cheating.

If there are not enough human players, then one of two types of bots - a Data Relay Bot which replays data from old games or a Tailored Response Bot which acts according to an opposing team's performance - will be introduced.

During each round a randomly chosen segmented CAPTCHA character appears and will shoot towards the target only when both players

correctly identify it before their opponents. Although the computer does not know which character each of the segments is, the answers given by the winning team can be accurate labels for the segments in the majority of cases.

The game also includes a high scoring table to encourage players to return to improve on a previous score.

"An average game session produced 25 correct labels per minute, giving 1,500 per hour," explained Dr Yan. "Although this is not particularly fast, if touch typists were used it would be noticeably improved, and also players need time to get to know how the game works.

"As this game supports a large number of parallel sessions, which are limited only by the network bandwidth and game server's CPU and memory, there is also a lot of scope to increase the labelling rate dramatically."

Dr Yan will be presenting his findings at the IJCAI'09, a leading artificial intelligence conference in Pasadena, CA, USA which runs from 11-17 July 2009.

Provided by Newcastle University