ISSN: 2321-7782 (Online)
Volume 2, Issue 2, February 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Classifier: A Real-Time Detection system for suspicious URLs in Twitter Stream

A. Jenefa[1]
M.E.
Network Engineering
Francis Xavier Engineering College
Tirunelveli - India

Dr. R. Ravi[2]
M.E., Ph.D
Professor & Head of CSE
Francis Xavier Engineering College
Tirunelveli - India

Abstract: ***Social Network is susceptible to malicious message containing URLs for spam, phishing, and malware distribution. Mainly we are dealing with Twitter, in that we are finding suspicious address by using application tool both by creating manually and also from that manual coded tool we are calling API. Conventional Spam detection schemes are ineffective against feature fabrications or consume abundant time and resources. In this paper, we propose WarningBird, a suspicious address detection system for twitter. Our system investigates correlations of URL redirect chains frequently share the same URLs. We develop methods to discover correlated URL by using application tool. We collect numerous tweets from the twitter public timeline and build a statistical classifier using them and also we are creating multiple accounts in that tool. We are finding suspicious URLs in that created account by analyzing URLs published in that account. Evaluation result shows that our classifier accurately and efficiently detects suspicious URLs. Also present our system work as a real-time detection system for classifying suspicious address in the twitter stream.***

Keywords: Conditional redirection, Classifier, Suspicious address, URL Shortening, Application tool, WARNINGBIRD.

I. INTRODUCTION

Twitter, Facebook, Google+ is a well-known social networking and information sharing service that allows users to exchange messages. In twitter mainly we can use message of fewer than 140 characters, also known as tweets, with their friends. When a user Jene Updates (or sends) a tweet, this tweet will be distributed to all of her followers, who have registered Jene as one of their friends. Instead of distributing her tweets to all of her followers, Jene can send a tweet to a specific twitter user Keerthi by mentioning this user by including @Keerthi in the tweet. Unlike status updates, mentions can be sent to users who do not follow Jene. When Twitter users want to share URLs with friends via tweets, they usually use URL shortening services to reduce the length of these URLs, because tweets can only contain a restricted number of characters [7]. Bit.ly and tinyurl.com are widely used services, and Twitter also provides its own shortening service [1][2]. Owing to the popularity of Twitter, malicious users often try to find a way to attack it. Because tweets are short in length, attackers use shortened malicious URLs that redirect Twitter users to external attack servers. To cope with malicious tweets, many Twitter spam detection schemes have been proposed. In this paper, we propose Application tool, a suspicious URL detection system for Twitter [2]. By using that tool we can add multiple twitter user account, initially we should enable follow, unfollow, tweet, retweet etc... And also by clicking tweet in that we can enter our message, it will be automatically send or update particular message in the twitter public timeline. If any suspicious URLs are there in that account means it can be easily detected and corrected by our Application tool. Because attackers' resources are limited and need to be reused, a portion of their redirect chains must be shared. So by using Application tool we are classifying suspicious URLS from normal URLs. We are using java as front end and MySQL as back end. In such a way we are designing system to efficiently classify suspicious URLs from normal URLs.

Whatever we are updating in application tool, it will be automatically updated to our corresponding twitter account so, it is more efficient one compare to all previous suspicious URLs detection system.

II. MALICIOUS SITES

A. Suspicious site-Blackraybansunglasses.com

We think about blackraybansunglasses.com that may be a suspicious website related to spam tweets. Here malicious server redirecting user, initially they will redirect both normal user and crawler [1]. After some correlation they will check whether the user is normal user or crawler, if the user is crawler means it will simply redirect them to Google page. Suppose if the user is normal user means it will redirect them again and make as to enter malicious page.

B. *Suspicious Site- 24newspress.net*

Let us conjointly discuss 24newspress.net that may be a suspicious website distributed via tweets. We tend to initially found this website at the tip of June 2011 and it remains active.

C. Observations concerning malicious sites

From the previous examples, we are able to determine meaningful characteristics of suspicious website. They use variety of various twitter accounts and shortened URLs to cloak constant suspicious URLs. They conjointly use long redirect chains to avoid investigation. These characteristics are the basis for the feature models we employ to classify suspicious address. They are having solely restricted resources, thus we are able to simply decide suspicious URLs.

III. PROPOSED SYSTEM

A. *Motivation and Basic plan concerning Classifier*

Our goal is to develop a suspicious address detection system for Twitter that is robust enough to safeguard against conditional redirection. An attacker creates a long URL redirect chain by using uniform resource locator shortening services, such as bit.ly and tinyurl.com, to redirect visitors to a malicious landing page. The attacker then uploads a tweet including the initial uniform resource locator 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 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 but we can avoid above problem using WARNINGBIRD tool. Therefore, if we analyze a number of correlated redirect chains instead of an individual redirect chain, we can find suspicious URLs in the correlated redirect chains. And also application tool is used to collect all information and classify URLS.

B. *System Details*

Our system is composed of three major components: data collection, feature extraction, WARNINGBIRD tool.

1. Information collection:

The information 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 WARNINGBIRD tool [1]. As we have seen, our crawler cannot 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. By adding individual accounts in Application tool, we can collect information from that particular account from tool itself.

2. Feature extraction:

The feature extraction main function is extracting feature vectors. This component monitors the tweet queue by using WARNINGBIRD Application tool to check whether a sufficient number of tweets have been collected. In feature extraction we are extracting features based on URL redirection, HTML content and also based on number of correlations.

3. WARNINGBIRD tool:

The WARNINGBIRD tool component has two subcomponents: retrieval of account statuses and the training classifier. We periodically update our classifier by using Offline Supervised Learning Algorithm. Here our WARNINGBIRD Application tool is efficiently trained to classify URLs.

The classification component executes our classifier that is WARNINGBIRD Application tool using input feature vectors to classify suspicious URLs [1]. When the classifier returns a number of malicious feature vectors, this component flags the corresponding URLs and their tweet information as suspicious. In classification we are classifying suspicious URLs from normal URLs. Attacker having only limited resources so that we can effectively classify suspicious URLs by using application tool.

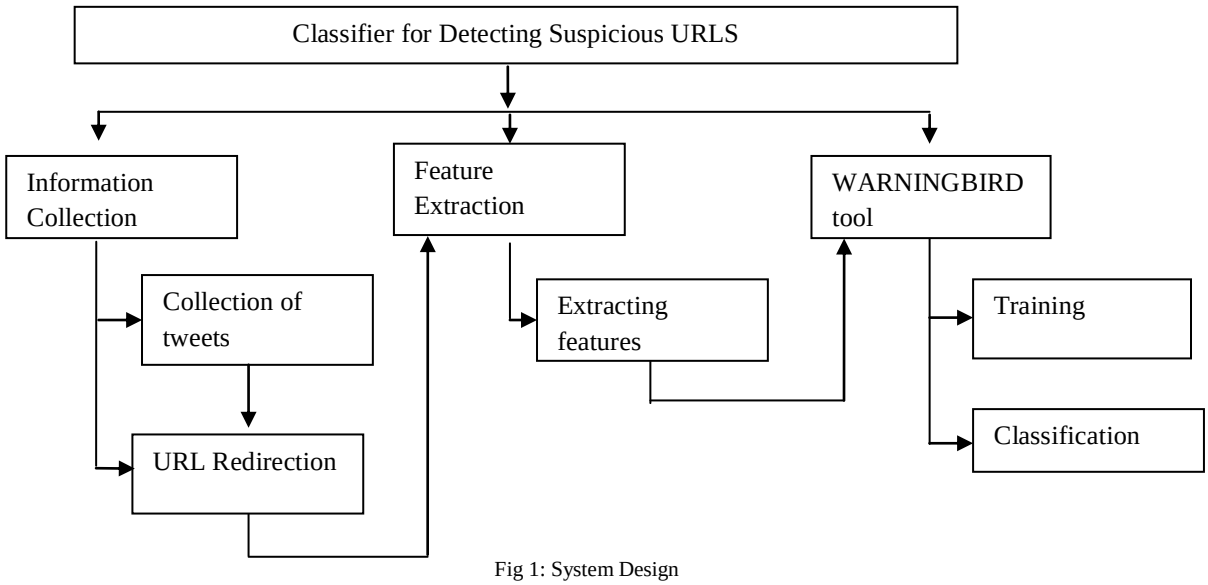

Fig 1: System Design

C. Options concerning Suspicious URLS

We introduce many features for classifying suspicious URLs on Twitter [1]. These features can be classified as features derived from correlated URL redirect chains and features derived from the related tweet context information using WARNINGBIRD Application tool.

1. Features Derived from Correlated URL Redirect Chains in Suspicious Site

We are using application tool that is available in www.tweetpro.com, from that tool we are identifying various features to find out suspicious URLs. Some of the features are as below.

2. URL redirects chain length in Suspicious Site

Attackers usually use long URL redirect chains to make investigations more difficult. By using our statistical classifier information we can easily find out URL redirect chain length. From that information we can classify URLs by using WARNINGBIRD Application tool both by using manual and real time tool.

3. Frequency of entry point URL in Suspicious Site

The number of occurrences of the current entry point URL within a tweet window is important. We are getting frequently occurring URLs from twitter public timeline and build a statistical classifier using them.

IV. EVALUATION

A. *System Setup and Information assortment*

To collect tweets, we are using website www.tweetpro.com by opening that we can get tweepro page from that we can add many account and the same time we can store previously detected suspicious URLs from that we can classify suspicious URLs [11].

B. *Real-time*

The real-time version of Application tool is used for achieving good latency and detection coverage. In real time we can detect suspicious URLs in twitter stream and add that particular suspicious URLs into banned account [11]. Because suspicious URL is very dangerous one, if suppose if we are clicking that means malicious server can easily able to hack our related account username and password. This type of application tool used to easily separate suspicious URL from normal URL in real time use.

C. *Comparison with Twitter*

We compare the efficiency of WARNINGBIRD Application tool with that of Twitter's detection system. Now by our Application tool we can efficiently detect suspicious URLs in real time. And also previously for spam detection we need to put lot effort for training network. But here by using Application tool we can accurately and efficiently detect suspicious URLs.

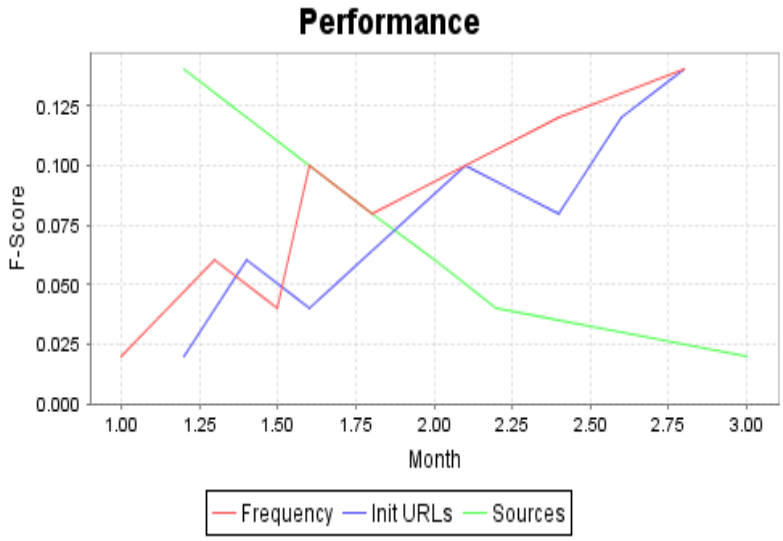

Fig 2: Performance of WARNINGBIRD tool

V. DISCUSSION

In this paper, we tend to discuss some limitations of our system and possible evasion techniques. Currently, WARNINGBIRD Application tool provide output for added account. So it can handle only that added account it won't find suspicious URLs from other twitter accounts. Conjointly its determination causes inaccuracy in some of the feature values, including the redirect chain lengths, positions of the entry point URLs, and the number of different landing URLs [11]. Therefore, in the future we will use customized Web browsers to retrieve redirect chains fully. If we are getting permission from twitter stream and can be able to see all twitter user, and finding suspicious URLs from unadded account too means it's much more efficient. And conjointly we didn't worked with Facebook and Google + but it is also possible to find suspicious URLs in Facebook and Google+ too.

VI. Related Work

A. *Manual Application Tool*

Many twitter suspicious URL detection schemes have been introduced. Most have focused on how to collect a large number of spam and non-spam accounts and extract the features that can effectively distinguish spam from non-spam accounts. To avoid feature fabrication, Manual tool relies on more robust features extracted from the Twitter public timeline. The extraction of these robust features, however, is time and resource consuming

B. *Real-time Application Tool*

Many suspicious URL detection schemes have been proposed. Here in real time Application tool by entering our account details we are automatically login to particular twitter account without using twitter account. From that tool we can add multiple accounts, in that account we should enable all task after that check each and every task and then run task to classify URLs. It is a real-time suspicious URLs detection system by building statistical classifier.

VII. Conclusion

Previous suspicious uniform resource locator detection systems are weak at protecting against conditional redirection servers that distinguish investigators from normal browsers and redirect them to benign pages to cloak malicious landing pages. In this paper, we propose a new suspicious URL detection system for Twitter, WARNINGBIRD Application tool [11]. Unlike the previous systems, WARNINGBIRD Application tool is robust when protecting against conditional redirection, because it does not rely on the features of malicious landing pages that may not be reachable. Instead, it focuses on the correlations of multiple redirect chains that share redirection servers. We introduced new features on the basis of these correlations, implemented a real-time classification system using these features, and evaluate the system's accuracy and performance. The evaluation results showed that our system is highly accurate and can be deployed as a real-time system to classify large samples of tweets from the Twitter public timeline.

In the future, we will extend our system to address dynamic and multiple redirections. We will also implement a distributed version of WARNINGBIRD Application tool to process all tweets from the Twitter public timeline. And also we will also implement technique as without adding multiple accounts itself we need to find suspicious URLs for all twitter users by using some additional capacity Application tool.

Acknowledgement

First of all I thank the almighty for giving us the knowledge and courage to complete the research work successfully. Next I express my gratitude to our respected Professor and Head of CSE Dr R. Ravi M.E., Ph.D., for following us to do research work intentially. Finally, I wish to express my gratitude to my family and my friends, for providing unending support in every aspect possible, to ease my path on this journey.

References

1. S. Lee and J. Kim, "WarningBird: Detecting Suspicious URLs in Twitter Stream," Proc. 19th Network and Distributed System Securitysymp. (NDSS), 2012.
2. H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?" Proc. 19th Int'l World Wide Web Conf.(WWW), 2010.
3. D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S.Ioannidis, E.P. Markatos, and T. Karagiannis, "we.b: The Web ofShort URLs," Proc. 20th Int'l World Wide Web Conf. (WWW), 2011.
4. D.K. McGrath and M. Gupta, "Behind Phishing: An Examination of Phisher Modi Operandi," Proc. First USENIX Workshop Large-Scale Exploits and Emergent Threats (LEET), 2008.
5. Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who Is Tweeting on Twitter: Human, Bot, or Cyborg?" Proc. 26th Ann. Computer Security Applications Conf. (ACSAC), 2010.
6. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting Spammers on Social Networks," Proc. 26th Ann. Computer Security ApplicationsConf. (ACSAC), 2010.

7. C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The Underground on 140 Characters or Less," Proc. 17th ACM Conf.Computer and Comm. Security (CCS), 2010.
8. S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru,"Phi.sh/$oCiaL: the Phishing Landscape through Short URLs,"Proc. Eighth Ann. Collaboration, Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS), 2011.
9. F. Klien and M. Strohmaier, "Short Links under Attack:Geographical Analysis of Spam in a URL Shortener Network,"Proc. 23rd ACM Conf. Hypertext and Social Media (HT), 2012.
10. K. Lee, J. Caverlee, and S. Webb, "Uncovering Social Spammers:Social Honeypots þ Machine Learning," Proc. 33rd Int'l ACMSIGIR Conf. Research and Development in Information Retrieval, 2010.
11. www.tweetpro.com
12. www.google.com

AUTHOR(S) PROFILE

A. Jenefa is presently studying M.E second year Network Engineering in Francis Xavier Engineering College. She has completed her B.E Computer Science and Engineering from Loyola Institute of Technology and Science. Her fields of interests are Network security, Data Mining, Neural Network, Computer Applications in medicine and Network Engineering.

Dr. R. Ravi is an Editor in International Journal of Security and its Applications (South Korea). He is presently working as a Professor & Head and Research Centre Head, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli. He completed his B.E in Computer Science and Engineering from Thiagarajar College of engineering, Madurai in the year 1994 and M.E in Computer Science and Engineering from Jadavpur Government research University, Kolkatta. He has completed his Ph.D in Networks from Anna University Chennai. He has 18 years of experience in teaching as Professor and Head of department in various colleges. He published 12 International Journals, 1 National Journal. He is also a full time recognized guide for various Universities. Currently he is guiding 18 research scholars. His areas of interest are Virtual Private networks, Networks, Natural Language Processing and Cyber security.

