Abstract

There has been an abrupt development and use of online transactions over the past decade. The increased sophistication of cyber criminals has lead to proliferation of phishing attacks. The continuous expansion of World Wide Web has led to the rapid spread of phishing, malware and spamming.

This paper proposes a feature based approach to classify URLs into phishing or non-phishing category. The usage of a variety of URL features is done by studying the anatomy of URLs. For classification of URLs, two different algorithms have been used. Random Forest machine learning algorithm is used to build an efficient classifier which would decide whether a given URL is phishing or not. In addition, a novel scheme has been proposed to detect phishing URLs by mining the publicly available content on the URLs.

Index Terms

Computer Science
Information Sciences
Keywords

Phishing URL, web mining, benign, machine learning