www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 3 Issue 5, May 2014, Page No. 5866-5872

Survey Paper for WARNINGBIRD: Detecting Suspicious URLs in Twitter Stream

Mr.Sulabh.S, Mr.Siva Shankar.S

Department of Computer Science and Engineering
Nehru College of Engineering and Research Centre, Pampady, Thrissur, Kerala.
sulabhls@gmail.com

Department of Computer Science and Engsineering
Nehru College of Engineering and Research Centre, Pampady, Thrissur, Kerala.
sss_siva85@yahoo.co.in

Abstract*—Today online social networks play an important role in daily life. There are various online social network(Twitter) are there and these shows tremendous growth in recent years. These kind of social networks allow users to make social connection with others. Apart from all these there are some security issues or security violations are there. This paperrelated to the system investigates correlations of URL redirect chains extracted from several tweets in Twitter. Because attackers have limited resources and usually reuse them, their URL redirect chains frequently share the same URLs. To develop methods to discover correlated URL redirect chains using the frequently shared URLs and to determine their suspiciousness. So collect numerous tweets from the Twitter public timeline and build a statistical classifier using them. Evaluation results show that our classifier accurately and efficiently detects suspicious URLs.

Keywords-Suspicious URL, twitter, URL redirection, conditional redirection, Classification

I. INTRODUCTION

Twitter is an online social networking website which allows its users to, among other things, micro-blog their daily activity and talk about their interests by posting short 140 character messages called tweets. Twitter is immensely popular with more than 100 million active users who post about 200 million tweets every day. Ease of information dissemination on Twitter and a large audience, makes it a popular medium to spread external content like articles, videos, and photographs by embedding URLs in tweets. However, these URLs may link to low quality content like malware, phishing websites or spam websites. Malware, short for malicious software, is software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems. Phishing is the act of attempting to acquire information such as usernames, passwords, and credit card details (and sometimes,indirectly, money) by masquerading as a trustworthy entity in an electronic communication.Spam is flooding the Internet with many copies of the same message, in an attempt to force the message onpeople who would not otherwise choose to receive it. Most spam is commercial advertisingRecent statistics show that on an average, 8% tweets contain spam and other malicious content. Figure 1.1 shows an example of a malicious tweet.The contributions of this paper can be summarized as follows:Since Twitter has limited tweet length, users make use of URL shortening services while posting long URLs. Owing to the popularity of Twitter, malicious users often try to find a way to attack it. WarningBird proposes a new suspicious URL detection system for Twitter which is based on the correlations of URL redirect chains, which are difficult to fabricate. The system can find correlated URL redirect chains using the frequently shared URLs and determine their suspiciousness in almost real time.Some new features of suspicious URLs are introduced. Some of the which are newly

Fig1.1. A malicious tweet saying White House has been attacked

discovered and while others are variations of previously discovered features.

The remainder of the paper is organized as follows. Section II reviews some related works.Proposed work is in Section III. Finally, the paper is concluded in Section IV.

II. RELATED WORK

In the recent past a lot of research work has been carried out for the design a better detection mechanism.

G. Stringhini, C. Kruegel, and G.Vigna in 2010 [1] used account features such as Friend-Follower ratio, URL ratio and message similarity to distinguish spam tweets. This paper analyzes to which extent spam has entered social network and how spammers who target social networking sites operate. To collect the data about spamming activity, a large and diverse set of "honey-profiles" are created on three large social networking sites and then analyzed the collected data and identified anomalous behavior of users who contacted honey-profiles. Features are developed based on the analysis of this behavior which are used for detection. A. Wang in 2010 [2] modeled Twitter as directed graph where vertices represent user accounts and the direction of edge determines the type of relationship between users, friend or follower. In this paper, detection mechanism is based on graph based features such as in-degree and out-degree of nodes and content based features such as presence of HTTP links and Trending topics in tweets. This work applies machine learning methods to automatically distinguish spam accounts from normal ones. A Web crawler is developed relying on the API methods provided by Twitter to extract public available data on Twitter website. Finally, a system is established to evaluate the detection method. J. Song, S. Lee, and J. Kim in 2011[3] viewed Twitter as an undirected graph and made use of Menger's theorem to calculate the values of message features such as distance and connectivity between nodes in order to perform detection. Here the messages are as spam or benign messages by identifying the sender. The relation featuresprototypesystem i such as distance and connectivity are unique features of social net-works and are difficult for spammers to forge or manipulate. This system identifies spammers in real-time, meaning that clients can classify the messages as benign or spam when a message is being delivered. C. Yang, R. Harkreader, and G. Gu (2011) [4] in theirwork used time based features such as following rate and tweet rate besides graph based features and content based features in order to perform detection.H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary [5] suggested a detection mechanism based on message features such as interaction history between users, average tweet rate, average number of tweets containing URL and unique URL number. In OSNs, multiple users are interacting via the message posting and viewing interface. The system inspects every message and evaluates the feature values before rendering the message to the intended recipients and makes immediate decision on whether or not the message under inspection are dropped.

Some previous works are based on URL detection schemes.J. Ma, L.K. Saul, S.Savage, and G.M. Voelker in 2009 [6] introduced a system which detects malicious websites by checking lexical features and host based features of URL. This application is particularly appropriate for online algorithms as the size of the training data is larger than can be efficiently processed in batch and because the distribution of features. Earlier works relied on batch learning algorithms. But online methods are far better for two reasons: (1)Online methods can process large numbers of examples far more efficiently than batch methods. (2)Changes in malicious URLs and their features over time can easily be adapted. D. Canali, M. Cova, G. Vigna, and C. Kruegel in 2011 found that HTML features, javascript features and UL based features can be used for effective detection of malicious websites [7].

Earlier works performed detection using honey client system. When page is loaded, the honeyclient system checks for artifacts and indicates a successful attack, such as executable files on the file system or unexpected processes. Major drawback of high interaction honeyclients is the fact that the analysis is expensive and analysis time directly limits the scalability. One approach to address the limited scalability of current analysis systems is to devise an efficientfilter that can quickly discard benign pages. Prophiler is such a fast and reliable filter thatuses static analysis techniques to quickly examine a web page for malicious content.The referrals from Twitter to understand the evolving phishing strategy isalso studied. The analysis revealed that most of the phishing tweets spread by extensive use of attractive words and multiple hashtags. In this paper, usage logs of a URL shortener service are studied that has been operated by a group for more than a year. It focuses on the extent of spamming taking place in logs, and provides first insights into the planetary-scale of this problem.

K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song[8] suggested Monarch, a real-time system that crawlsURLs as they are submitted to web services and determines whether the URLs direct to spam in 2011. This paper analyses fundamental differences between email and Twitter spam and presents a novel feature collection and classification architecture that employs an instrumented browser for detection. This system to act as a first layer of defense against spam content targeting web services, including social networks, URL shorteners, and email.

C. Whittaker, B. Ryner, and M. Nazif [9] demonstrated a system in 2011 which says that a scalable machine learning classifier can be used to automatically maintain a blacklist of phishing pages and it can achieve a very high accuracy despite a noisy training set. Phishing as a fraudulent attempt usually made through email, to steal personal information. Large amount of data is collected and system extract sand analyzes a number of features that describe the composition of the webpage's URL, the hosting of the page, and the page's HTML content as collected by a crawler. A logistic regression classifier makes the final determination of whether a page is phishing on the basis of these features.Phishing attacks have been increasing at an alarming rate and can cause damages in the form of identity theft, financial losses, and compromised security for organizations and governmental institutions.

Comparisons of above described papers are shown in the table1 below.

Existing Systems	**Scheduling Method**	**Advantage**	**Disadvantage**
Spam detection in Twitter	Native Bayesian classifier algorithm	Simple, Faster and Very efficient algorithm is used	Presence of noisy data, Features can be easily fabricated.
Large- Scale Detection of Malicious Web Pages	Prophiler analysis approach	Reduce the work load, Fast and Reliable	Need updated daily,Do not expect our filter to be accurate.
Evading high interaction honeyclients	Honeyclient mechanism	Fast and Accuracy	Honeyclients easily attack by attacker.
URL shortening services'	Blacklisting	Sensitive, Reduce work load.	Chances for hacking the URL Shortening Services, Not optimal.
Detecting Spammers on Social Networks	Random Forest Algorithm	Accurate and secure	Timeconsuming,Identify single spam profile

Table 1. Comparison table of existing systems

III. GENERAL SYSTEM MODEL

Twitter is a famous social networking and information sharing service that allows users to exchange messages of fewer than 140-character, also known as tweets, with their friends Twitter, malicious users often try to find a way to attack it. The most common forms of web attacks, including spam, scam, phishing, and malware distribution attacks, have also appeared on Twitter. Because tweets are short in length, attackers use shortened malicious URLs that redirect Twitter users to external attack WARNINGBIRD, a suspicious URL

detection system for Twitter.Twitter users want to share a URL with friends via tweets, they usually use URL shortening services to reduce the URL length because tweets can contain only a restricted number of characters. bit.ly and tinyurl.com are widely used services, and Twitter also provides a shortening service t.co. as well as the attacker's own private redirection servers used to redirect visitors to a malicious landing page. The attacker then uploads a tweet including the initial URL of the redirect chain to Twitter.

Later, when a user or a crawler visits the initial URL, he or she will be redirected to an entry point of the intermediate URLs that are associated with private redirection servers. Some of these redirection servers check whether the current visitor is a normal browser or a crawler. If the current visitor seems to be a normal browser, the servers redirect the visitor to a malicious landing page. If not, they will redirect the visitor to a benign landing page. Therefore, the attacker can selectively attack normal users while deceiving investigators corresponding IP addresses. The crawling thread appends these retrieved URL and IP chains to the tweet information and pushes this extended tweet information into a tweet queuecannot reach malicious landing URLs when they use conditional redirections to evade crawlers.However, because our detection system does not rely on the features of landing URLs, it works independently of such crawler evasions.

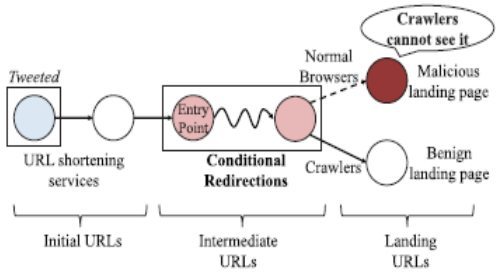

Fig 1:Conditional redirection

In investigators, cannot fetch the content of malicious landing URLs, because attackers do not reveal them to us. And also cannot rely on the initial URLs, as attackers can generate a large number of different initial URLs by abusing URL shortening services attackers may reuse some of their redirection servers when creating their redirect chains because they do not have infinite redirection servers. Therefore, if it analyze several correlated redirect chains instead of an individual redirect chain, It can be find the entry point of the intermediate URLs in these chains. In correlated redirect chains the entry point has different initial URLs and two different landing URLs, and participates in redirect chains that are several URLs long. These are the characteristics of the suspicious URLs. Therefore this correlation analysis can help to detect suspicious URLs even when they perform conditional redirection, because the suspiciousness of the two landing URLs is not important to the correlation analysis.

A. SUSPICIOUS URL DETECTION SYSTEM

WARNINGBIRD is composed of four major components:

- Data collection
- Feature extraction
- Training
- Classification

- **Data collection**:

The data collection component has two subcomponents: the collection of tweets with URLs and crawling for URL redirections. To collect tweets with URLs and their context information from the Twitter public timeline, this component uses Twitter Streaming APIs.

Whenever this component receives a tweet with a URL from Twitter, it executes a crawling thread that follows all redirections of the URL and looks up the. In crawler

- **Feature extraction**

The feature extraction component has three subcomponents: grouping identical domains, finding entry point URLs, and extracting feature vectors. This component monitors the tweet queue to check whether a sufficient number of tweets have been collected. Specifically, our system uses a tweet window instead of individual tweets. When more than w tweets are collected, it pops w tweets from the tweet queue. First, for all URLs in the w tweets, this component checks whether they share the same IP addresses. If some URLs share at least one IP address, it replaces their domain names with a list of those with which they are grouped. For when http://123.com/hello.html and http://xyz.com/hi.html, this component replaces the URL with http://['123.com','xyz.com']/hello.html and http://['123.com','xyz.com']/hi.html, respectively. This grouping process allows the detection of suspicious URLs that use several domain names to bypass blacklisting. Next, the component tries to find the entry point URL for each of the w tweets. First, it measures the frequency with which each URL appears in the w tweets. It then discovers the most frequent URL in each

URL redirect chain in the w tweets. The URLs thus discovered become the entry points for their redirect chains. If two or more URLs share the highest frequency in a URL chain, this component selects the URL nearest to the beginning of the chain as the entry point URL. Finally, for each entry point URL, this component finds URL redirect chains that contain the entry point URL, and extracts various features from these URL redirect chains and the related tweet information. These feature values are then turned into real-valued feature vectors. When it group domain names or find entry point URLs, is ignore whitelisted domains to reduce false-positive rates. Whitelisted domains are not grouped with other domains and are not selected as entry point URLs.

- **Training**

The training component has two subcomponents: retrieval of account statuses and the training classifier. Because, it use an offline supervised lea vectors relative.

To label the training vectors, use the Twitter account status; URLs from suspended accounts are considered malicious and URLs from active accounts are considered benign. If periodically update our classifier by using labeled training vectors.

- **Classification**

The classification component executes our classifier using input feature vectors to classify suspicious URLs. When the classifier returns a number of malicious feature vectors, this component flags the corresponding URLs and their tweet information as suspicious. These URLs, detected as suspicious, will be delivered to security experts or more sophisticated dynamic analysis environments for in-depth investigation.

B. FEATURES

In warning bird, following features are used for classifying suspicious URLs on Twitter. These features can be classified as features derived from correlated URL redirect chains and features derived from the related tweet context information. Describe how to normalize these feature values to real values between zero and one.

- **Features Derived from Correlated URL Redirect Chains**

URL redirect chain length: Attackers usually use long URL redirect chains to make investigations more difficult and avoid the dismantling of their servers. Therefore, when an entry point URL is malicious, its chain length may be longer than those of benign URLs. To normalize this feature, to choose an upper-bound value of 20, because most of the redirect chains have seen over the four-month period have had fewer than 20 URLs in their chains. If the length of a redirect chain is l, this feature can be normalized as min(l,20)/20.

Frequency of entry point URL: The number of occurrences of the current entry point URL within a tweet window is important. Frequently appearing URLs that are not whitelisted are usually suspicious. When the window size is w and the number of occurrences is n, this feature can be normalized as n/w.

Position of entry point URL: Suspicious entry point URLs are not usually located at the end of a redirect chain, because they have to conditionally redirect visitors to different landing URLs. If the position of an entry point of a redirect chain of length l is p, this can be normalized as p/l.

Number of different initial URLs: The initial URL is the beginning URL that redirects visitors to the current entry point URL. Attackers usually use a large number of different initial URLs to make their malicious tweets, which redirect visitors to the same malicious URL, look different. If the number of different initial URLs redirecting visitors to an entry point URL that appears n times is i, this feature can be normalized as i/n.

Number of different landing URLs: If the current entry point URL redirects visitors to more than one landing URL, It can assume that the current entry point URL performs conditional redirection behaviours and may be suspicious. If an entry point URL that appears n times redirects visitors to λ different landing URLs, this feature can be normalized as λ/n.

- **Features Derived from Tweet Context Information**

The features derived from the related tweet context information are variations of previously discovered features. Our variations focused on the similarity of tweets that share the same entry point URLs.

Number of different sources: Sources are applications that upload the current entry point URL to Twitter. Attackers usually use the same source application, because maintaining a number of different applications is difficult. Benign users, however, usually use various Twitter applications, such as TweetDeck and Echofon. Therefore, the number of different sources may be

small when the current entry point URL is suspicious. If the number of different sources of an entry point URL that occurs n times is s, this feature can be normalized as s/n.

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. If the number of Twitter accounts uploading an entry point URL that occurs n times is α, this feature can be normalized as α/n.

Standard deviation of account creation date: Attackers usually create a large number of Twitter accounts within a relatively short time period. Therefore, if the creation dates of the accounts that upload the same entry point URL are similar, it might indicate that the current entry point URL is suspicious. It use the standard deviation of account creation date as a similarity measure. To normalize the standard deviation, assume that the time difference between any account creation dates is less than or equal to one year. Therefore, this feature can be normalized as

$$\min\left(\frac{\text{std(a set of account creation date)}}{(1\text{ year})\sqrt{n}}, 1\right).$$

Standard deviation of the number of followers and number of friends: The numbers of followers and friends of attackers' accounts are usually similar, because attackers use certain programs to increase their numbers of followers and friends. If again use standard deviations to check for similarities in the numbers of followers and friends. To normalize the standard deviations, Assume that the number of followers and friends is usually less than or equal to 2,000, which is the restricted number of accounts Twitter allows one can to follow. Therefore, these features can be normalized as

$$\min\left(\frac{\text{std(\#followers or \#friends)}}{2000\sqrt{n}}, 1\right).$$

Standard deviation of the follower-friend ratio: Define the follower-friend ratio as below:

$$\frac{\min(\#\text{followers}, \#\text{friends})}{\max(\#\text{followers}, \#\text{friends})}.$$

Like the numbers of followers and friends, the follower friend ratios of attackers' accounts are similar. It use a normalized standard deviation to check the similarity as

$$\min\left(\frac{\text{std(a set of follower-friend ratios)}}{\sqrt{n}}, 1\right).$$

Because attackers accounts usually have more friends than followers, the follower-friend ratios of malicious accounts are usually different from the follower-friend ratios of benign accounts. Attackers, however, can fabricate this ratio, because they can use Sybil followers or buy followers. Therefore, instead of using an individual follower-friend ratio, it use the standard deviation of follower-friend ratios of accounts that post the same URLs and assume that fabricated ratios will be similar.

Tweet text similarity: The texts of tweets containing the same URL are usually similar. Therefore, if the texts are different, It can assume that those tweets are related to suspicious behaviors, because attackers usually want to change the appearance of malicious tweets that include the same malicious URL. If it measure the similarity between tweet texts as

$$\sum_{t,u \in \text{a set of pairs in tweet texts}} \frac{J(t,u)}{|\text{a set of pairs in tweet texts}|},$$

where J(t; u) is the Jaccard index, which is a famous measure that determines the similarity between two sets t and u, and is defined as below:

$$J(t,u) = \frac{|t \cap u|}{|t \cup u|}.$$

remove mentions, hashtags, retweets, and URLs from the texts when measure their similarity, so that if only consider the text features.

IV CONCLUSION

Suspicious URL detection system for Twitter called WARNINGBIRD.WARNINGBIRD is robust when protecting against conditional redirection, besides existing features some new features namedcorrelation features are introduced. These features help distinguishing malicious and benign URLs in a better way. This projects works in real time. Hence the time taken for detection is very less. Results show that WarningBird is much faster and efficient compared to twitter's detection system. But this work cannot discard or block when a page is detected malicious.This work can be enhanced in such a way that it can help identifying phishing pagessuccessfully.

REFERENCES

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

[2]A. Wang, "Don't Follow Me: Spam Detecting in Twitter," *Proc. Int'l Conf. Securityand Cryptography (SECRYPT), 2010.*

[3] J. Song, S. Lee, and J. Kim, "Spam Filtering in Twitter Using Sender-ReceiverRelationship," *Proc.14th Int'l Symp. Recent Advances in Intrusion Detection.*(RAID), 2011.

[4] C. Yang, R. Harkreader, and G. Gu, "Die Free or Live Hard? Empirical Evaluationand New Design for Fighting Evolving Twitter Spammers*," Proc. 14th Int'l Symp.Recent Advances in Intrusion Detection (RAID), 2011.*

[5] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards Online SpamFiltering in Social Networks," Proc. 19th Network and Distributed System SecuritySymp. (NDSS), 2012.

[6] J. Ma, L.K. Saul, S. Savage, and G.M. Voelker, *"Identifying Suspicious URLs: AnApplication of Large-Scale Online Learning,"Proc. 26th Int'l Conf. Machine Learning(ICML), 2009.*

[7] D. Canali, M. Cova, G. Vigna, and C. Kruegel, "Prophiler: A Fast Filter for theLarge-Scale Detection of Malicious Web Pages*," Proc. 20th Int'l World Wide Web Conf. (WWW), 2011.*

[8] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and Evaluation of aReal-Time URL Spam Filtering Service," *Proc. IEEE Symp. Security and Privacy (S&P), 2011.*

[9] C. Whittaker, B. Ryner, and M. Nazif, "Large-Scale Automatic Classificationof Phising Pages," *Proc. 17th Network and Distributed System Security Symp.(NDSS), 2010.*