ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

An Authentication for Social Network from Cautious URLs

N.Aravindhu[1], M.Arun Yokesh[2] , T.Manoharan[3], M.Sivasubramanian[4]

Assistant Professor, Christ College of Engineering and Technology , Puducherry, India[1]

UG Student, Christ College of Engineering and Technology , Puducherry, India[2]

UG Student, Christ College of Engineering and Technology , Puducherry, India[3]

UG Student, Christ College of Engineering and Technology , Puducherry, India[4]

Abstract: Social Network contains many URLs for spam, cautious and suspisious circulation. Conservative Social Network spam discovering schemes make use of account features such as the ratio of status update containing URLs and the account creation information and relation features in the Social Network chat. These Discovering schemes are unproductive beside feature fabrications or put away much time and resources. Conservative suspicious URL detection schemes make use of many features along with URLs lexical features , URL redirection mechanism, HTML web content and dynamic behavior. Evading techniques such as point based evasion and crawler evasion exist. In this paper, we propose detection and blocking scheme, a suspicious URL discovering system for Social Network. Our system investigates correlations of URL redirect chain extracted from several status update. Because attackers have limited resource and usually recycle them, their URL redirecting chains often share the same URLs. We expand methods to discover interrelated URL redirects chains using the often shared URLs and to decide their suspiciousness. We collect serveral status update from the Social Network public timeline and built a numerical classifier using them. Estimation results show that our classifier exactly and powerfully detects suspicious URLs. The Detected malicious status update that containing URLs are blocked using a Gibraltar prototype.

Index Terms: Social Network, Suspicious URL, Blocking, URL Redirection.

I. INTRODUCTION

In a famous social networking and information sharing services that allows users to exchange messages of less than 140-character, also known as status update, with their friends. When a user Alice updates a status update, it will be distributed to all of her followers and friends who have registered Alice as one of their friends. Instead of distributing a status to all of her followers, Alice can also send a status to a specific user Bob by mentioning this user by including @Bob in the social network. Unlike status updates, mentions can be sent to users who do not follow Alice. When social network users want to share a URL with friends via status, they usually use URL shortening services to reduce the URL length because status update can contain only a restricted number of characters. Bit.ly and tinyurl.com are widely used services. Owing to the popularity of social network, malicious users frequently try to find a way to attack it. The most common forms of web attacks, including spam, scam, phishing, and malware spreading attacks, have also appeared. Because status is short in length, attackers use shortened malicious URLs that redirect social network users to other external attack servers. To manage with malicious status, several social network spams discovering schemes have been proposed earlier and that can be classified into user account feature - based, friends feature-based, and message feature-based schemes. A number of cautious URL detection schemes in social network also have been discovered. They use static or dynamic crawlers, and they may be executed in virtual machine honeypots, such as Capture-HPC, HoneyMonkey, and Wepawet to investigate newly observed URLs. These schemes classify URLs according to several features including lexical features of URLs, DNS information, URL redirections, and the HTML content of the landing pages. Nevertheless, cautious servers can get around a research by selectively giving benign pages to crawlers users. For example, because static graph of crawler's users usually cannot handle JavaScript or Flash, suspisious servers can use them to deliver suspisious content only to standard browsers. Even if investigators use dynamic crawlers with all of the functionalities of real browsers, malicious servers may be able to recognize them through their IP addresses, user interaction, browser fingerprinting, or honeyclient detection techniques. A topical scientific statement from Google has also discussed techniques for escaping current web malware discovering systems. Suspisious servers can also employ temporal behaviors providing different content at different times to evade an investigation. In this paper, we propose blocking scheme, a suspicious URL detection system for social network. The qualified classifier is shown to be precise and has low mistakes and negatives. The offering of this paper are as

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

follows: We present a new suspicious spam URL discovering system for social network that is based on the interconnected of URL redirect chains, which are hard to construct. The system can find interconnected URL redirect chains using the often shared URLs and decide their suspiciousness in almost real time on the social network. We launch new features of cautious URLs: Some of which are newly discovered and while others are variations of formerly discovered features. We present the outcome of investigations conducted on cautious URLs that have been widely spreader through social network over several months.

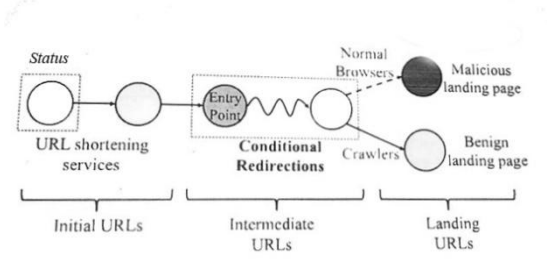

Fig. 1. URL Redirection

II. EXISTING SYSTEM

In the existing system attackers use shortened malicious URLs that redirect social network users to external attack servers. To manage with malicious status, several social network spams discovering schemes have been proposed earlier and that can be classified into user account feature - based, friends feature-based, and message feature-based schemes. A number of cautious URL detection schemes in social network also have been discovered on more robust features that spam or unauthorized users cannot easily construct such as the space and connectivity apparent in the graph.

Extracting these relation features from a graph, however, requires a important amount of time and resources as a social network graph is incredible in size. The message feature-based scheme alert on the lexical features of the status messages. However, spammers and suspicious users can easily change the shape of their messages. A number of suspicious URL detection schemes have also been introduced.

A. Drawbacks

- Malicious servers can bypass an investigation by selectively providing benign pages to crawlers.
- For instance, because static crawlers usually cannot handle JavaScript or Flash, malicious servers can use them to deliver malicious content only to normal browsers.
- A recent technical report from Google has also discussed techniques for evading current Web malware detection systems.
- Malicious servers can also employ temporal behaviors— providing different content at different time to evade an investigation

III. PROPOSED SYSTEM

In this paper, we propose detecting and blocking scheme, a suspicious URL detection system for social network. In the place of probing the landing pages of individual URLs in each status, which may not be successfully fetched, we considered connections of URL redirect chains extracted from a number of status update. Because invader's resources are generally limited and need to be reused, their URL redirect chains usually share the same URLs. We therefore created a method to detect simultaneous URL redirect chains using such frequently shared URLs. By analyzing the simultaneous URL redirect chains and their status context information, we discover several features that can be used to classify suspicious URLs. We collected a large number of status update from the social network public timeline and trained a statistical classifier using the discovered features.

A. Advantage

- We present a new suspicious URL detection system for social network that 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.
- We introduce new features of suspicious URLs: some of which are newly discovered and while others are variations of previously discovered features.
- We present the results of investigations conducted on suspicious URLs that have been widely distributed through Twitter over several months.

IV. MODULES

- Data collection
- Feature extraction
- Training and classification
- Blocking

A. Data collection

- The collection of status with URLs.
- Crawling for URL redirections.

The data collection component has two subcomponents: The collection of status with URLs and crawling for URL

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

redirections. To collect status with URLs and their context. Whenever this component obtains a status with a URL, it executes a crawling thread that follows all redirections of the URL and looks up the corresponding IP addresses. At last the URL was pushes it into a URL queue.

B. Feature Extraction

- Grouping of identical domains
- Finding entry point URLs
- Extracting feature vectors.

The feature extraction component has three subcomponents: grouping of identical domains, finding entry point URLs, and extracting feature vectors. This component checks whether they share the same IP addresses. If several URLs share at least one IP address, it replaces their domain names with a list of domains with which they are grouped. If several URLs share at least one IP address, it replaces their domain names with a list of domains with which they are grouped. This grouping process enables the detection of suspicious URLs that use several domain names to bypass the blacklisting. The component finds URL redirect chains that contain the entry point URL, and extracts various features from these URL redirect chains. Ignore white listed domains to reduce false-positive rates. White listed domains are not grouped with other domains and are not selected as entry point URLs.

C. Training and Classification

- Retrieval of account statuses
- Training of the classifier.

The training component has two subcomponents: Retrieval of account statuses and training of the classifier. To label the training vectors, we use the account status; URLs from suspended accounts are considered malicious, whereas URLs from active accounts are considered benign. 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 status information as suspicious.

D. Blocking

All the above modules explain only how to detect the suspicious URL but this module used to block the suspicious URLs. Using Gibraltar prototype we block the URL. It obtains the corresponding memory page from the target. The specification for URLs limits the allowed characters in a Request-URI to only a subset of the ASCII character set. This means that the query parameters of a request-URI beyond this subset should be encoded. Because a malicious payload may be embedded in the request-URI as a request parameter.

V. SYSTEM ARCHITECTURE

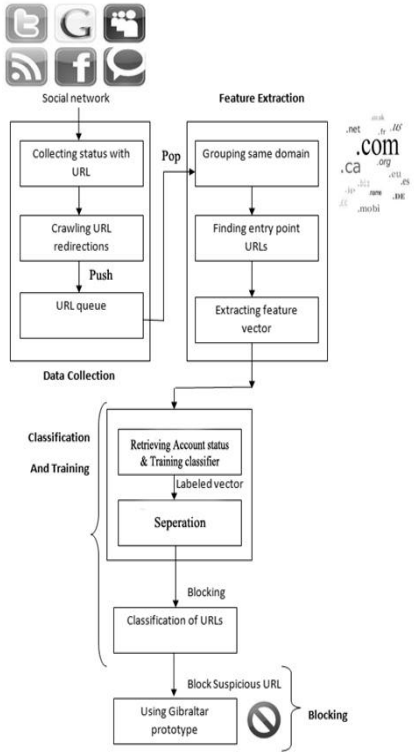

Fig. 2. System Architecture Design

VI. CONCLUSION

Conventional suspicious URL detection systems are ineffective in their protection 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 proposed a new suspicious URL detection system for social network, called WARNINGBIRD. Unlike the conventional systems, WARNINGBIRD 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 the same redirection servers. We introduced new features on the basis of these correlations, implemented a near real-time classification system using these features, and evaluated the system's accuracy and performance. The evaluation results show that our system is highly accurate and can be deployed as a near real-time system to classify large samples of status from the social network 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 to process all status from the user public timeline.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

REFERENCES

[1] S. Lee and J. Kim, "WarningBird: Detecting Suspicious URLs in Twitter Stream," Proc. 19th Network and Distributed System Security Symp. (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 of Short 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 Applications Conf. (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 ACM SIGIR Conf. Research and Development in Information Retrieval, 2010.

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

[12] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting Spammers on Twitter," Proc. Seventh Collaboration, Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS), 2010.

[13] 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.

[14] C. Yang, R. Harkreader, 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.