Journal of Network and Computer Applications 34 (2011) 1717–1723

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Social network analysis of web links to eliminate false positives in collaborative anti-spam systems

Zac Sadan, David G. Schwartz [*]

Graduate School of Business Administration, Bar-Ilan University, Ramat Gan, Israel

ARTICLE INFO

Article history:
Received 26 October 2010
Received in revised form
17 May 2011
Accepted 12 June 2011
Available online 23 June 2011

Keywords:
Anti-spam
Whitelister
Collaborative anti-spam
URL-based filters
Betweenness
Reputation

ABSTRACT

The performance of today's email anti-spam systems is primarily measured by the percentage of false positives (non-spam messages detected as spam) rather than by the percentage of false negatives (real spam messages left unblocked). One reliable anti-spam technique is the Universal Resource Locator (URL)-based filter, which is utilized by most collaborative signature-based filters. URL-based filters examine URL frequency in incoming email and block bulk email when a predetermined threshold is passed. However, this can cause erroneous blocking of mass distribution of legitimate emails. Therefore, URL-based methods are limited in sufficient prevention of false positives, and finding solutions to eliminate this problem is critical for anti-spam systems. We present a complementary technique for URL-based filters, which uses the betweenness of web-page hostnames to prevent the erroneous blocking of legitimate hosts. The technique described was tested on a corpus of 10,000 random domains selected from the URIBL white and black list databases. We generated the appropriate linked network for each domain and calculated its centrality betweenness. We found that betweenness centrality of whitelist domains is significantly higher than that of blacklist domains. Results clearly show that the betweenness centrality metric can be a powerful and effective complementary tool for URL-based anti-spam systems. It can achieve a high level of accuracy in determining legitimate hostnames and thus significantly reduce false positives in these systems.

1. Introduction

The proportion of email that is spam has significantly increased in recent years (Boykin and Roychowdhury, 2005; Goodman et al., 2007). Anti-spam systems filter incoming email either at the level of the email server or at the level of the email client program (Georgiou et al., 2008). These filtering techniques are generally judged according to the dual metrics of false positives and false negatives. A false positive indicates that a legitimate email message has been falsely identified as spam, potentially causing the intended recipient to miss the message. A false negative indicates that a spam message has gone undetected and passed through to the recipient. End user tolerance for false negatives is generally higher than for false positives, because users would rather accept a few spam messages that pass through their filter, than contemplate missing a single legitimate message (Yih et al., 2006). Misclassification of an email can even be prohibitively expensive in the real world. Therefore, the elimination of false positives is paramount to the success of an anti-spam system. However, although much work has been done to improve specific algorithms for detecting unwanted messages, less work has reported on leveraging multiple algorithms and correlating models for prevention of false positives (Hershkop and Stolfo, 2005). Many of today's anti-spam systems have reached a high level of accuracy in identifying actual spam (close to 98% (Snyder, 2009)), thus the most important attribute that differentiates between the various systems is their percentage of false-positives.

In this paper we begin by describing the collaborative spam filtering technique and its URL-based sub-set. We then present some of the recent URL-based anti-spam research. This is followed by an explanation of social network analysis in general, and betweenness centrality in particular. We subsequently depict recent work, which exploited social network metrics for the battle against spam. Finally, we present our technique for determining the degree of email spamminess, based on URL betweenness centrality.

2. Background and related work

2.1. Collaborative spam filtering technique

One of the major types of anti-spam systems are collaborative signature-based filters. These filters generate a unique signature for each known spam message by counting the frequency of

[*] Corresponding author. Tel.: +972 54 4890060.
E-mail address: David.Schwartz@biu.ac.il (D.G. Schwartz).

1084-8045/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2011.06.004