Journal of Theoretical and Applied Information Technology
15th September 2016. Vol.91. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FEATURE-BASED PHISHING DETECTION TECHNIQUE

[1]XIN MEI CHOO, [2]KANG LENG CHIEW, [3]DAYANG HANANI ABANG IBRAHIM, [4]NADIANATRA MUSA, [5]SAN NAH SZE, [6]WEI KING TIONG
Faculty of Computer Science and Information Technology,
University Malaysia Sarawak, 94300 Kota Samarahan, MALAYSIA
E-mail: [1]xmchoo@gmail.com, [2]klchiew@unimas.my, [3]hananii@unimas.my, [4]nadia@unimas.my, [5]snsze@unimas.my, [6]wktiong@unimas.my

ABSTRACT

Phishing is an Internet fraud to entice unsuspecting victims. The tactic of phishing is to impersonate the trusted entities by employing both social engineering and technical subterfuge. Moreover, phishing is a form of online identity theft that creates a fake copy of popular site. There are many types of anti-phishing techniques available. However, they are mostly still in the infancy stage which may give false alarm to the user. Therefore, this research aims to develop a feature-based phishing detection technique to overcome the limitation. The proposed method involves aggregating new features with several existing features to form a sensitive features set. Based on the features set, the proposed method will utilise support vector machine to perform the classification. The experimental results show convincing performance with 95.33 percent of accuracy.

Keywords: *URL Features, Website Features, Phishing Detection, Anti-Phishing, Feature Extraction, Classification*

1. INTRODUCTION

Phishing attacks usually target user confidential information such as username, password and financial ID. Phishers would use their sophisticated attack vector such as emailing, or pop up window notification to lure the victim to visit the phishing website which has legitimate-looking layout. This will allow the phishers to harvest the victim credentials and sell them in the black market (as depicted in Figure 1).

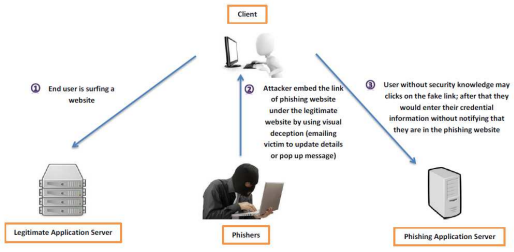

Figure 1: Example of Phishing Attack Scenario

Anti-phishing refers to the method that is employed to prevent and defend phishing attacks. There are many techniques that offer protection at different domain. Some techniques work on emails, while others work on website attributes [2]. The proposed method works on the later. The contribution of this paper is twofold: first, it identifies and analyses attributes exhibited in phishing websites. Second, it proposes several new features and integrates to an existing method to enhance the overall detection performance. The works start by analysing and looking for abnormal attributes of a phishing website. The abnormal attributes that usually appear on phishing website include some uncommon symbols in the URL and some irregular HTML form and title elements. Therefore, extracting features from these attributes will enhance the phishing detection ability.

2. LITERATURE REVIEW

Many anti-phishing techniques focus on enabling the client to recognize and filter various types of phishing attacks [1, 4, 9]. In general, anti-phishing techniques can be classified into four categories, which are content filtering, blacklisting, symptom-based prevention and domain binding. Content filtering uses machine learning techniques, such as Bayesian Additive Regression Trees (BART) or Support Vector Machines (SVM) to filter phishing email that matches phishing attributes. Blacklist is a collection of known phishing websites that are published by a trust entity like Google and Microsoft. However, blacklist requires constant