
Detection of Phishing Attacks: A Machine Learning Approach

Ram Basnet, Srinivas Mukkamala, and Andrew H. Sung

New Mexico Tech, New Mexico 87801, USA
{ram,srinivas,sung}@cs.nmt.edu

1 Introduction

Phishing is a form of identity theft that occurs when a malicious Web site impersonates a legitimate one in order to acquire sensitive information such as passwords, account details, or credit card numbers. Though there are several anti-phishing software and techniques for detecting potential phishing attempts in emails and detecting phishing contents on websites, phishers come up with new and hybrid techniques to circumvent the available software and techniques.

Phishing is a deception technique that utilizes a combination of social engineering and technology to gather sensitive and personal information, such as passwords and credit card details by masquerading as a trustworthy person or business in an electronic communication. Phishing makes use of spoofed emails that are made to look authentic and purported to be coming from legitimate sources like financial institutions, ecommerce sites etc., to lure users to visit fraudulent websites through links provided in the phishing email. The fraudulent websites are designed to mimic the look of a real company webpage.

The phishing attacker's trick users by employing different social engineering tactics such as threatening to suspend user accounts if they do not complete the account update process, provide other information to validate their accounts or some other reasons to get the users to visit their spoofed web pages.

Why is it important to tackle the problem of phishing? According to the Anti-Phishing Working Group, there were 18,480 unique phishing attacks and 9666 unique phishing sites reported in March 2006. Phishing attacks affect millions of internet users and are a huge cost burden for businesses and victims of phishing (Phishing 2006). Gartner research conducted in April 2004 found that information given to spoofed websites resulted in direct losses for U.S. banks and credit card issuers to the amount of \$1.2 billion (Litan 2004). Phishing has become a significant threat to users and businesses alike.

Over the past few years, much attention has been paid to the issue of security and privacy. Existing literature dealing with the problem of phishing is scarce. Fette et al proposed a new method for detecting phishing emails by incorporating features specific to phishing (Fette et al. 2006).

We applied different methods for detecting phishing emails using known as well as new features. We employ a few novel input features that can assist in discovering phishing attacks with very limited a-prior knowledge about the adversary or the method used to launch a phishing attack. Our approach is to classify phishing emails by incorporating key structural features in phishing emails and employing different