International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 4 (2015) pp. 8499-8503

http://www.ripublication.com

URL ATTACKS: Classification Via Analysis And Learning

M.Rajesh[1] and A.Safiya Parvin [2]

Computer science Department, Sathyabama University Chennai, India.
[1] rajesh.manoharan89@gmail.com
[2] aspkareem@gmail.com

Abstract

Social Networks such as Twitter, Facebook plays a remarkable growth in recent years. The ratio of tweets or messages in the form of URLs increases day by day. As the number of URL increases, the probability of fabrication also gets increased using their HTML content as well as by the usage of tiny URLs. So it is important to classify the URLs by means of some modern techniques. Conditional redirection method is used here by which the URLs get classified and also the target page that the user needs is achieved. Learning methods also introduced to differentiate the URLs and there by the fabrication is not possible. Also the classifiers will efficiently detect the suspicious URLs using link analysis algorithm.

Keyword-: URL, Tiny URL, Link Analysis

I. INTRODUCTION

SOCIAL NETWORKING plays an important role in information sharing service for transferring of messages in the form of tweets or any other modes .When the social users need to share a URL with their close once then they formally use some of the shortening services

The proliferation of social networking [6] lead to increase in spam activity. The spammers send unsolicited messages for various purposes. Hash tags [4] and shortened URLs [7] [9] like t.co are frequently abused by the spammers. Hash tags are used to denote the topic or latest trend and they are abused by the spammers. The ability to disguise URL destination has made twitter or other social networks as an attractive target for the spammers.

In the first study focusing on spam detection [5], we collect a number of users account. The users are considered as spammers by use of special methods and algorithms and to determine the false positive rate. Here we collect a specific number

of users account such as in small environment like colleges or small scale industries to detect their spamming .This will act as the stand alone application for spam detection.

II. RELATED WORK

Zachary Miller et al [9] discussed about data stream clustering by introducing two clustering algorithms such as StreamKM++ and DenStream to facilitate spam identification. Here clustering is considered for spam detection also K-Means algorithm is used to cluster the related number of users who use specific URLs.

Sangho Lee et al [4] discussed a suspicious URL detection system called warning bird .Here the authors had a detailed study about correlated URL redirect chains using frequently shared URLs .Friends follower ratio is used to find the similarity.SVC algorithm is used here to determine the low false positive rate.

Dhanalakshmi Renganayakulu et al [1] concluded that Bayes Classifier will be efficient enough for spam filters. Bayes theorem is used to calculate the probability of hypothesis by comparing the event made by the user and the training data.Here the cumulative scores and threshold value are compared to differentiate phishing and legitimate URLs.

Nazpar Yazdanfar et al [5] hash tags in measuring the relevancy of URLs. He used similarity measures such as Euclidian, Cosine, Jaccard and dice coefficient. He compared this similarity with matrix factorization method. Also concluded that aggregation of hash tags and user similarities will improve the accuracy.

Jelena Isacenkova et al and Oliver Thonnard et al [2] used multidimensional clustering technique for grouping similar emails. The role of phone numbers are considered as an important identifiers where they forward the victims to specific URLs.

G.Stringhini and C.Kruegel et al [6] analyzed to which extent the spamming has happened in social networking. Also analyzed how they operate while they enter into the sharing environment. They created a large set of honey profiles to analyze the set of data received. Based on the analysis the spammers are identified.

F.Klien et al and M.Strohmaier et al [3] had a overall study about the URL shortening services by which the users redirected to malicious sites. They discussed about a real time websites like tinyurl.com.

J,Song et al and S.Lee et al [7] came across two problems. One is that the accounting features that can be abused by the spammers and second is that the accounting features. They discussed about a particular account that sends or involved in spamming activity can be detected.

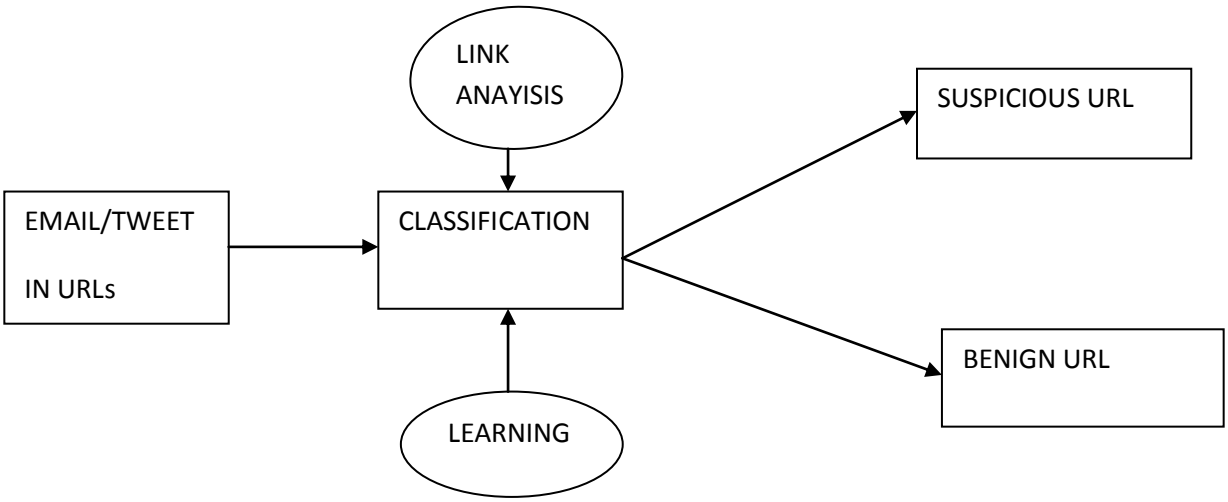

Figure: URL Redirection scheme

III. IDEAS AND FEATURES

4.1 Data sets

Dataset for this study include number of users accounts with an e-mail or a message from each account in the form of URLs. Since larger data sets may give inaccurate results, here we have chosen smaller data sets to facilitate more and thorough analysis for an stand alone application operated in offline mode. To have more conclusive results, we elected to use smaller data sets that could be manually created. Content based features were used in the previous studies such as hash tags that uses "@" and "#" respectively.

4.2 Machine Learning

In our study, we take supervised learning approach with trained datasets as mentioned above and this predictive technique will give a reasonable response for classification that is it can able to differentiate the URLs based on the environment we work upon. Suppose we take a case that number of people will have heart attack within a year. These can be predicted by taking a data sample that comprise of age, height, weight, blood pressure. The problem will combine all the existing data into a model hat can predict a new person will have heart attack or not.

This supervised leaning splits into two broad categories such that classification for responses and regression for responses. Classification for responses will have known values such as true or false. Regression for responses that will denote a real number based on which the values can be predicted.

4.3 Link analysis:

By extending the concept of page ranking we propose power iteration in our study in which each page is considered as 'i' and the count increases gradually as the user visits the particular site .As we consider a smaller environment each count for every page increases and thereby we conclude that highly visited pages are less suspicious.

Spammers may try to fabricate through some suspicious sites with a motive to increase the prestige of the page which can be eradicated by redirecting conditionally by our study.
Algorithm: Link Analysis
Initial Phase:
Generate data sets that comprise of URLs with suspicious and unsuspicious links
Classification Phase:
Initially consider page count as 1.
Increment the count as each user visit the page.
Attempt to compare with the datasets obtained
If successful, consider as normal
Else, consider as spam.

IV. CONCLUSION

Due to increase in popularity of social networks, the number of spammers also gets rapidly growing in recent years. This may lead to several spam detection techniques. We believe that this ranking mechanism by power iteration will result in good accuracy by reducing the false rate. This study can be extended by testing in web based application that will be capable of handling multiple URLs at a time.

REFERENCES

[1] Dhanalakshmi Renganayakulu, ”Detecting Malicious URLs in E-mail, AASRI Procedia,Elsevier,vol4,pp.125-131,April 2013.

[2] Jelena Isacenkova and Oliver Thonnard, “Inside Scam Jungle: a closer look at 419 scam email operations.URASIP journal of information security, Springer open journal, April 2014.

[3] Kelin.F and Strohmaier.M, “Short links under Attack: Geographical Analysis of Spam in URL Shortener Network, Proc.23 ACM Conf.Hypertext and Social media(HT),2012.

[4] Lee.S and Kim.J, Warning Bird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream IEEE transactions on secure computing, vol 10,no.3,May/June 2013.

[5] Nazpar Yazdanfar, Alex Thomo, ”Collaborative-Filtering for Recommending URLs to Twitter Users,Procedia. Elsevier, vol 19,pp.412-419,April 2013.

[6] Stringhini.G, Kruegel.C and Vigna.C, ”Detecting Spammers on Social Networks”,Proc.26th Ann. Computer Security Applications Conf.(ACSAC),2010.

[7] Song.J, Lee.S and Kim J, “Spam Filtering in Twitter Using Sender-Receiver Relationship, “Proc.14th International Symp.Recent Advances in Intrusion detection(RAID),2011.

[8] Twitter Developers, ”The t.co URL Wrapper, ”http://dev.witter.com/docs/t.co-url-wrapper,2013.

[9] Zachary Miller, ”Twitter spammer detection using data stream clustering, ”Information sciences, Elsevier, vol 260,pp.64-73,March 2014.