

Mini social networking website

Abstract

With an average of 80% length reduction, the URL shorteners have become the norm for sharing URLs on Twitter, mainly due to the 140-character limit per message. Unfortunately, spammers have also adopted the URL shorteners to camouflage and improve the user click-through of their spam URLs. In this paper, we measure the misuse of the short URLs and analyze the characteristics of the spam and non-spam short URLs. We utilize these measurements to enable the detection of spam short URLs. To achieve this, we collected short URLs from Twitter and retrieved their click traffic data from Bitly, a popular URL shortening system. We first investigate the creators of over 600,000 Bitly short URLs to characterize short URL spammers. We then analyze the click traffic generated from various countries and referrers, and determine the top click sources for spam and non-spam short URLs. Our results show that the majority of the clicks are from direct sources and that the spammers utilize popular websites to attract more attention by cross-posting the links. We then use the click traffic data to classify the short URLs into spam vs. non-spam and compare the performance of the selected classifiers on the dataset. We determine that the Random Tree algorithm achieves the best performance with an accuracy of 90.81% and an F-measure value of 0.913.