Information Sciences 289 (2014) 133–147

Contents lists available at ScienceDirect

Information Sciences

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

# Feature set identification for detecting suspicious URLs using Bayesian classification in social networks

Chia-Mei Chen *, D.J. Guan, Qun-Kai Su

*National Sun Yat-sen University, Kaohsiung, Taiwan, ROC*

ARTICLE INFO

*Article history:*
Received 30 August 2013
Received in revised form 12 June 2014
Accepted 24 July 2014
Available online 9 August 2014

*Keywords:*
Social network
Anomaly detection
Bayesian classification

ABSTRACT

Social network services (SNSs) are increasing popular. Communicating with friends forms a social network that can be used to promptly share information with friends. In targeted attacks, SNSs are often used to collect personal information and craft attacks based on a specific user profile. Malware can be used to facilitate social relationship, sends messages containing malicious URLs, lures users to click on these URLs by employing social engineering techniques; then replicates through the social network over and over again. Because users are curious and trust in their friends, they typically click on malicious URLs without verification. In this study, a feature set is presented that combines the features of traditional heuristics and social networking. Furthermore, a suspicious URL identification system for use in social network environments is proposed based on Bayesian classification. The experimental results indicate that the proposed approach achieves a high detection rate.

## 1. Introduction

Based on advances in information technology, websites offer various convenient web services such as information retrieval, chat rooms, Web 2.0-based services, blogs, albums, and multimedia sharing. Social network services (SNSs), such as Facebook, Twitter, and MySpace, have recently proliferated offering interactive information platforms that allow users to share and to interact. Based on Facebook statistics [6], Facebook owns 845 active members monthly. A TrendMicro report [19] indicated that the percentage of employees using SNS increased from 19% in 2008 to 24% in 2010. In addition, based on a survey Pew Research [17], the trend toward using SNSs has annually increased among all age groups; thus, SNSs are increasingly popular and essential.

Incident investigation reports have indicated that cybercrimes, such as targeted attacks or advanced persistent threats (APTs) often use SNSs to collect personal information and launch social engineering attacks [25,29,12]. In other words, the convenience of SNSs facilitates potential cyber attacks on SNS platforms. For example, a social-network-based worm spreads by attempting to steal account information and infect additional users by using a social engineering trick that sends malicious URL posts or emails. Because SNS users typically trust their friends, they are often breached by these worms, which rapidly spread through the friendship connections of victims.

Malware programs often leverage short URL and blog services are often used by malware to disguise original URLs and evade security inspections, such as blacklist filtering applications. URL shortening transfers original URLs through shortened URLs by using redirection. Because URL shortening is often abused, these providers may find themselves blacklisted.

* Corresponding author.

http://dx.doi.org/10.1016/j.ins.2014.07.030

SNSs comprise wide range of interactive and sharing services is offered by SNS. Walls in Facebook which combine bulletin and message boards, enable board members post messages. In instant message environments, only those in a contact list can send messages, whereas in SNSs, a cluster of friends formed through friendship links or a cluster of members who share common interests are not limited to contact lists, and can post messages that include malicious URLs.

Walls are used on personal, club, and fan pages; access to personal and club pages is protected using access controls, whereas fan pages cannot. All users can access fan pages without requesting permission from the page owner. A Websense survey [30] indicated that 10% of URLs posted to in Facebook were malicious links; thus, users who access popular fan pages may risk security breaches. Hackers can spread attacks by simply posting messages containing malicious links on the most popular fan walls; multiple fans are likely to click on such links.

Compared with spam, posts containing malicious URLs are faster and more effective. As long as the content of the post addresses hot topics, it can catch the attention of numerous users. In this study, a set of heuristic features and Bayesian classification are proposed for detecting malicious URLs in SNSs. According to the findings, malicious web links in a post exhibit domain and social anomalies that differ from those of typical links. The proposed detection method involves using a naive Bayesian model to detect social network posts that contain malicious URLs based on anomalies in the URL domain and unusual posting behaviors.

## 2. Related work

This section introduces spam, phishing, and methods of detecting suspicious URLs in online messages and social networks. Classification algorithms are introduced along with their related network security applications in detecting anomalies.

Zhang et al. [32] proposed a content-based method for detecting phishing web sites, suggesting that phishing sites are created based on minor modifications from the authenticated sites and exhibit low page ranks in the Google search results. A set of heuristics was proposed based on domain name, lexical signatures of web links, and the HTML content of web pages. Five keywords were extracted from each web page based on TF-IDF (term frequency/inverse document frequency) algorithm and the Google search was applied to verify the website authenticity.

According to McGrath et al. [18], a brand name should appear in the URL of a web site. They collected and analyzed the URLs of phishing and non-phishing websites, determining that diverse countries host phishing sites, phishing domains are rarely hosted in their registered country, and phishing domains last approximately 3 days.

Fette et al. [8] extracted email features such as HTML tags, the number of links, use of javascript, and number of domains, to distinguish phishing emails by using a support vector machine (SVM). Bergholz et al. [3] proposed using email features namely, the structure of the email body, web link properties, and a keyword list, which were generated using dynamic Markov chain training and class-topic models. Their results indicated that using these features improved the detection rate. Abu-Nimeh et al. [1] chose the 43 most popular keywords as features and evaluated the performance level by using various machine learning classification algorithms.

Ma et al. [14] adopted semantic features from McGrath [18] and bag-of-words features from Kolari et al. [13]. In addition, Ma et al. addressed features specific to the hosted machine such as the IP address, WHOIS information, domain name, and geographic location. Machine learning algorithms namely, the native Bayesian theorem, Support Vector Machine (SVM), and logistic regression, were applied to evaluate the detection of suspicious URLs when using various combinations of features and data sets. Their subsequent studies [15,16] have yielded the similar conclusions, indicating that the features of URL semantics and host information are essential for identifying malicious web links when a suitable machine learning technique is applied.

Online messaging is a real-time communication service. Morse et al. [21] noted that online messaging networks are scale-free, and the distribution of network node connectivity follows power law. In other words, highly connective nodes are likely to be connected with other nodes. Therefore, worms that infect highly connected nodes can rapidly propagate throughout a network, making it difficult to completely remove such worms. Thus, identifying methods of effectively detecting or preventing online messaging worms is critical.

Guan et al. [10] proposed a suspicious URL detection method used in instant message environments, noting the discrepancies between human and robot interactions and proposing features based on anomalous user behaviors, such as response time entropy, delay time entropy, and the domains of attached links (e.g. domain rank, lexical features, and WHOIS information). The experimental results demonstrated that the features were effective and the scoring model attained a sufficient performance level. Guan el at. demonstrated that distinct interaction timing features are distinguishable in online messages. Although the proposed features might not be applicable for detecting malicious URLs in social networks, the unique social communication patterns of a user might be essential for detecting malicious activities.

Social networks exhibit scale-free and small-world network characteristics [20]. Small-world networks consist of multiple clusters that exhibit small average shortest path lengths and high clustering coefficients. Therefore, the nodes within a cluster are densely connected. A cluster in a social network may be a friend group or a club of members sharing common interest. In these networks worm infection may propagate through friend groups and clubs, spreading more rapidly compared with infections in instant message environments.