A deep learning-based framework to detect phishing websites

December 1 2021, by David Bradley

Credit: Pixabay/CC0 Public Domain

Most of us will have received a scam email that looks like it has come from our bank or an online store or other company or organization. They can look genuine but usually hidden within are malicious links that once clicked take you to a third-party server that either steals login details you

enter or drops malware on your device. These are phishing emails. The deliberate misspelling of "fish" with a "ph" is related etymologically to the term "phreak" which is an abbreviated portmanteau from the 1960s meaning "phone freak" and alluding to a person who hacked phone systems for pleasure or personal gain.

Some phishing emails may have poor grammar and spelling are rarely perfect or the layout may be askew and not exactly what one would expect from a legitimate organization. Such phishing attacks are relatively easy to spot, but the close-to-perfect ones may well not be and protective systems on one's device are then needed to avoid the user being duped into clicking a malicious link.

Writing in the *International Journal of Information Privacy, Security and Integrity* a team from China has developed a deep learning-based framework that might be used to detect phishing websites. Huanhuan Wang, Debin Cheng, and Hui Peng of the Fifth Electronic Research Institute of Ministry of Industry and Information Technology in Guangzhou, China, explain how their framework can extract descriptive and statistical features from a website and then determine whether these features are indicative of a phishing website. The detection of such sites could then be used in online security research and perhaps even be incorporated into browsers to protect unwary users from being phished.

The team has tested their system against two databases, one containing the website address (uniform resource locators, URLs) of 10,000 legitimate and otherwise benign sites and 13,000 URLs found in the PhishTank public dataset of sites that have previously been themselves hooked and identified as malicious. The team has demonstrated a detection accuracy of almost 99 percent, which they say is a significant improvement on earlier phish detection methods. The approach they have taken might also point to new areas of research in this area and the development and optimization of detection systems that can be

incorporated into security systems for mobile and desktop devices.

More information: Huanhuan Wang et al, Phishing website detection method based on CNAIR framework, *International Journal of Information Privacy, Security and Integrity* (2021). DOI: 10.1504/IJIPSI.2021.119167

Provided by Inderscience