Knowledge-Based Systems 000 (2017) 1–15

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Early detection of spamming accounts in large-Scale service provider networks

Yehonatan Cohen, Daniel Gordon, Danny Hendler*

Department of Computer Science, Ben-Gurion University of The Negev, Be'er Sheva 84105, Israel

ARTICLE INFO

Article history:
Received 5 October 2016
Revised 26 November 2017
Accepted 29 November 2017
Available online xxx

Keywords:
Email spam detection
Outgoing spam
Low-volume spamming
Email service provider

ABSTRACT

We present *ErDOS* — an algorithm for the Early Detection Of Spamming accounts. The detection approach implemented by ErDOS combines content-based labelling and features based on inter-account communication patterns. We define new account features, based on the ratio between the numbers of sent and received emails, the distribution of emails received from different accounts, and the topological features of the network induced by inter-account communication. We also present *ErDOS-LVS* — a variant of *ErDOS* that targets the detection of low-volume spammers. The empirical evaluation of both detectors is based on a real-life data set collected by an email service provider, much larger than data sets previously used for outgoing spam detection research. It establishes that both are able to provide effective early detection of the spammers population, that is, they identify these accounts as spammers before they are detected as such by a content-based detector. Moreover, both detectors require only a single day of training data for providing a high-quality list of suspect accounts.

© 2017 Published by Elsevier B.V.

1. Introduction

With a daily rate of messages sent and received exceeding 200 billion in 2016, email is probably still the most popular Internet-based application [1]. It is anticipated that over one third of the worldwide population will be using email by the end of 2019 [1]. Due to its widespread use, email has become a fertile ground for cyber attacks such as phishing, spreading of viruses and the distribution of spam mail, consisting of unsolicited messages mostly of advertisement contents. Recent statistics show that, on average, approximately 97.4 billion spam emails were sent every day during the 1st quarter of 2013 [2]. This is almost two thirds of all email traffic [3].

Email Service Providers (ESPs) suffer from spam email and must therefore combat it. First, vast amounts of spam email that are being sent from ESP domains or being sent to these domains overload ESP servers and communication infrastructure [4,5]. In addition, ESPs from which large numbers of spam messages are sent are likely to become blacklisted, thereby preventing the legitimate users of these ESPs from exchanging email and disconnecting them from external domains. Indeed, ESPs that fail to deploy effective spam filtering mechanisms provide poor user experience and thus hurt their popularity and reputation [6–11].

A number of techniques for detecting spamming accounts and filtering out spam mail have been developed. *Content-based filters* are programs that learn and identify textual patterns of spam messages. They are widely used by ESPs. Many content-based spam filter proposals are based on machine-learning classification algorithms. Examples include filters based on Artificial Neural Networks [12–14], support vector machines (SVM) [15–19] and Bayesian classifiers [20–22].

Zhang et al. [23] proposed a spam detector that employs the binary particle swarm optimization (PSO) algorithm for feature selection [24]. The detector focuses on reducing the false positive rate by using a cost matrix that assigns higher weight to false positive errors as compared with false negative errors. Idris et al. [17] employ a combination of the PSO and the negative selection algorithm (NSA) [25] for detecting spam email. Content-based spam filtering was also studied for other types of spam, such as SMS spam [26] and social media spam [27].

Spammers have developed their own techniques to outsmart content-based filters, such as sending image-spam mails [28–31], poisoning the filter by inserting random strings into spam messages [32–35], and more.

Whereas content-based filters consider the properties of individual messages, a different approach examines the social interactions of email accounts, reflected by inter-account communication patterns, since, in many cases, the social interactions of spammers and legitimate accounts are significantly different.

* Corresponding author.
E-mail addresses: yehonatc@cs.bgu.ac.il (Y. Cohen), gordonda@cs.bgu.ac.il (D. Gordon), hendlerd@cs.bgu.ac.il (D. Hendler).

https://doi.org/10.1016/j.knosys.2017.11.040
0950-7051/© 2017 Published by Elsevier B.V.

Table 1
Our data set vs. data sets used by previous studies.

	Our data set		SUNET	NTU	Enron
#Mails	9.86E7	2.13E8	2.40E7	2.86E6	5.17E5
#Edges	7.40E7	12.90E7	2.16E7	-	3.68E5
#Accounts	5.63E7	5.81E7	1.05E7	6.37E5	3.67E4
Time period	4 days	26 days	14 days	10 days	3.5 years
Contents	Spam & ham		Spam & ham	Spam & ham	Ham

Social interactions can be modeled using a communication graph in which a vertex is introduced for each email account appearing in the data set and an edge connecting two nodes is introduced if there was email exchange between the accounts represented by the two nodes [36]. These communication graphs can be either directed or undirected. Edges may be weighted, e.g. by using a weight function that assigns to each edge the number of emails communicated between the two accounts represented by its endpoints.

After modeling social interactions by a communication graph, network-level features that distinguish between legitimate accounts and spamming accounts can be extracted. These features are then used to construct a feature-vector representing the account.

Our detector is machine-learning based and uses features based on communication patterns. We use real spamming accounts for training, identified as such by a content-based spam filter. The data set we use in this study was collected by a large, well-known, ESP, and encompasses 58 million email accounts and almost a quarter-billion email transactions. This information is stored in log files that were collected by the ESP in the course of a period of 26 days. Each of the messages in the dataset was classified by the ESP as spam/ham using Cyren's eXpurgate content-based filter [37].

Our goal is to achieve *early detection*, that is, to detect spammers before they are detected by a content-based filter and possibly even if they are not detected by the filter at all.

Our focus in this work is on the detection of spamming accounts that are hosted by large ESPs. ESPs are in an excellent position to determine which of their hosted accounts send spam, since they are able to log every email that is sent or received by these accounts. Large ESPs, such as the one whose dataset we use in this work, host a large number of email accounts (see Table 1) and so have visibility into the communication patterns between these accounts. It is much more difficult for the ESP to identify sources of spam that are outside of its network perimeter. A key reason for this is that the ESP typically receives only a (typically small) fraction of the messages sent by external spamming accounts. Moreover, external accounts may spoof their "From" address. Internal accounts, on the other hand, are associated with a static ESP identifier, hence the origin of messages sent by hosted accounts to other hosted accounts cannot be spoofed.

Pathak et al. [38] distinguish between mail accounts that send large amounts of spam, which we henceforth refer to as *High Volume Spammer* (HVS) accounts, and accounts that attempt to remain "under the radar" by sending spam messages at a relatively slow rate, which we henceforth refer to as *Low Volume Spammer* (LVS) accounts. LVS accounts are often operated by bot machines which are a part of a botnet [38–40].

LVS-spamming botnets manage to relay significant quantities of spam messages by generating and/or hijacking numerous email accounts and sending a small number of spam emails from each of these accounts. From the spammer's perspective, LVS spamming is advantageous since LVS accounts are less conspicuous and harder to identify than HVS accounts. Consequently, providing effective early detection of LVS spammers is more challenging.

1.1. Our contributions

This study is based on a large real-life data set, consisting of both outgoing and incoming mail logs involving tens of millions of email accounts hosted by a large, well-known, ESP. It was made available to us after having undergone privacy-preserving anonymization pre-processing. This data set is much larger than data sets used by previous outgoing-spam detection research.

Using this data set, we evaluated previously published outgoing-spam detection algorithms. Our experimental evaluation finds a large drop in their accuracy on this data set as compared with the results on the data sets used in their evaluation, indicating that algorithms optimized for small and/or synthetic data sets are not necessarily suitable for real-life mail traffic originating from large ESPs. New approaches are therefore needed in order to efficiently detect outgoing spam in large ESP environments.

Our emphasis in this work is on *early detection of spamming accounts hosted by ESPs*. We present two detectors: *ErDOS*, an algorithm for the Early Detection Of Spamming accounts; and *ErDOS-LVS*, a variant focusing on the early detection of Low-Volume Spammers. The detection approach implemented by both ErDOS and ErDOS-LVS combines content-based labelling and features based on inter-account communication patterns.

ErDOS uses email-account features that are based on the ratio between the numbers of sent and received emails and on the distribution of emails received from different accounts. ErDOS-LVS also uses features computed based on the 2-hop neighborhoods of accounts. By using the output of a content-based spam detector as a means for obtaining initial labeling of email accounts, we manage to avoid the use of synthetically-generated spam accounts as done by some prior work.

ErDOS uses the account labels induced by the output of the content-based detector for supervised learning of a detection model based on the features we define. Empirical evaluation of ErDOS shows that it provides significantly higher accuracy as compared with previous outgoing-spam detectors. Moreover, by using only a single day of training data, it is able to provide high quality early detection of spamming accounts.

Careful examination of the detection results obtained by ErDOS reveals that, although it provides good detection results for medium to high-volume spammers, its detection accuracy on LVS accounts is lower. We therefore implemented and evaluated ErDOS-LVS, a variant of ErDOS that is optimized for the detection of LVS accounts. The evaluation we conducted establishes that ErDOS-LVS succeeds in providing high detection accuracy and early detection of LVS spamming accounts.

2. Related work

The approach of social network-based spam detection was introduced by Boykin and Roychowdhury [41]. In their work, they constructed a network based on the information available in a user's inbox. The spam-detection method they proposed is graph-based and aims to identify *incoming* spam messages. In this section, we survey previous work dealing with social-network based outgoing spam detection and the detection of low-volume spammers.

2.1. Social-network based outgoing spam detection

The social behavior of spamming email accounts was studied and identified by several studies (see, e.g., [9,27,42–45]). Based on the findings of these studies, a few outgoing spam detection schemes were proposed.

Lam and Yeung [46] present a machine-learning based outgoing spammers detector that uses inter-account communication pat-

Please cite this article as: Y. Cohen et al., Early detection of spamming accounts in large-Scale service provider networks, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.11.040

Download English Version:

https://daneshyari.com/en/article/6861886

Download Persian Version:

https://daneshyari.com/article/6861886

Daneshyari.com