Scientists help Microsoft and Yahoo improve online security

October 21 2008

(PhysOrg.com) -- Computer scientists at Newcastle University have cracked the security behind the biggest names in global email services.

If you've noticed a reduction in the amount of email spam in your inbox lately, it could be thanks to computer scientists at Newcastle University.

Dr Jeff Yan and PhD student Ahmad Salah El Ahmad recently became the first people to crack the security behind the biggest names in global email services, exposing widespread vulnerability.

Yahoo and Microsoft believed they had systems in place that were secure enough to stop widespread abuse by spammers, but the scientists discovered that even the best on the market offered little more than a 'false sense of security'.

But, unlike the hackers who exploit cracks in the system for their own gain, they used their knowledge for the greater good and took their findings straight to the companies.

The security system in question is CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart), designed to prevent automated hacker attacks where a computer is set up to constantly bombard an online system with junk.

Anyone surfing the web will have encountered a picture of wavy, distorted letters which have to be deciphered and typed into a box before

accessing email accounts, joining social networking sites such as Facebook, or posting a comment on a website: this is a CAPTCHA scheme.

It is standard technology used to defend against malicious automated 'bots' - which can grab thousands of free email accounts in order to continuously spread junk emails or post adverts on blogs - and is used by Microsoft, Yahoo, Google and many other commercial websites.

However, in the last two year there has been a noticeable increase in spam originating from free email providers' domains.

'There were suggestions that cheap labour was behind this increase, and that CAPTCHA security was good enough, but low-paid people in developing countries were being hired to decode it manually,' explained Dr Yan, who will be presenting his findings at the ACM Computer and Communications Security Conference next week (27-31 October). 'Our research showed that computers, not people, were able to break this code much easier than previously thought.'

Dr Yan's team's methods were initially tested in 2007 on a high-profile CAPTCHA designed and widely deployed by Microsoft, with surprisingly good results. Microsoft has been using this CAPTCHA technology since 2002 for many of its online services, including Hotmail, MSN and Windows Live, and it has been fine-tuned by its designers over the years.

The latest CAPTCHA used by Yahoo, which was designed to be more hacker-proof, has also fallen foul of Dr Yan's technique. 'In our view, unfortunately all the different versions only provided a false sense of security as they were all open to our simple, low-cost segmentation attacks,' he said.

One of the hardest parts of CAPTCHA to break is separating the letters and putting them in the right order, a process known as segmentation. Warped letters confuse machines, but humans are much better at visually removing extraneous lines.

Using an ordinary desktop computer, Dr Yan and Mr El Ahmad used a seven-step method – which took less than 80 milliseconds - to remove arcs in the Microsoft scheme that link letters and make them hard to isolate, and then identify all the characters in the right order. Key to their success was an innovative colour filling method, which proved extremely powerful when combined with more traditional vertical histogram analysis.

They could isolate each of the eight characters in over 90 per cent of the challenges generated by the Microsoft scheme and, by combining this with character recognition techniques, they were able to solve them over 60 per cent of the time. The aim of CAPTCHA is to not allow bots to be more successful than 1 in 10,000 attempts (a success rate of 0.01%).

These findings were not released until the companies concerned were able to address the issues raised by Dr Yan's research.

'It is not a trivial task to design a CAPTCHA scheme that is both usable and robust,' said Dr Yan. His team's critical analysis of the security of current schemes has contributed to an immediate improvement to existing systems and will also help to create a next generation of CAPTCHAs that are both secure and useable.

Early research suggests that computers are very good at recognising single characters, even if they are highly distorted. 'Once the positions of the characters are known, breaking the scheme is purely a recognition problem, which is a trivial task with standard machine learning techniques such as neural networks,' explained Dr Yan.

The best line of defence, says Dr Yan, appears to be letting characters touch or overlap with each other, juxtaposing characters in any direction to make it harder to tell real characters and other 'noise' apart, and randomising the width of those characters.

However, by making it harder for computers to solve it also becomes more difficult for humans to decipher. 'It's a question of striking the right balance,' said Mr Yan. 'I actually think the idea of CAPTCHA is a good one, but the devil is in the detail and this is where future work needs to focus.'

Dr Yan and Mr El Ahmad are currently designing a 'tool box', which will contain a collection of algorithms and attacks to allow companies to evaluate the strength of future CAPTCHAs.

Provided by Newcastle University