Coordinator Based Suspicious Url Detection and Prevention of Malicious Url Attacks in Twitter's

Sandra G, Shanavas K A, Saheeda A

Abstract— Online Social Networks (OSNs), a fruitful product of technological advancement, have now become an integral part of modern social life. Twitter holds a top position among the most popularly used online social networks. As their popularity increases, security becomes a compromising factor. Many methods have been developed to prevent attacks on them. Twitter is also subjected to these attacks, which is constantly increasing. In this paper we suggest a coordinator based twitter attack detection and prevention method. The key factor of our approach is real time attack detection method which not only detects attacks, but also prevents them. Ours is a coordinator based system which always makes users land on safe sites.

Index Terms—: OSNs, Twitter, Urls, Attacks, Redirection Chain, Coordinator System, Real Time Detection.

1 INTRODUCTION

Today the network infrastructure, known as internet, consists of heterogeneous transport systems, which interconnect different persons and organizations using common network protocol. A prominent among them is social network system, which refers to a social structure consists of social actors, such as persons or organization, with a set of dyadic ties between them. The social networks [13] are basically an interdisciplinary academic field, contributed by many branches of knowledge, such as sociology, social psychology, information science, geography etc. All the social science and physical science enriched the development of social networks. They are primarily used to find out and evaluate the relationship between individuals, organization and even entire social groups. The social networks have different levels, which may be grouped into micro-level, meso-level and macro-level[15]. In micro-level, the social network begins with an individual or a small group of individuals in a particular social context. It consists of (1) dyadic level, that is social relationship between two individual, (2) triadic level , that is one individual add to a dyad and (3) actor level, which concentrates on an individual and its role in social settings. In macro level, instead of small groups of relations, large group relationship is taken into consideration. The meso- level lies between the micro and macro level connections.

At present we have large-scale networks and complex networks systems which are synonymous with macro-level system. The large numbers of social networking websites, which are extensively used throughout the world, represent the above pattern of social network relationship. Its importance and popularity may vary between countries, number of users and the focal point of its relationship. Our study is concentrated on one among the popular global social networking website, namely, Twitter [11]. It always stands among the top 20 most popular OSNs. The security issues related with twitters are a real matter of concern. The existing security methods do not eradicate all the security issues that the twitters are facing now. Though they are effective to a great extend, we cannot fully depend on these methods because all these methods have its own disadvantages. The prevailing security methods are just to detect the suspicious urls, but do not provide any real time detection or prevent users landing on the attacking sites. Real time detection methods [1] as well as methods for preventing attacks such as phishing [8] are not yet developed for twitters. In this paper we propose a system which provides a near real time detection as well as elimination of attacks by making users land only on the original websites. Section 3 gives a detailed over view of our system.

2 RELATED WORKS

Since twitter is a very popular social networking site, many active researches are carried out for identifying suspicious url detection and some of such efforts are briefed below,

2.1 Filter Based Approach

This work [2] focus on a filter based spam detection. In this approach the supervised machine learning technique is used for classifying the normal page from the malicious one. Two kinds of features are extracted here. One is html based feature and the other one is url and host based features. These features help us to distinguish between benign and malicious pages in a more accurate way. But the feature extraction is a very complex process and the time for extracting these features is also a problem.

.

2.2 Online Learning Approach

Online learning approach suggested [3] a good mechanism for identifying suspicious urls. In this approach lexical and host based features are extracted and these features are given as input to the online algorithms. This work also analyses and compares between online and batch algorithms. It suggests that online algorithms are better than batch algorithms, In online algorithms any new features can be added with the change in the environment and also online algorithm can eliminate memory overhead issues.

2.3 Redirection Chain Based Approach

A near real time based approach is suggested in [1]. In this method suspicious urls are identified on the basis of redirection chains. It consists of four major modules such as Data Collection, Feature Extraction, Training and Classification. It is the first near real time suspicious method developed so far. Its experimental results also show a good performance

rate. Its main advantages are,
(1) no need to access twitter graphs and (2) fabrication is impossible. But its main disadvantages are; (1) dynamic redirection cannot be handled and (2) its coverage is less.Taking all the advantages of above method, we propose a new system using its basic idea. Our system is a coordinator based syste em to detect the suspicious urls in twitter. Our system not only identifies the suspicious urls, but also prevents the users from landing on malicious urls sites. Hence our system aims on developing a safe environment for twitter users, which is free from any kind of malicious url attacks.

3 Proposed System

Proposed system aims on detecting suspicious urls using a coordinator system. Detection process starts with data collection, followed by feature extraction and classification.

3.1 Data Collection

Twitter steam is the input, when input is received, tweets with its urls are collected. Twitter Streaming APIs are used for collecting information about tweets and its urls. After the collection of information, API's runs crawling threads following all the url redirection chains and obtains its IP addresses. Then tweets along with urls are pushed into the tweet queues.

3.2 Feature Extraction

As suggested in [1], feature extraction module consists of three parts, entry url detection, feature vector extraction and grouping of same domains. In the tweet queue when more than 'w' tweets are collected, tweets are popped out. On these 'w' tweets domain checks are carried out and the tweets with same domains are grouped into one. From the collected urls along with the tweets, the most frequently appearing url is taken as the entry point url. Final step is the extraction of features. Here we are extracting ten different vector features such as Length of url redirect chain, Entry point url's frequency, Number of different initial URLs, Number of different landing URLs, Number of different Twitter accounts, Account creation data's standard deviation, Number of followers and number of friend's standard deviation, Follower-friend ratio's standard deviation, Number of friend's standard deviation and finally similarity of tweet texts .Out of these ten features, first six are derived from the correlated url redirection chains and the next four features are from tweet context information. After the extraction of these features, each one is normalized and F-score[1], as shown in Table 1 is used for comparing and analyzing each feature.

TABLE 1
F-SCORE OF FEATURES

FEATURES	F-SCORE
Length of url redirect chain	0.0963
Entry point url's frequency	0.0374
Number of different initial URL	0.0117
Number of different landing URLs	0.015
Number of Different Twitter accounts	0.0008
Account creation date's standard deviation	0.0680
Number of followers and number of friend's standard deviation	0.0085
Follower-friend ratio's standard deviation	0.0321
Similarity of tweet texts	0.0060

3.3 Coordinator System

Coordinator system proposed here serves the same function as other normal coordinators. The process of collecting information starts at the moment the user logins twitter. Then each of the ten vector features mentioned above are analyzed. Coordinator collects information from different twitter users and maintains it F-score as a data set. It should be noted that, when the user logs on, only new urls along with its tweets contained in his account are analyzed. Previously checked urls are not analyzed again. Coordinator system can be represented diagrammatically as in figure 1.

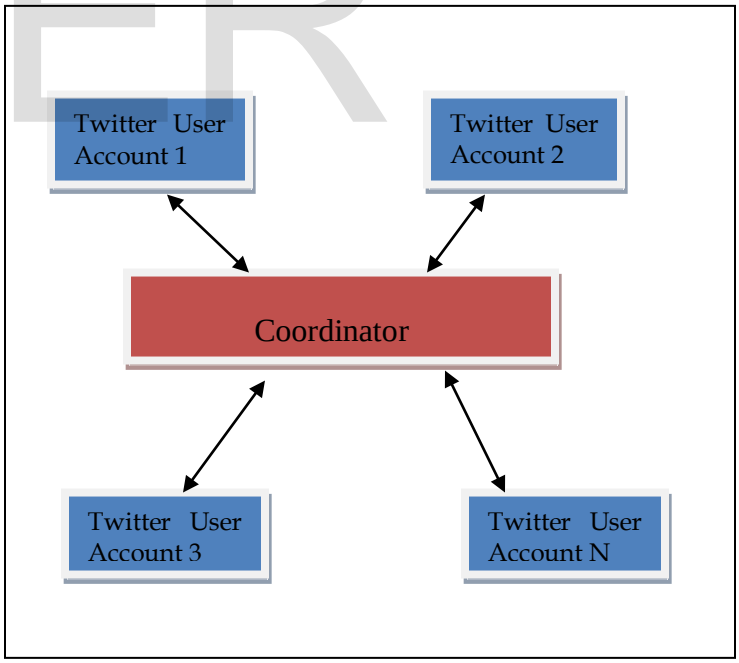

Fig 1. Coordinator System

3.4 Working

When a user logins to his twitter account, data collection pro-

cess starts at once. It is followed by feature extraction and normalization. Normalized value is analyzed with the F-score. If value of any of the feature exceeds its F-score, it is again cross checked with all the values available in the existing data set. If same feature value exceeds in the existing data sets too, then it is assigned with the value 0, else it is assigned with -1. Likewise a matrix is constructed with all the ten features across all the urls obtained from the twitter account of the logged user. From the matrix constructed, those urls whose number of features exceeding its F-score value more than or equal to 5, are black listed as suspicious ones. After preparing the black list, we perform a "Detection Test". In the "Detection Test" the blacklisted urls along with its tweets are searched in Google. For this we use Google API. By checking these links in Google we get the account creation date of the original site. Then this account creation date is compared with the account creation date of the blacklisted urls (here we take the account creation date of the site which these blacklisted urls lands in). If the account creation dates are same, urls are benign ones, and the user is redirected to the same urls and these urls are removed from the blacklist. If the account creations dates are different, urls are malicious ones and the users are blocked from being redirected to these urls by replacing the malicious url links with the original link found in Google Search. The entire system can be explained with the following example,

Let x be the user logged on to Twitter. Data collection and url feature extraction process starts as he logs in. Let the ten features be F1, F2, F3… F10 and the number of urls be U1, U2, U3, U4 and U5. After comparing normalized feature values of urls with F-score, the coordinator checks these values with the one already present in the data set. After the cross checking process, the matrix is constructed. Let the constructed matrix be the one shown in figure 2,

Fig 2. Constructed Matrix

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
U1	0	-1	0	0	0	-1	0	0	0	0
U2	0	-1	0	-1	0	-1	0	-1	0	-1
U3	-1	-1	-1	-1	-1	0	-1	-1	-1	0
U4	-1	0	0	-1	-1	-1	-1	0	0	-1
U5	0	0	0	0	0	0	-1	0	-1	0

From the matrix above the urls with 5 or more features with value '0' are listed under black list. So here urls U1,U2 and U5 are black listed. Next the 'Detection Test'is carried out on these blacklisted urls. It's performed by browsing the urls along with its tweets in Google and finding out the original sites based on the account creation date. If the urls pass the "Detection Test", the user is redirected to the same urls, otherwise urls are replaced by the original ones. Thus user always land only on original sites.

4 Experimental Analysis

Experimental analysis was done in the same method as suggested in [1] with an addition of a Coordinator System and a Detection Test. A clear division was made in the collected data set in order to have a classification between new and old data. In this experiment, we created a phishing site for "eBay" called "messbss.com", and this phishing link was posted in one of the twitter user accounts. As our system processed all the urls links, the phishing link created was also taken into account and its features were extracted. Its F-scores were measured and was compared with the database. The limit for the number of features for each url that could exceed its threshold value (F-Score) was kept as 5. After the feature extraction process, it was found the number of features that exceeded the threshold value was more than 5 in the case messbss.com(the phishing site we added) ,hence its url link was sent to the coordinator system. The coordinator system found out its fakeness by searching and comparing its account creation date with original eBay site links using the Google API search. In comparison, it was found that account creation date of phishing eBay(messbss.com) site was very new as compared with the original eBay site (shown in figure 3) and hence this phishing link was replaced with the original link. We also created a window for showing the real and phishing site in our experimental setup. In addition to this, a browser named, 'Social Phishing Media', was created to show the original site. Since our system completely prevents the users from landing in malicious urls, we can claim almost 100% efficiency for our system. We repeated this experiment for about 180 days by creating few fake links and posting it in twitter accounts .We got the same results with a good performance rate.

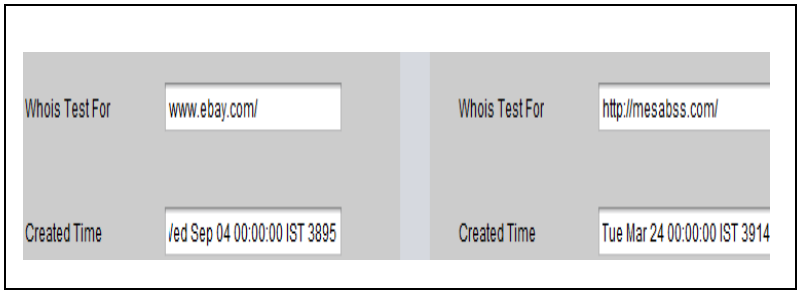

Fig 3. Detection Test's Screen Shot

5 Conclusion And Future Scope

Popular OSNs attacks are very common now days. Twitter is one of those OSNs, which is also very famous for its spamming and phishing attacks. In this paper we have proposed a real time technique using a coordinator system for detection as well prevention of malicious url attacks in Twitter. Experimental results show a good performance for attack detection as well as prevention of spamming and phishing attacks. Coordinator detection speed limit is a major constrain factor which we need to focus more on. We also aim to implement this in other Online Social Networks.

References

[1] S. Lee and J. Kim, "WarningBird: Detecting Suspicious URLs in Twitter Stream," Proc. 19th Network and Distributed System SecuritySymp. (NDSS), 2012.

[2] Davide Canali, Marco Cova, Giovanni VignaChristopher Kruegel. "Prophiler: A Fast Filter for the Large-Scale Detectionof Malicious Web Pages" In Proceedings of the 20th international conference on World wide web,2010.

[3] Justin Ma ,Lawrence K. Saul ,Stefan Savage Geoffrey M. Voelker." Identifying Suspicious URLs: An Application of Large-Scale Online Learning" .In Proc. of the International Conference on Machine Learning (ICML), 2009 .

[4] Georgios Kontaxis, Iasonas Polakis, Michalis Polychronakis and Evangelos P. Markatos "Dead. Drop: URL-based Stealthy Messaging" Proceedings of the 2011 Seventh European Conference on Computer.

[5] D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S. Ioannidis, E. P. Markatos, and T. Karagiannis, "we.b: the web of short urls," in Proceedings of the ACM international conference on World Wide Web, 2011.

[6] K. Thomas, C. Grier, V. Paxson, and D. Song. "Suspended accounts in retrospect: An analysis of twitter spam". In Internet Measurement Conf. (IMC), 2011.

[7] J. Song, S. Lee, and J. Kim. "Spam filtering in Twitter using sender-receiver relationship". In Int. Symp. Recent Advances in Intrusion Detection (RAID), 2011.

[8] M. A. Rajab, L. Ballard, N. Jagpal, P. Mavrommatis, D. No-jiri, N. Provos, and L. Schmidt. "Trends in circumventing web-malware detection". Technical report, Google, 2011.

[9] N. Nikiforakis, M. Balduzzi, S. Van Acker, W. Joosen, and D. Balzarotti, "Exposing the lack of privacy in file hosting services," in Proceedings of the 4th USENIX conference on Large-scale xploits and Emergent Threats, 2011.

[10] "Online Spam Filtering in Social Networks," Proc. 19th Network and Distributed System Security Symp. (NDSS), 2012.

[11] Tweet Attacks. Twitter marketing software that breaks the limits. http://tweetattacks.com.

[12] C. Yang, R. Harkreader, and G. Gu. "Die free or live hard? empirical evaluation and new design for fighting evolving Twitter spammers". In Int. Symp. Recent Advances in Intru-sion Detection (RAID), 2011.

[13] Scott, W.Richard , Davis,Gerate F.(2003) "Networks In and Around Organizations". Pearson Printice Hall.

[14] Scott, John P.(2000). "Social network Analysis: A Handbook" (2nd edition). Thousand Oaks, CA:Sage Publications.

[15] Granovetter,M.(176)."Network Sampling:Some First Steps". American Journal of Sociology 81(6).PP. 1287-1303.