Indian Journal of Research in Applied Sciences Engineering (IJRASE)
Vol.2.No.1 2014pp 33-35.
available at: www.goniv.com
Paper Received :05-03-2014
Paper Published:28-03-2014
Paper Reviewed by: 1. John Arhter 2. Hendry Goyal
Editor : Prof. P.Muthukumar

EXPLORATION AND BLOCKING APPREHENSIVE IN TWITTER

A.Deepak Sharma[1], P.Savaridassan[2]
SRM University, kattankulathur, Tamil Nadu, India
[1] imdeepaksharma222@gmail.com, [2] savaridassan.p@ktr.srmuniv.ac.in

ABSTRACT

Twitter can suffer from malicious tweets containing suspicious URLs for spam, phishing, and malware distribution. Previous Twitter spam detection schemes have used account features such as the ratio of tweets containing URLs and the account creation date, or relation features in the Twitter graph. Malicious users, however, can easily fabricate account features. Moreover, extracting relation features from the Twitter graph is time and resource consuming. Previous suspicious URL detection schemes have classified URLs using several features including lexical features of URLs, URL redirection, HTML content, and dynamic behavior. However, evading techniques exist, such as time-based evasion and crawler evasion. In this project we are focusing on the landing pages of individual URLs in each tweet, we consider correlated redirect chains of URLs in a number of tweets. Because attackers have limited resources and thus have to reuse them, a portion of their redirect chains will be shared. We focus on these shared resources to detect suspicious URLs. We have created our own twitter like application to collect some amount of tweets and trained a statistical classifier with features derived from correlated URLs and tweet context information. Finally, we block the user, who generates malicious URL.

Keywords— Twitter, spam,tweets, suspicious url,crawlers, classification.

1. INTRODUCTION

Twitter is famous social networking and microblogging service that enables users to post and read tweets. Twitter was 20th most visited site on internet. This uses to create a username and brief about profile like name, photograph, location information, short note and web address. Many businesses, government agencies, news media outlet, popular public figure and other joined the twitter. They use social media for reporting natural disasters and in business and marketing it used for product and service review. There are 175million tweets per day and more than one million new accounts are added everyday. It is high number of influencers. That allows users to exchange messages fewer than 140-characters. When a user1 updates a tweet, it will be distributed to all of user1's followers who have registered user1 as one of their friends. Instead of distributing a tweet to all of user1's followers, Unlike status updates, mentions can be sent to users who do not follow user1. When 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. Twitter uses **bit.ly** and **tinyurl.com** are widely used for shortening services.

Popularity of twitter, malicious users often try to find a way to attack it. The most common attacks can be created is web attacks, spam, scam, phishing, and malware distribution attacks. Tweets are short in length, so attackers use shortened malicious URLs that can be redirect twitter users.

In this crawler can be used for collecting the tweets. They can be static and dynamic crawlers, and they may be executed in virtual machine honeypots. A number of suspicious URL detection schemes have also be introduced. This schemes classify features like DNS information, URL redirections and HTML content. So malicious servers can delivery only to normal browsers. After it investigators use dynamic crawlers with functionalities of real browsers. Malicious servers can identify through their IP addresses, user interaction.

We present new suspicious URL detection system for twitter based correlations of URL redirect chain. In paper, we propose the landing pages of individual URLs in each tweet, we consider correlated redirect chains of URLs in a number of tweets. Because attackers have limited resources and thus have to reuse them, a portion of their redirect chains will be shared. We focus on these shared

resources to detect suspicious URLs. We have created our own twitter like application to collect some amount of tweets and trained a statistical classifier with features derived from correlated URLs and tweet context information. Finally, we block the user, who generates malicious URL.

Existing system:

A number of suspicious URL detection schemes have also been introduced. They use static or dynamic crawlers. An attacker creates a long URL redirect chain by using public URL shortening services, such as bit.ly and t.co, and his or her own private redirection servers to redirect visitors to a malicious landing page. The attacker then uploads a tweet including the initial URL of the redirect chain to Twitter

Rank	July 23	July 24	July
1	24newpress.net	24newspress.net	24newpress.ne
2	blackraybansunglasses.com	blackraybansunglasses.com	blackraybansu
3	software-spot.com	cheapdomainname.info	bigfollow.net
4	ustream.tv	ustream.tv	twitmais.com
5	10bit.info	twitmais.com	jbfollowme.cc
6	blackreferrer.com	bigfollow.net	addseguidores
7	tweetburner.com	jbfollowme.com	elitebrotherh
8	livenation.com	10bit.info	livenation.cc
9	twitmais.com	addseguidores.com.br	naturesoundcc
10	bigfollow.net	wayjump.com	all-about-leg

Fig1: some of URL redirect chains

. Later, when a user or a crawler visits the initial URL, he or she will be redirected to an entry point of Intermediate URLs that are associated with private redirection servers. Some of these redirection servers will check whether the current visitor is a normal browser or a crawler. If the current visitor seems to be a normal browser, they will redirect the visitor to a malicious landing page. If not, they will redirect the visitor to a benign landing page.

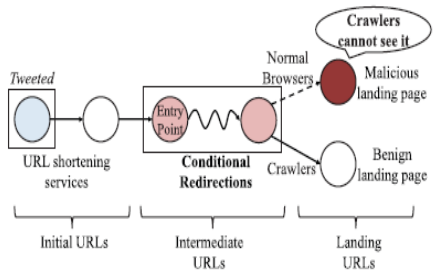

Fig2: Conditional redirect

Motivation:

we propose a suspicious URL detection system for Twitter. Instead of investigating the landing pages of individual URLs in each tweet, which may not be successfully fetched, we considered correlated redirect chains of URLs included in a number of tweets. Because attackers resources are limited and need to be reused, a portion of their redirect chains must be shared. We are going to find a number of meaningful features of suspicious URLs derived from the correlated URL redirect chains. And we can enhanced privacy and security for online URL posting.

2. CASE STUDY AND OBSERVATION

2.1 Redirection URL chains

Blackraybansunglass.com is one of the suspicious site associated with spam tweets. Then this was come upon this April 2011 and it active till August 2011. It can also uses redirect.php, which conditionally redirects. And it generates random spam pages with different twitter accounts and distribute its URL to other twitter users. When user clicks on one of the shortened URLs, example if bit.ly/raCz5i spread by hashahruel09, he or she will be redirected to other site. Twitter spam program like tweetattacks.com which uses web interface to deceive spam receivers will be sold.

24newspress.net is also a kind of suspicious site. But in this does not perform conditional redirection to avoid investigate. It also uses a number of IP address and domain names for cloaking like IP fast flux and domain flux methods.

2.2 Observation

From identify suspicious URLs a number of different twitter accounts and shortened URLs or domain names and IP addresses to cloak the same suspicious URLs. To avoid investigations they may also use long redirect chain. It may appear frequently in posting tweets by with or without urls.

3. SYSTEM DETAILS

In this overview the redirect chains are grouping the detection of suspicious URLs that use several domain names.

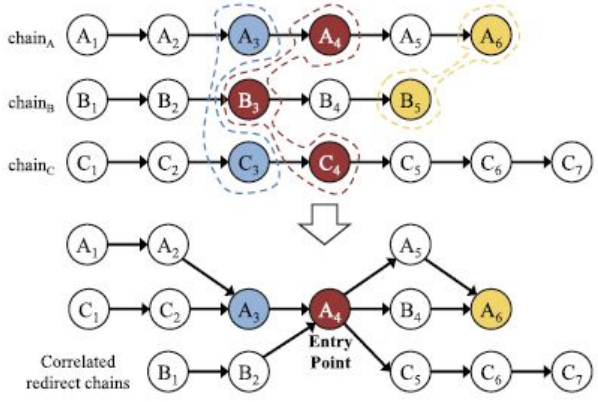

Fig3: Redirect chains and their correlation.

It depends on the data collection to the collection of tweets with URLs and crawling for URL redirections. To collect the context information from twitter public timeline, this is used by twitter streaming APIs. Crawling thread executes the following redirect URLs and looks up the corresponding IP addresses. When retrieved URL and IP chains from tweet information and into tweet queue, crawler cannot reach malicious landing URLs when they use conditional redirections to evade crawlers.

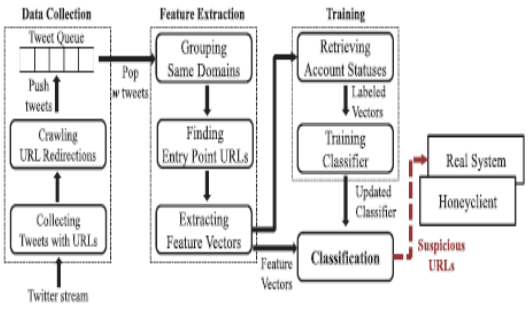

Fig4: System overview

From feature extraction the several domains is been grouped. So it method is to find to correct entry point of suspicious URLs domain. After finding entry point the trained component and sub component is used in training function. It can be done in offline controlled feature vectors for training are relatively older than feature vectors for classification.

4. TRAINING AND TESTING CLASSIFIERS

To implement our classifier we compare several algorithms L2-regularized L1-loss support vector classification(SVC) algorithm, it compares highest AUC and lowest FP with the training data set.

5. RELATED WORK

5.1 Twitter spam detection

In twitter spam detection shows how to collect a large number of spam and nonspam accounts and extract the features. To detect manually analyze the twitter public timeline. Some sample for spam are honey-profile and blacklisted URLs and it monitor twitter's official account for spam reporting. The number of followers, account creation dates, tweets which can be efficiently collected but easily made. In this we foscused on relation between spam nodes and their neighboring nodes such as redirect link. In additional, increasing spammers social influences using link farming.

5.2 Suspicious URL Detection

Most of suspicious URL detection classified static and dynamic detection system. Some times static detection system focus on the lexical and takes information. Additionally extract features from HTML content and JavaScript codes to detect attacks. So static detection cannot detect suspicious URLs with dynamic content. So we need dynamic system that use virtual machines and web applications for analysis of suspicious URLs. But still it fails to detect suspicious sites by using dynamic detection.

6. CONCLUSION

Twitter uses usually URL shortening services to reduce the URL length because tweets can contain only a restricted number of characters. Twitter uses **bit.ly** and **tinyurl.com** are widely used for shortening services. . They can be static and dynamic crawlers, and they may be executed in virtual machine honeypots. Instead of focusing on the landing pages of individual URLs in each tweet, we consider correlated redirect chains of URLs in a number of tweets. Because attackers have limited resources and thus have to reuse them, a portion of their redirect chains will be shared. We focus on these shared resources to detect suspicious URLs. To avoid suspicious URL the landing page is concentrated so that the traffic will be reduced from redirect pages. In future we can enhanced privacy and security for online shorten URL posting in twitter and can be create Add-ons for web applications.

REFERENCE

[1] Twitter Developers, "Streaming API," https://dev.twitter.com/docs/streaming-apis

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

[3] K. Lee, J. Caverlee, and S. Webb, "Uncovering Social Spammers: Social Honeypots Machine Learning," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, 2010.

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

[5] http://www.cse.lehigh.edu/~chuah/publications/atc11_spam_camera.pdf,"Detecting Spammers on Twitter,"