

Anti-Phishing Method for Detecting Suspicious URLs in Twitter

Salu Sudhakar¹, Narasimhan T²

P.G. Scholar, Dept of Computer Science, Mohandas College of engineering and technology Anad, TVM¹

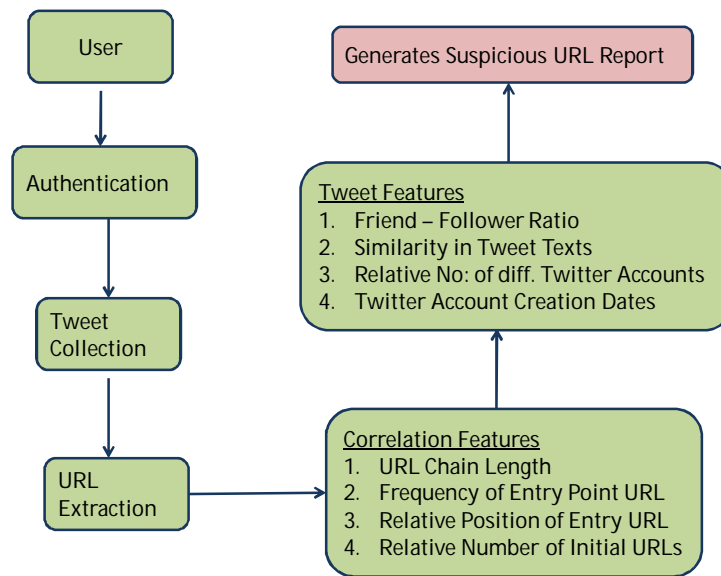
Assistant Professor, Dept of IT, Mohandas College of engineering and technology Anad, TVM¹²

ABSTRACT: Phishing is a type of web attack and it attempt to acquire information's from a website. Many popular websites are the target for this attack. Twitter is an online social networking and micro blogging service that enables users to send and read "tweets", which are text messages limited to 140 characters. Owing to this popularity many types of attacks are occur here. Attackers use shortened malicious urls that redirect users to external attack servers. Here proposed a method for detecting suspicious urls using some correlation features of urls

KEYWORDS: twitter stream, urls, correlation features

I. INTRODUCTION

Phishing is the act of attempting to acquire information such as usernames, passwords etc by masquerading as a trustworthy entity in an electronic communication. Communications purporting to be from popular social web sites, auction sites, banks, online payment processors or IT administrators are commonly used to lure unsuspecting public. Phishing emails may contain links to websites that are infected with malware .Phishing is typically carried out by email spoofing or instant messaging, and it often directs users to enter details at a fake website whose look and feel are almost identical to the legitimate one. Phishing is an example of social engineering techniques used to deceive users and exploits the poor usability of current web security technologies. Attempts to deal with the growing number of reported phishing incidents include legislation, user training, public awareness, and technical security measures. Phishers are targeting the customers of banks and online payment services. Emails, supposedly from the Internal Revenue Service, have been used to glean sensitive data from U.S. taxpayers. While the first such examples were sent indiscriminately in the expectation that some would be received by customers of a given bank or service, recent research has shown that phishers may in principle be able to determine which banks potential victims use, and target bogus emails accordingly. Social networking sites are now a prime target of phishing, since the personal details in such sites can be used in identity theft; in late 2006 a computer worm took over pages on MySpace and altered links to direct surfers to websites designed to steal login details. Experiments show a success rate of over 70% for phishing attacks on social networks



5/22/2014

11

Fig 1: flow chart of the system

II . RELATED WORKS

Most of the existing techniques make use of the various methods to detect the suspicious urls. The works in [2] uses various online algorithms such as perceptron, confidence weight and passive regression algorithm. These algorithms are used to detect malicious urls based on its lexical and host based features. But classification has some limitation. Because new features are introduced daily. In [3], [4], authors considered various spam bot activities and also different strategies to deliver spams. Developed six features which are friend follower ratio, URL ratio, message similarity ratio, friend choice ratio, message send and friend number ratio. But it takes large amount of time and expensive. It is hard to analyze each profile, because millions of users are in twitter. Then detect the anomaly via online oversampling [9] can successfully use the variation of the dominant principal direction to identify the presence of abnormal data. But it has high computational cost and it is less preferable because of its less computational efficiency.

Various black list and white list domain detections in [8] that present the characterization of spam's including its behavior. Supervised machine learning and incremental clustering module are discussed in [7]. To collect the clusters and trained a classifier to make binary decisions. Thus the system shows the malicious urls based on their average time interval based feature. It does not work for the image based spams. In [5] proposed various features about the sender receiver relationships. Such as the distance and connectivity of each accounts. Connectivity means to measure the number of paths. Detecting spammers using distance and connectivity is difficult to evade. Then design the various features based on the spam's account [6] such as graph based features, distance based features, automation based features and timing based features. Graph based features consider each account as graph and take distance between each node, then also calculate local clustering coefficient and bidirectional link ratio. The graph based features are difficult to evade and

also it is very expensive In this paper proposed a method for detecting the conditional redirections of the malicious urls based on various correlation features. Unlike the existing systems it focuses on various page level redirections. It considers the correlations of multiple redirect chains that share the same redirection server's .Here introduce various tweet based features and correlation features. Calculate threshold value for each feature and use a machine learning classifier to detect whether the incoming URL is genuine or not.

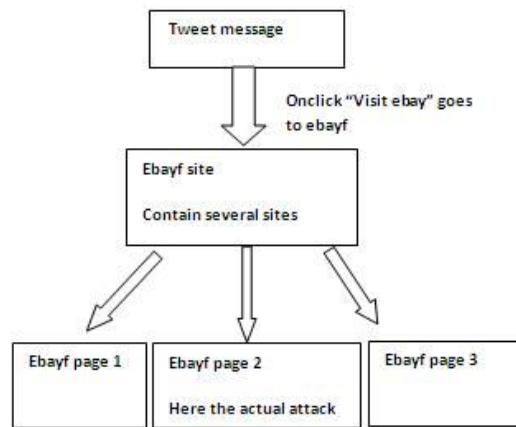


FIG 2: PAGE LEVEL REDIRECTION OF A LINK

Currently system only supports anchor tag, (a tag in html code) E.g.: href=www.ebay.com> Visit EBay doesn't support the cases of the script like Window. Location="www.ebay.com/html (this can be used in different event handling mechanism e.g. When a button click, lost focus of a textbox, this variety is the main issue for solving this) Means doesn't support dynamic script .In proposed system, add a module for processing these types of script. In this module develop rich regular expression for identify these types of script. And a text processing mechanism for iterating urls from this script Multiple Redirection. Currently phishing is detected by analyzing URL specified in the tweet. This system will not navigate to links present in that URL.

III. THE PROPOSED PRIVACY PROTECTION SYSTEM

The proposed work detecting suspicious URLs and also it blocks them by redirecting it to a genuine page, for instance google.com. The proposed system intends to detect phishing pages if included in tweets. Mainly the system consists of five steps. Tweet extraction, link extraction, calculation of correlation features, calculation of tweet features and decision component. The system is divided into several modules.

- 1) *Tweet Extraction:* Tweets are extracted from Twitter using Twitter API. An application programming interface (API) specifies how some software components should interact with each other. An API can be used to ease the work of programming graphical user interface components. Twitter API comes in the form of a library that includes specifications for routines, data structures, object classes, and variables with which an external program can access Twitter account information of a registered user. Using Twitter API, tweets are extracted and from which tweets with URLs are filtered out.
- 2) *Link Extraction:* Each URL which is being filtered out is thoroughly checked. If it is a shortened URL, URL expanding service is used to expand it. A URL expanding service does exact opposite of a URL shortening

service. Short URL and its associated expanded URL are stored in a database for future use. Using a java routine, contents of each URL is taken and again URLs are filtered out if present. In this way, such chain of URLs are identified for each URL and stored in a database.

3) *Calculation of URL correlation Features:* Since we have chain of URLs corresponding to each URL, correlation features can be calculated. 6 features are introduced in this section. There are certain terminologies associated with this module. An entry point URL is the most frequent URL that occurs when whole URL chains are considered. An initial URL is the starting URL from which that URL chain is expanded. Landing URL is the last URL in a chain. The features are as follows:

- Occurrences of starting URL: The starting point link is most main point that occurs in the window. So number of times the link that occur in the window is considered as suspicious.
- Number of times of Initial links: The beginning of the link is the initial URL. This particular link that redirect visitors to an error page. Normally attackers use different number of starting urls that redirect the visitors into multiple pages. From that feature we can classify the malicious urls based on the number of times it occur .
- Relative position of an entry point URL: The position of a suspicious links that not located at the end of a link chain. Mostly their position is at the middle stage of the link. Thus the attackers have to redirect visitors to various malicious pages. So the relative position is an important feature of the detection.

4) *Calculation of tweet features:* Some other features are obtained from similar information of tweets. Similar information's mainly include the characteristics of accounts that send the similar starting links. Resemblance checking is an effective method because there are so many twitter accounts that are not similarity for distributing spam URLs becomes a burden to attackers.

- Relative number of different source applications: Maintaining different applications are difficult for hackers. So they use the same source applications. Sources are applications that upload the current entry point URL to Twitter. Benign users, however, typically use various. Twitter applications such as Tweet Deck and Echofon. Therefore, the number of different sources may be small when the current entry point URL is suspicious.
 - Relative number of different Twitter accounts: The number of different Twitter accounts that upload the current entry point URL can be used to detect injudicious attackers who use a small number of Twitter accounts to distribute their malicious URLs.
 - Resemblance of tweet texts: Normally in retweets that containing the same links are usually same. So classify these tweets are malicious, if tweet text associated with same URL are different. To make detection more complex, attackers usually want to change the appearance of malicious tweets.
 - Similarity in the account creation dates: During short period hackers create more number of accounts. So the account creation date is similar in that accounts. Thus based on the account creation dates we have to classify each links from malicious links.
- 5) *Decision:* With the help on a threshold value, a tweet is categorized as malicious or not which is based on the presence of malicious URL in the tweet.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

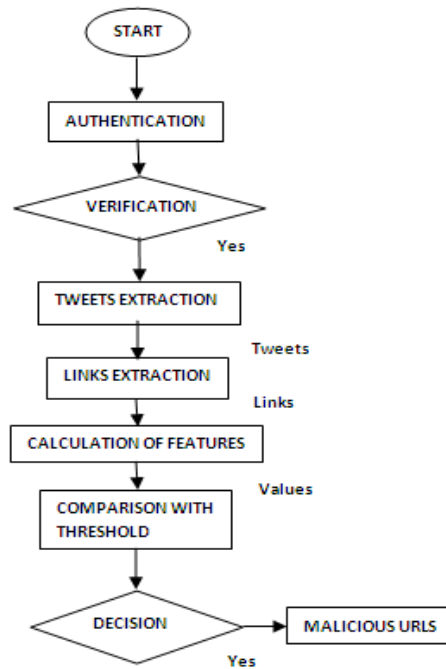


Fig 3. Proposed system architecture

Dynamic redirections can handle both http redirections and java script. Coverage and scalability will improved by using distributed accounts Future enhancements are linking twitter with tumblr, phishing detection and redirection using who is information.

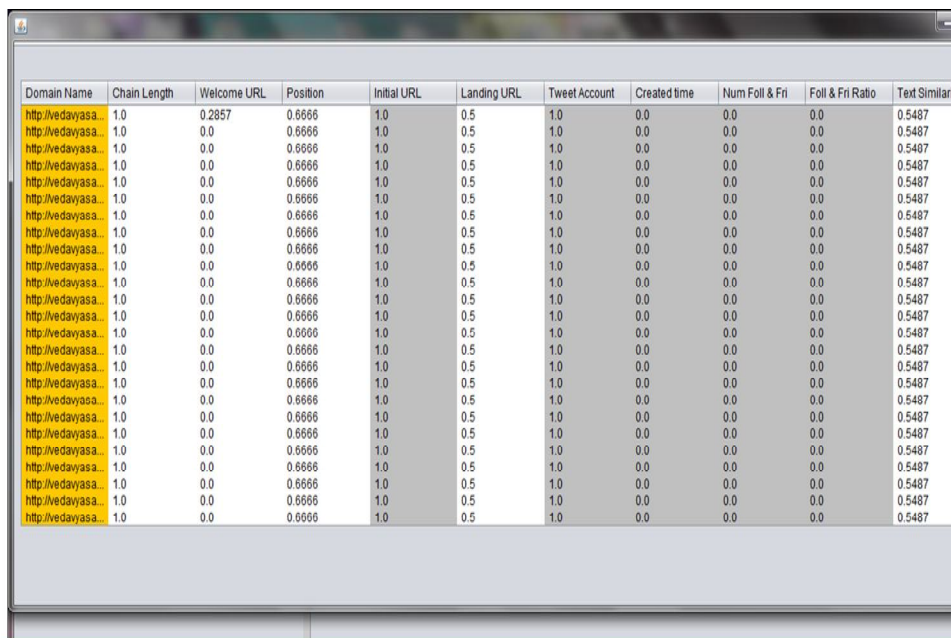
```

Domain Name: google.com
Registrant: Google Inc.
Administrative Contact:
  DNS Admin
  Google Inc.
  1600 Amphitheatre Parkway
  Mountain View CA 94043
  US
  dns-admin@google.com +1.6506234000 Fax: +1.6506188571
Technical Contact, Zone Contact:
  DNS Admin
  Google Inc.
  2400 E. Bayshore Pkwy
  Mountain View CA 94043
  US
  dns-admin@google.com +1.6503300100 Fax: +1.6506181499
Created on.....: 1997-09-15.
Expires on.....: 2020-09-13.
Record last updated on...: 2012-01-29.
  
```

Fig.4 Sample generated whois report

IV. RESULTS

Detecting suspicious urls based on correlating features of urls have completed. Next phase is to linking the twitter with tumblr, by using message extraction ,link extraction, perform correlation and text features. Then detection of phishing sites in multiuser system. For that, client system send suspicious reports to coordinator. Coordinator performs whois test to detect phishing sites. Then generates phishing url information to all the client systems. Then finally the redirection of phishing sites based on the report from coordinator. Thus phishing sites are redirected to genuine urls.



Domain Name	Chain Length	Welcome URL	Position	Initial URL	Landing URL	Tweet Account	Created time	Num Foll & Fri	Foll & Fri Ratio	Text Similarity
http://vedajasa	1.0	0.2857	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487
http://vedajasa	1.0	0.0	0.6666	1.0	0.5	1.0	0.0	0.0	0.0	0.5487

Fig: Generated suspicious URLs report

V. CONCLUSION

This work was to implement a detection mechanism which detects malicious harmful URLs from Twitter stream. Existing system used account based features and other tweet features for detection. But they are inefficient working against malicious servers that employ temporal behaviour. In this work, besides existing features some new features named correlation features are introduced. These features help distinguishing malicious and benign URLs in a better way. This system provides solution for unauthorized web pages.

REFERENCES

[1] S. Lee, J. Kim, "Warning Bird: A Near Real-time Detection System for Suspicious URLs in Twitter Stream," in IEEE Transactions on Dependable and Secure Computing, May/June 2013.

[2] Yuh-Jye Lee, Yi-Ren Yeh, "Anomaly detection via online oversampling principal component analysis IEEE transactions on knowledge and engineering 2011.

[3] J. Song, S. Lee, and J. Kim , "Spam filtering in twitter using sender receiver relationship" Proc. 14th Int'l Symp. Recent Advances in Intrusion Detection (RAID), 2011.

[4] C. Yang, R. Hark reader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers". Proc. 14th Int'l Symp. Recent Advances in Intrusion Detection (RAID), 2011.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th - 18th July -2014

- [5] F. Benevento, G. Magno, T. Rodriguez, and V. Almeida, "Detecting spammers on twitter" Proc.seventh collaboration, electronic messaging and spam conf, 2010.
- [6] K. Lee, J. Caver lee, and S. Webb, "Uncovering social spammers: Social Honey pots + machine learning "Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, 2010.
- [7] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The Underground on 140 Characters or Less" (University of California, Berkeley _University of Illinois, Champaign-Urbana Grier, vern, czhang@cs.berkeley.edu).
- [8] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards Online Spam Filtering in Social Networks"(Hongyu Gao North western University Evanston, IL, USA hyago@northwestern.edu.
- [9] J. Ma, L.K. Saul, S. Savage, and G.M. Volker, "Identifying suspicious urls: an application of large-scale online learning" proc.26th Int'I Conf.Machine Learning (ICML), 2009.