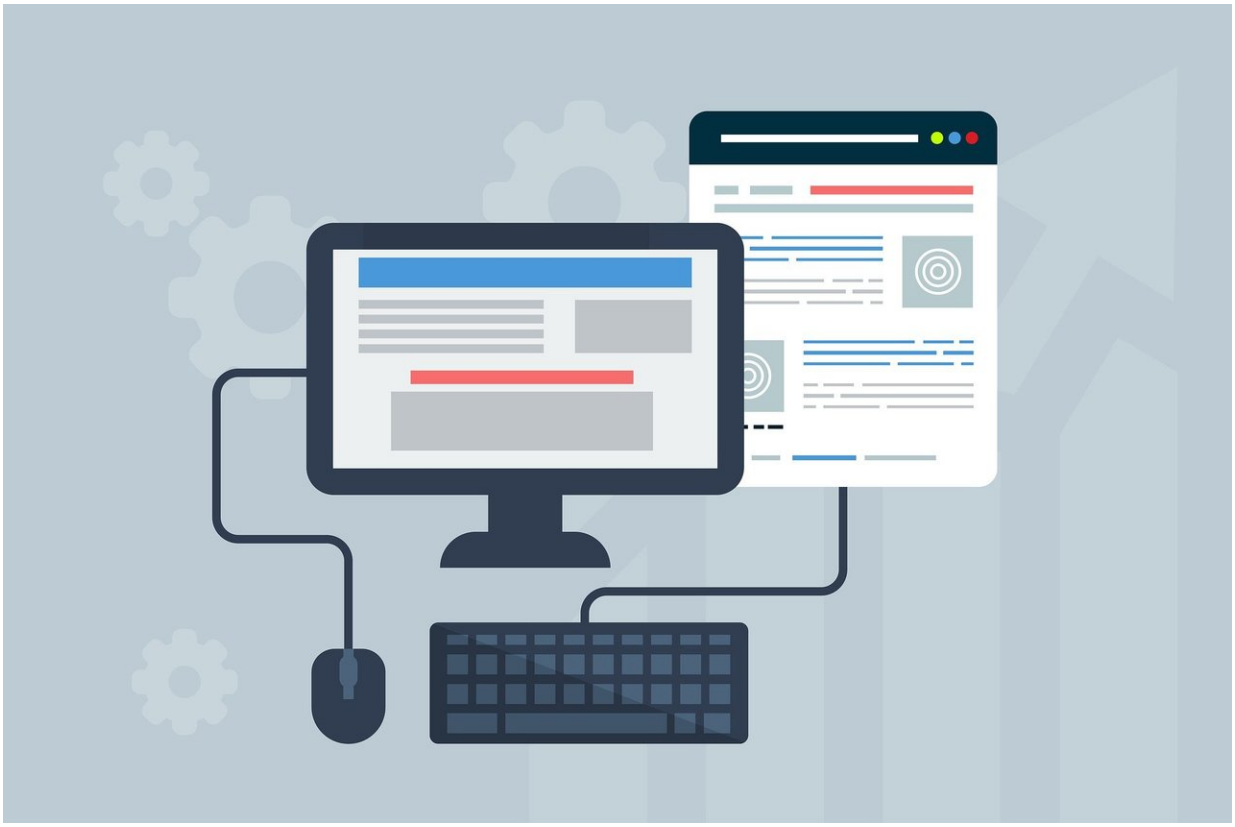
New attack could make website security captchas obsolete

December 5 2018



Credit: CC0 Public Domain

Researchers have created new artificial intelligence that could spell the end for one of the most widely used website security systems.

The new algorithm, based on deep learning methods, is the most effective solver of captcha security and authentication systems to date and is able to defeat versions of text captcha schemes used to defend the majority of the world's most popular websites.

Text-based captchas use a jumble of letters and numbers, along with other security features such as occluding lines, to distinguish between humans and malicious automated computer programmes. It relies on people finding it easier to decipher the characters than machines.

Developed by computer scientists at Lancaster University in the UK as well as Northwest University and Peking University in China, the solver delivers significantly higher accuracy than previous captcha attack systems, and is able to successfully crack versions of captcha where previous attack systems have failed.

The solver is also highly efficient. It can solve a captcha within 0.05 of a second by using a desktop PC.

It works by using a technique known as a 'Generative Adversarial Network', or GAN. This involves teaching a captcha generator programme to produce large numbers of training captchas that are indistinguishable from genuine captchas. These are then used to rapidly train a solver, which is then refined and tested against real captchas.

By using a machine-learned automatic captcha generator the researchers, or would be attackers, are able to significantly reduce the effort, and time, needed to find and manually tag captchas to train their software. It only requires 500 genuine captchas, instead of the millions that would normally be needed to effectively train an attack programme.

Previous captcha solvers are specific to one particular captcha variation. Prior machine-learning attack systems are labour intensive to build,

requiring a lot of manual tagging of captchas to train the systems. They are also easily rendered obsolete by small changes in the security features used within captchas.

Because the new solver requires little human involvement it can easily be rebuilt to target new, or modified, captcha schemes.

The programme was tested on 33 captcha schemes, of which 11 are used by many of the world's most popular websites—including eBay, Wikipedia and Microsoft.

Dr. Zheng Wang, Senior Lecturer at Lancaster University's School of Computing and Communications and co-author of the research, said: "This is the first time a GAN-based approach has been used to construct solvers. Our work shows that the security features employed by the current text-based captcha schemes are particularly vulnerable under deep learning methods.

"We show for the first time that an adversary can quickly launch an attack on a new text-based captcha scheme with very low effort. This is scary because it means that this first security defence of many websites is no longer reliable. This means captcha opens up a huge security vulnerability which can be exploited by an attack in many ways.

Mr Guixin Ye, the lead student author of the work said: "It allows an adversary to launch an attack on services, such as Denial of Service attacks or spending spam or fishing messages, to steal personal data or even forge user identities. Given the high success rate of our approach for most of the text captcha schemes, websites should be abandoning captchas."

Researchers believe websites should be considering alternative measures that use multiple layers of security, such as a user's use patterns, the

device location or even biometric information.

The research is published in the paper 'Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach' which was presented at the ACM Conference on Computer and Communications Security (CCS) 2018 in Toronto.

More information: Guixin Ye et al, Yet Another Text Captcha Solver, *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security - CCS '18* (2018). DOI: 10.1145/3243734.3243754