SaTC: CORE: Small: Phishing Website Detection

Matthew Wright and Josephine Wolff and Richard Zanibbi

In the United States alone, hundreds of millions of dollars are lost to phishing attacks each year [29]. Thus, despite substantial research into anti-phishing mechanisms, attackers continue to find the attacks to be successful and profitable. While a variety of information could be used to detect phishing sites, such as DNS WHOIS records and IP addresses of the servers, much of it can be manipulated by an intelligent adversary. One aspect that is harder to manipulate, however, is the visual branding that the attacker uses to help convince the user that the site is the legitimate target site, e.g. an online bank like Chase.com. Without these visual cues, or with heavily modified cues, users will find the site's appearance to be suspicious, lowering the attacker's success rate.

Thus, a well-studied approach to combating phishing attacks is to look for *visual similarity* between an unknown website and popular phishing targets, like online banks and e-commerce sites. Unfortunately, prior work in using visual similarity is either too slow for real-time use or not robust to attackers manipulating parts of the page.

Intellectual Merit. We propose to design and evaluate a framework for phishing site detection that is fast enough to use in web browsing, accurate, and difficult to fool without making the user suspicious. Our approach focuses on logos, as they are they key branding elements that users expect to see on the sites they visit. To ensure that our approach is fast enough for real-time use in the browser, we propose a *multi-phase framework*, in which we first pass images from an untrusted page through a *shallow detection* phase that looks for potential matches between a newly seen logo and any of the target logos. The potential logo is then more carefully checked against the matching target logo, if any, in a process that requires more effort but has only one point of comparison.

To evaluate our proposed framework, we will perform the following tasks:

- *Design algorithms for fast detection.* We have explored a promising approach for shallow detection using Histograms of Gradients (HoGs), and we will extend this approach to ensure robustness to attacker manipulation, low error rates, and speed.
- *Create and evaluate a complete detection tool.* We will evaluate algorithms for both object detection and deep detection, examine adding intermediate phases between shallow and deep detection, and build a prototype browser plug-in for full evaluation of the approach.
- *Design and evaluate for usability.* We will examine whether users respond with appropriate suspicion to pages with logos that have been modified in various ways. We will also evaluate the design of warnings and their effect on the browsing experience.
- *Interaction with Training.* We will perform the first long-term study to evaluate the impact of phishing training on the incidence of account compromises. We will also evaluate how we can use signals from phishing training in warnings produced by our tool for greater effectiveness.

Broader Impacts. With novel approaches to both improved technology and training, we can make major strides in reducing the impact of phishing attacks. Our phishing detection tool will be made available both as an open source project and as a browser plug-in. We will disseminate our findings in top conferences and journals and also in pieces for the popular press intended to convey best practices and new tools for phishing prevention to the larger public. Building on our ideas on combining training with technology, we will develop a new interdisciplinary graduate course on Users and Security. We will also leverage our work to design activities for our GenCyber summer camps for K-12 students.