Phishing Email Detection Technique by using Hybrid Features

Lew May Form*, Kang Leng Chiew†, San Nah Sze‡ and Wei King Tiong§
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia
*mayform19kt@gmail.com
†klchiew@fit.unimas.my
‡snsze@fit.unimas.my
§wktiong@fit.unimas.my

***Abstract*—Phishing emails is growing at an alarming rate in this few years. It has caused tremendous financial losses to internet users. Phishing techniques getting more advance everyday and this has created great challenge to the existing anti-phishing techniques. Hence, in this paper, we proposed to detect phishing emails through hybrids features. The hybrid features consist of content-based, URL-based, and behavior-based features. Based on a set of 500 phishing emails and 500 legitimate emails, the proposed method achieved overall accuracy of 97.25% and error rate of 2.75%. This promising result verify the effectiveness of the proposed hybrid features in detecting phishing email.**

***Keywords*-Anti-phishing; emails; behavior-based; classification;**

I. Introduction

Email has made the communication process become easier, faster and cheaper. It has became increasingly popular. However, phishing email is one of the major security threats. The phishing email can lead to financial loss. Attacker always sending email tends to make user believe that they are comminicating with trusted entity and deceive them into providing personal credentials in order to access service, such as credit card numbers, account login credential or identity information. Phishing email causes a serious threat to information security and internet privacy. Forrester Research that 20 percent of consumers refuse to open email or attachment even the email look legitimate, due to their loss of trust [9].

According to APWG phishing attack trends reports [2], the number of phishing email increase from 28,897 unique phishing reports in December 2009 to 45,628 unique phishing reports in December 2012. Phishing activity is growing at an alarming rate. Gartner survey December 2007 [7] estimated that phishing attacks $3.2 billion losses in business and 3.3 percent of consumers claim that they lost money because of email-based phishing attacks.

Many machine-learning techniques have been proposed in the literature to detect and filter phishing email. From an overview in previous work [1] shown that existing machine-learning techniques have limitation on consumes memory, consumes time, weak in detecting zero-day attack and feed continuously . Despite the constant improvement in anti-phishing techniques, phishers are more advanced to challenge existing anti–phishing techniques by introducing sophisticated techniques in order to bypass the detection. High changing rate of phishing attack techniques increase the difficulty of detecting and filtering phishing email attacks. Besides, the phishers are able to convince and deceive users by creating replica of a website that is identical to the original legitimate websites. User who is unfamiliar with browser security indicators can become a victim in these attacks.

In this paper, we enhance the method proposed in [8] by combining the URL-based, content-based and behavior-based features. In order to extract the features, we analyse the email header information such as message-ID, sender email and return path. Analysing these information is rational as these information can not be easily camouflaged by phishers. Furthermore, email header is globally unique identification which is not visible to most users, but it is a useful indicator in determing where the message really originated from though routing information. The remainder of the paper is structured as follow. In the next section, we will discuss some of the related works. In Section 3 further discuss the proposed method. In Section 4 presents the experimental results and gives the analysis. In Section 5 conclude the paper.

II. Related Work

Ease of communicating with advent of email has caused the problem of unsolicited bulk email, especially issues of phishing attacks through emails. Various anti-phishing techniques have been developed to solve the phishing attacks problem. Many email clients (eg. Thunderbird) have some sort of built-in junk filter, but it is insufficient to overcome the phishing email problem, due to the ever changing phishing attacks techniques.